ФГБОУ ВО СИБИРСКАЯ
ПОЖАРНО-СПАСАТЕЛЬНАЯ
АКАДЕМИЯ ГПС МЧС РОССИИ

С.В. Бабенышев
Е.Н. Материов

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Учебное пособие

Железногорск

МИНИСТЕРСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ ПО ДЕЛАМ ГРАЖДАНСКОЙ ОБОРОНЫ, ЧРЕЗВЫЧАЙНЫМ СИТУАЦИЯМ И ЛИКВИДАЦИИ ПОСЛЕДСТВИЙ СТИХИЙНЫХ БЕДСТВИЙ

ФГБОУ ВО СИБИРСКАЯ ПОЖАРНО-СПАСАТЕЛЬНАЯ АКАДЕМИЯ
ГПС МЧС РОССИИ

С.В. Бабенышев, Е.Н. Матеров

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Учебное пособие

«Допущено Министерством Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий в качестве учебного пособия для курсантов, студентов и слушателей образовательных организаций МЧС России»

Железногорск
2022

УДК 519.87 + 519.248
ББК 22.18
Б12

Авторы:
Бабенышев Сергей Валерьевич, канд. физ.-мат. наук
Матеров Евгений Николаевич, канд. физ.-мат. наук

Рецензенты:
Тараканов Д. В., доктор технических наук
(ФГБОУ Академия ГПС МЧС России)

Рыженко А. А., кандидат технических наук, доцент
(ФГБОУ Академия ГПС МЧС России)

Бабенышев С.В. Системный анализ и исследование операций. [Текст]: учебное пособие / С.В. Бабенышев, Е.Н. Матеров. – Железногорск: ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России, 2022. – 122 с.: ил.

В пособии приведены необходимые теоретические сведения и примеры решения задач, необходимые для преподавания дисциплины «Системный анализ и исследование операций» по направлению подготовки 38.03.04 «Государственное и муниципальное управление», профиль «Управление в кризисных ситуациях». В пособии используются примеры, иллюстрирующие применение классических методов исследования операций в задачах распределения ресурсов, планирования маршрутов, эффективности вложений и других. Изложение материала направлено на понимание основ применяемых методов, что позволит адаптировать их к широкому кругу практических задач оптимизации.

УДК 519.87 + 519.248
ББК 22.18

ISBN 978-5-906874-80-1

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

«Системный анализ и исследование операций» – это учебная дисциплина, изучающая общие методы моделирования и получения оптимальных управленческих решений в задачах, возникающих в экономике, логистике, планировании и инженерных приложениях. Оптимальное решение задачи рассматривается относительно некоторого выбранного числового критерия – целевой функции, каким может быть, например, наименьший расход материалов, наименьшая стоимость производства, наименьшая дальность перевозки, наибольшая по стоимости загрузка транспортного средства, минимум ошибки и т.п. Содержание дисциплины «Системный анализ и исследование операций» в значительной степени связана с рядом родственных разделов и научных дисциплин, таких как «Поддержка методов принятия решений», «Методы оптимизации», «Теория оптимального управления» и других.

Изучаемые в дисциплине методы, применяются в системах поддержки логистики и организации сервиса и производства, на транспорте, при программировании и проектировании промышленных и технологических установок и бытовых приборов.

Целью данного пособия является изложение идей и методов дисциплины «Системный анализ и исследование операций» на уровне, достаточном для усвоения материала обучающимися без специального математического образования. Для иллюстрации теоретического материала широко используются рисунки и подробный пошаговый разбор решения основных примеров. В целях сопровождения изучения дисциплины с помощью этого пособия, рекомендуется использовать комплект индивидуальных однотипных задач для каждого рассмотренного в пособии вида основных примеров. Примеры подобраны так, чтобы при самостоятельном прорешивании были усвоены основные идеи и ограничения методов, рассмотренных в теоретической части.

Структурно пособие разделено на пять тем: «Основные понятия системного анализа и исследования операций», «Принципы динамического программирования», «Задачи линейного программирования», «Элементы теории игр» и «Основы теории массового обслуживания». Глава 1 излагает важнейшие понятия системы, абстрагирования и моделирования. Главы 2 и 3 связаны общими подходами и подкрепляют усвоение базовых понятий методов оптимизации. Главы 4 и 5 рассматривают моделирования в условиях неопределённости, в том числе, статистически описываемой неопределённости для Главы 5. Внутри подразделов каждой темы используется отдельная нумерация рисунков, формул и примеров.

ГЛАВА 1. ОСНОВНЫЕ ПОНЯТИЯ СИСТЕМНОГО АНАЛИЗА И ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

Основы моделирования

Модели и моделирование

Как практическая, так и познавательная деятельность человека и общества широко опирается на процесс ***моделирования***. Создаваемые для этого ***модели*** – это обычно упрощенные образы объектов или процессов, которые, тем не менее, должны с приемлемой точностью отражать основные черты исследуемых более сложных систем. Методологически, объекты и процессы могут считаться специфическими подклассами более широкого класса – класса *явлений*.

Если полученная модель достаточно ***адекватна***, то её можно использовать при исследовании вместо самого объекта. Таким образом, создаваемая при моделировании модель должна по возможности удовлетворять двум часто противоречащим требования:

- быть максимально простой;
- достаточно точно отражать моделируемой явление.

Например, лежащая в основе теории Ньютона посылка, что в механической системе тела можно представить материальными точками с приложенными к ним векторами-силами, заложила основы для моделирования широчайшего спектра механических явлений, от движения планет вокруг звезд, до полета снарядов и динамики механизмов. Теория Ньютона, опирающаяся всего на три базовых принципа – ***законы Ньютона***, из-за её простоты и приемлемой точности, до сих пор применяется для моделирования механических систем, несмотря на существование её уточнений – квантовой механики и теории относительности. С другой стороны, критически важное с точки зрения сохранения жизни и имущества людей, точное моделирование лесных пожаров требует учёта направления ветра, рельефа (огонь распространяется быстрее вверх по склону), влажности почвы и растительного покрова, учёта высыхания и нагревания воздуха под действием самого пожара. Необходимость учитывать эти факторы в большом количестве точек на местности приводит к тому, что временные затраты на точное моделирование приближаются к времени развития самого пожара.

Итак,

> ***Правильный выбор баланса между простотой и точностью применяемой модели является важнейшей задачей исследователя.***

Понятия «модель» и «моделирование» особенно широко используются в сфере науки, образования, инженерном и технологическом проектировании, в серийном техническом производстве. В этих областях термин «модель» обычно используется для обозначения:

- устройства, воспроизводящего строение и/или действие какого-либо другого устройства (уменьшенного, увеличенного или в натуральную величину),
- символического аналога (чертежа, плана, схемы, описания, ...) конкретного явления, объекта или процесса,
- системы математических уравнений или соотношений, связывающих наблюдаемые параметры исследуемой системы.

Важное место при составлении моделей принадлежит исходным ***гипотезам*** и наблюдаемым ***аналогиям***:

- ***гипотеза*** – это предположительное суждение о свойствах исследуемого явления, основанное на эмпирических данных, наблюдениях или догадках;
- ***аналогия*** – это представление о каком-либо частном сходстве, существенном или несущественном, в зависимости от уровня абстрагирования, определяемого конечной целью исследования.

Гипотезы и аналогии, в определенной мере отражающие реальный, объективно существующий мир, должны обладать наглядностью или сводиться к удобным для человека логическим схемам.

Обобщая вышесказанное, приходим к следующим определениям:

Модель – такой материальный или идеальный (мысленно представляемый) объект, который в процессе познания (изучения) замещает оригинал, сохраняя при этом некоторые его типичные черты, важные для данного исследования.

Моделирование – процесс построения и использования модели.

Краткая характеристика моделей

Все модели и методы моделирования с определенной условностью могут быть разделены на:

материальные (реально существующие);
идеальные (мысленно воображаемые).

Примерами материальных моделей служат лабораторные установки, устройства-демонстраторы, макеты автомашин, строений, городов, а идеальных – описание или представление любых явлений, процессов и предметов с помощью графических и математических символов, и слов.

Среди идеальных моделей иногда еще выделяют ***когнитивный*** тип модели, понимая под ним мысленный образ конкретных объектов.

Наряду со сложностью и адекватностью, ***важнейшей*** характеристикой модели является ее ***предсказательная сила***, то есть ее пригодность для получения новых знаний об объекте-оригинале, например, предсказания его поведения в новых или запроектных ситуациях.

Для достижения предсказательной силы, как уже было отмечено, используются два основных подхода к моделированию сложных объектов или процессов: моделирование как ***системы*** и моделирование как ***черного ящика***.

При первом подходе, объект или процесс методологически разбивается на существенные более простые части – ***элементы системы***, и устанавливаются качественные и количественные характеристики взаимосвязей этих элементов, то есть устанавливается ***структура*** исследуемого явления. При адекватном представлении системой, реакцию моделируемого объекта на заданный набор входных параметров можно просчитать по системе, исходя из взаимодействия её элементов. Системный подход применим к моделированию как физических, так и более сложных технологических объектов, организационных структур и процессов, но хуже применим к моделированию целостных явлений с трудом поддающихся анализу, например, явлений в биологии или социологии, например, в силу свой *динамичности* (например, в силу изменчивости структуры за счет создания/смерти элементов или изменчивости и перераспределения связей), *целостности* (так называемые, *холистические* объекты, описание функционирования которых не удается свести к описанию взаимодействия их элементов) или *многообразности* (то есть объектов, число важных элементов или взаимосвязей которых или их взаимосвязей существенно превышает то, что может удержать человеческий разум ($\gg$7), как например, биологические, экономически или социальные системы, из более простых систем – веб-поисковики).

Второй подход, подход к исследуемого объекту как к *черному ящику*, не берется устанавливать структуру исследуемого объекта. Взамен рассматривается многообразие известных реакций объекта на какой-то класс входных воздействий. Развитие этого подхода для моделирования сложных объектов, стало возможно благодаря развитию технологии программных нейронных сетей, имитирующих работу биологических систем. При таком подходе описание работы исследуемой системы, например, распознавание изображений водителем автомобиля, выбирается из достаточно широкого класса функций, которые могут быть реализованы нейронными сетями, путем процесса, называемого *обучением*. Одним из недостатков этого безусловно перспективного подхода, однако является плохое понимание функционирования такой модели, например, то на какие аспекты исследуемого явления она опирается при моделировании, и как результат невозможность надежно предсказать поведение модели за пределами обучающей выборки.

В целом, системный подход является более распространенным и имеет то преимущество, что дает еще и понимание работы исследуемого объекта, например, какой элемент или взаимосвязь является ключевым, где лежат пределы применимости модели и так далее.

Таким образом, мы ознакомились с наиболее существенными признаками и свойствами моделей. В частности, установили, что модели и моделирование нужны для того, чтобы

а) понять, как устроен конкретный объект-оригинал; каковы его структура, основные свойства и/или установить закономерности его функционирования и развития, в случае, когда структуру и детали функционирования объекта установить трудно или принципиально невозможно;

б) научиться управлять объектом и процессом его функционирования, в том числе определять *оптимальные* (наилучшие) для него управляющие воздействия при заданных целях и критериях;

в) прогнозировать прямые и косвенные последствия реализации конкретных способов и форм воздействия на моделируемый объект.

Контрольные вопросы

1. Что такое абстрактная модель и чем занимается моделирование?
2. Каким двум противоречащим требованиям должна удовлетворять модель явления?
3. Для описания каких трех видов подобий исследуемой системы используется термин «модель»?
4. На основе каких двух видов суждений и предположений строится модель явления, объекта или процесса?
5. Чем материальные модели отличаются от идеальных? Приведите примеры каждого вида.
6. Каков основной критерий адекватности построенной модели?
7. Какие два принципиально различных подхода используются при моделировании?

Классификация моделей и методов моделирования

Далее рассмотрим классификацию моделей и методов, используемых при моделировании. Неоднозначность термина «модель», огромное количество видов моделей и способов их использования, а также их быстрое развитие в настоящее время затрудняют построение единой логически стройной и удовлетворяющей всех классификации. Взамен, мы далее покажем, как модели классифицируются по ряду важных независимых признаков.

Классификация по природе и типу представления

Систематизация известных к настоящему времени моделей и методов их использования приводит к следующей классификации по природе и типу формализованного представления.

МОДЕЛИРОВАНИЕ

- **Идеальное**
  - Интуитивное
    - *Мысленный эксперимент*
    - *Метод сценариев*
    - *Операционная игра*
  - Семантическое
    - *Вербальное*
    - *Графическое*
  - Семиотическое
    - *Математическое*
      - *Аналитическое*
      - *Алгоритмическое*
        - *Численное*
        - *Имитационное (симуляционное)*
- **Материальное**
  - Аналоговое
  - Физическое

Рис. 1.1. Классификация методов моделирования. Отступ перед названием вида моделирования означает, что он является подклассом вышестоящего вида.

Поясним особенности всех основных видов моделирования:

Материальные (реальные, натурные или предметные) модели являются вторичными по отношению к идеальным и основаны на использовании свойства подобия между ними и прототипом.

Физические модели обычно являются геометрически подобными оригиналам, а ***аналоговые*** – напротив, физически.

Методы ***физического*** (натурного, предметного) моделирования нашли широкое применение в авиа-, автомобиле-, ракето- и судостроении, а также в других отраслях промышленности и транспорта. Дополнительный импульс физическое моделирование получило в связи с развитием аддитивных технологий.

Аналоговое моделирование использует совпадение (преимущественно – качественное) математического описания различных предметов, процессов и явлений.

Пример. Механические и электрические колебания, которые подчинены одним и тем же законам, но относятся к качественно различным физическим процессам; при некоторых допущениях аналогичными можно считать большинство процессов, протекающих в газе и жидкости, включая обтекание их потоками различных тел, а также явления теплопереноса и диффузии примесей.

Основное удобство аналоговых моделей заключается в том, что изучение одних процессов можно проводить в других, более удобных условиях. Например, изучение тех же механических колебаний можно вести с помощью электрической схемы, а обтекание жидкости заменить обтеканием газом, и наоборот.

Под ***интуитивным*** (иногда называемым «ненаучным») обычно подразумевают моделирование, использующее не обоснованное с позиций формальной логики представление объекта исследования, которое к тому же не поддается формализации или не нуждается в ней.

Такое моделирование осуществляется в сознании человека, в форме мысленных экспериментов, сценариев и игровых ситуаций с целью его подготовки к предстоящим практическим действиям за счет заблаговременной преднастройки к ним.

Семантическое (смысловое) моделирование логически обосновано с помощью некоторого числа исходных предположений. Сами эти предположения нередко принимают форму гипотез, создаваемых на основе наблюдения за объектом моделирования или какими-либо его аналогами.

Главное отличие этого вида моделирования от интуитивного заключается не только в умении выполнять и воспроизводить для других его действия, но и в знании внутренних механизмов, которые используются при этом.

В группу ***семантических*** методов входит ***вербальное*** (словесное) и ***графическое*** моделирование. При этом первый тип моделей образуется с помощью слов, из которых составляются высказывания, суждения и умозаключения относительно моделируемого объекта.

Графическое моделирование использует материальные носители информации – бумага, классная доска или монитор компьютера, на которых размещаются различные рисунки, чертежи, структурно-функциональные схемы или диаграммы причинно-следственных связей.

Семиотическое, или ***знаковое***, моделирование является наиболее формализованным, поскольку использует не только общеизвестные слова или довольно наглядные изображения (как в семантических моделях), но и специализированные знаки, ассоциированные с исследуемым явлением или понятиями – буквы, иероглифы, нотные знаки, цифры. Более того, в последующем все они объединяются с помощью специфических правил, по которым принято оперировать как отдельными элементами, так и создаваемыми из них знаковыми образованиями (формулы).

Основным и наиболее представительным подвидом данного моделирования считается математическое моделирование.

Математическим моделированием будем называть

> ***<u>идеальное</u> <u>знаковое</u> <u>формальное</u> моделирование, при котором описание объекта-оригинала осуществляется на языке математики, а исследование модели проводится с использованием тех или иных математических методов.***

Классификация по цели моделирования

По признаку ***цели моделирования*** математические модели делятся на:

- ***дескриптивные*** (***описательные***);
- ***нормативные*** (***оптимизационные***);
- ***ситуационные*** (***управленческие***).

Дескриптивные модели являются описаниями признаков моделируемых объектов и объясняют законы изменения их параметров с помощью слов, рисунков или формул.

Нормативные или ***оптимизационные*** модели имеют целью не столько отражение действительности, сколько определение желательного способа поведения системы.

Ситуационные или ***управленческие*** модели предназначены для выявления наиболее существенных для моделируемого объекта факторов и ***априорной*** оценки его основных количественных характеристик.

Следующие области применения предпочтительны для использования моделей каждого из трех классов:

- ***дескриптивные*** модели – для словесной, графической и математической интерпретации объекта системного анализа и моделирования процессов в техносфере;
- ***нормативные*** – для обоснования или уточнения значений показателей безопасности их проведения;
- ***ситуационные*** – для исследования явлений и процессов, оказывающих наиболее существенное влияние на возникновение и предупреждение техногенных происшествий.

Классификация по способу исследования

По ***способу исследования*** математические модели делят на два основных класса: ***аналитические*** и ***алгоритмические***.

Аналитическое моделирование позволяет получить выходные результаты в виде аналитических выражений (формул), использующих конечное или счетное число арифметических операций и переходов к пределу по натуральным числам.

Широко известные примеры включают алгебраические модели и модели, основанные на дифференциальных уравнениях.

Пример 1.1 (*Движение тела, брошенного под углом к горизонту.*)

При движении тела, брошенного со скоростью v под углом α к горизонту, его координаты меняются следующим образом:

$$\begin{cases} x(t) = x_0 + (v \cdot \cos\alpha) \cdot t \\ y(t) = (v \cdot \sin\alpha) \cdot t - \dfrac{g \cdot t^2}{2} \end{cases}$$

где: x_0 – начальное смещение тела вдоль оси Ox;

g = 9,81 $м/с^2$ – ускорение свободного падения.

Таким образом, так как уравнения связывают только переменные (x, y) и степени переменных (t^2), то есть являются алгебраическими, то мы имеем ***алгебраическую модель*** движения тела.

Пример 1.2 (***Задача управляемого одномерного движения тела.***)

Допустим, что тело движется вдоль горизонтально расположенной прямолинейной рельсовой дорожки, причем движением тела можно управлять, придавая ему в любой момент времени некоторое ускорение (тягу) $u(t)$ вперед или назад. Тогда движение тела задается дифференциальным уравнением (включающим производные переменных величин):

$$\dot{x} = u, \qquad x(0) = x_0;$$

где: x_0 – начальное смещение тела вдоль оси Ox;

$\dot{x}$ – производная смещения вдоль оси Ox по времени.

Таким образом, здесь мы имеем модель движения управляемого тела, с ***дифференциальными уравнениями***.

В более сложных и более приближенных к практике случаях, в одной модели возможны комбинации алгебраических, дифференциальных и интегральных уравнений, выражений и неравенств, как, например, в следующей задаче.

Пример 1.3 (***Простейшая задача оптимального управления***)

Пусть материальная точка массой $m = 1$, движется без трения по горизонтальной прямой (рис. 1.2). Пусть эта точка снабжена двигателем, развивающим силу тяги $u(t)$ такую, что $|u(t)| \le 1$.

Рис. 1.2. Движение материальной точки вдоль прямой.

Пусть требуется привести точку в начальное положение с конечной нулевой скоростью (например, под загрузку) за наименьшее время.

Формальное описание модели:

Введем обозначения

$$x_1(t) = x(t), \; x_2(t) = \dot{x}(t) = v(t).$$

Тогда уравнения движения имеют вид:

$$\begin{cases} \dot{x}_1(t) = x_2(t), & x_1(0) = x_1^0, \\ \dot{x}_2(t) = u(t), & x_2(0) = x_2^0. \end{cases}$$

а вся модель задачи оптимизации имеет вид:

$$J(x_1(t), x_2(t), u(t)) = \int_0^T dt = T \to \inf,$$

$$|u(t)| \le 1, \; x_1(T) = 0, \; x_2(T) = 0 \; .$$

Рассматривая классификацию далее, ***алгоритмические*** модели, в свою очередь, используют программные алгоритмы для численного или имитационного моделирования исследуемой системы.

Алгоритмические модели часто строятся на основе уже имеющихся аналитических моделей. При этом алгоритмическое моделирование вносит свои специфические погрешности, связанные, прежде всего, с конечностью представления вещественных чисел в компьютерах (ошибки приближения) и дискретностью представления времени, связанной с дискретной природой современных цифровых вычислительных сред, а также связанной с вышеуказанными погрешностями, устойчивостью/неустойчивостью модели по начальным данным.

Классификация по виду входных данных и результатов

По виду входных данных и результатов моделирования математические модели делятся на:

- детерминированные;
- стохастические;
- случайные;
- интервальные;
- нечеткие.

Основное отличие ***детерминированных*** моделей заключается в том, что каждому фиксированному входному набору параметров однозначно соответствует результат моделирования (имеется *функциональная зависимость*).

В ***стохастической*** модели значения всех или отдельных параметров определяются случайными величинами, то есть как входные данные, так результаты моделирования задаются возможными значениями с соответствующими плотностями вероятностей.

В отличии от стохастической, в ***случайной*** модели все параметры являются конечным случайными величинами, найденными в результате статистической обработки ограниченной выборки (т.е. эмпирически) и представленными в виде оценок соответствующих плотностей вероятности, а потому и потенциально менее точными, по сравнению со стохастической моделью.

Заметно более неопределенные параметры имеют ***интервальные*** модели, в которых вместо точечных оценок их значений (как в предыдущем случае) используются интервальные.

Примерно этот же способ представления параметров применяется и в ***нечетких*** моделях, которые уже оперируют нечеткими величинами или числами, также заданными на некоторых интервалах возможных значений.

Особое положение занимают модели, основанные на моделировании с помощью нейронных сетей. Основное их достоинство в том, что для их построения не требуется знать внутреннее устройство моделируемой системы, а достаточно иметь представительный объем наблюдений поведения этой системы.

Так, например, можно создавать машины технического зрения для распознавания объектов или даже ситуаций и сцен, просто используя отклики экспертов-людей. Существенный недостаток таких систем на данном этапе их развития, это большой объем требуемой обучающей выборки (>100 000 откликов, что, например, невозможно получить при моделировании сложных и уникальных технических систем, вроде атомного реактора или авиационного двигателя). Второй недостаток нейронных сетей тесно связан с их достоинствами – даже при достаточно хорошем соответствии поведения обучения нейронной сети на обучающей выборки, нет гарантии, что на редко встречающихся ситуациях (входах), модель поведет себя адекватно.

Список неучтенных выше типов моделей можно продолжить также за счет включения в него математических моделей, параметры которых имеют различное отношение, допустим:

а) ко времени – «статическая», «динамическая»;

б) к размерности пространства – «одномерная», «многомерная»;

и так далее.

Контрольные вопросы

1. Чем отличаются идеальное и материальное моделирование?
2. Какие два вида моделирования относятся к материальному моделированию?
3. Какие виды моделирования называют интуитивными?
4. Какие виды моделирования относятся к семантическому классу моделирования?
5. Какое моделирование называется математическим?
6. К какому виду моделирования относится математическое моделирование, семантическому или семиотическому?
7. По целям моделирования на какие классы делят математические модели?
8. Какие цели преследуют дескриптивная модели?
9. Какие цели преследуют нормативные модели?
10. Какие цели преследуют ситуационные модели?
11. Чем отличаются аналитические и алгоритмические (имитационные) математические модели?
12. Как разделяются математические модели по виду входных данных и результатов моделирования?
13. Дайте краткую характеристику каждого из классов математических моделей: детерминированные, стохастические, случайные, интервальные, нечеткие?
14. Как можно использовать нейронные сети для моделирования?

Системный анализ и моделирование

Анализ, синтез, оптимизация

Моделирование какого-либо сложного объекта можно свести к следующим трем шагам: анализ, синтез, оптимизация.

Анализ сводится к расчленению, выделению и рассмотрению отдельных существенных для функционирования частей объекта, т.е. элементов.

Синтез устанавливает связи между зафиксированными элементами этого объекта, т.е. синтезирует его интегральные (целостные) свойства.

Оптимизация вырабатывает рекомендации, реализация которых может способствовать повышению качества исследуемого объекта.

Другими словами, анализ делает известными отдельные составляющие части (элементы, аспекты) сложного объекта и свойства этих частей как уже самостоятельных образований. Синтез же систематизирует представления, добытые в результате анализа. При этом, именно, анализ выделяет и рассматривает те отличительные и существенные признаки и отношения между компонентами объекта, в силу которых они могут считаться частью какого-то целостного образования, и которые, следовательно, являются существенными для синтеза. При этом, при анализе выделенные свойства и связи должны быть *стабильными*, т. е. сохраняться при внешних и внутренних возмущениях. Оптимизацией, т.е. отысканием оптимальных (наилучших в некотором смысле) или рациональных решений и управлений занимается научная дисциплина «Исследование операций», а основными подходами по их отысканию служат эвристический поиск и нахождение экстремума методами математического анализа или математического, динамического, линейного, нелинейного или выпуклого программирования.

Общая методология исследования и совершенствования больших и сложных систем базируется на их рассмотрении по таким аспектам:

а) системно-элементный, качественно и количественно характеризующий состав системы;

б) системно-структурный, концентрирующий внимание на способах связи и организации взаимодействия ее элементов;

в) системно-функциональный, учитывающий задачи основных компонентов системы;

г) системно-коммуникативный, рассматривающий ее вертикальные и горизонтальные связи с другими объектами;

д) системно-интегративный, определяющий факторы самосохранения и самосовершенствования сложной системы;

е) системно-исторический, выявляющий условия ее возникновения, развития и гибели.

По сравнении с альтернативными исследовательскими инструментами – статистическим и экспериментальным, такой подход обладает тремя особенностями:

а) статистический подход требует отлаженной системы сбора и обработки конкретной информации, а также малоэффективен в тех случаях, когда отсутствуют данные, необходимые для оценки эффективности принципиально новых проектов, и затруднителен из-за невозможности учета всего опыта, накопленного в других сложных системах, по причине их существенного различия;

б) экспериментальный же подход не обеспечивает требуемой оперативности выявления интересующих исследователя закономерностей и требует больших затрат на проведение натурных испытаний; хуже того, он не может быть использован для опасных технологических процессов и объектов, поскольку это связано с угрозой здоровью людей, крупным ущербом материальным и природным ресурсам;

в) системное моделирование лишено части перечисленных недостатков, хотя и требует определенного времени – для подготовки высококвалифицированных специалистов, разработки моделей интересующих их процессов, а затем и для качественного и количественного анализа этих моделей.

В то же время статистический анализа и непосредственный эксперимент могут использоваться как средство получения и обработки исходных данных, необходимых для моделирования либо проверки достоверности полученных с помощью моделирования результатов.

Процесс анализа может приводить к существенному упрощению представления исследуемого объекта, с потерей некоторых деталей описания несущественных для функционирования объекта в поставленной задачи. Однако при другой постановке опущенные детали могут оказаться существенными.

Пример 1.4 Для обеспечения бесперебойного функционирования группировки сил и средств в зоне ЧС формируется план загрузки автомобиля, с учетом того, что движение будет проходить по дорогам общего пользования и по участку дороги с ограничением максимальной грузоподъемности до 600 кг (аварийный мост). К доставке планируются (для наглядности условия задачи намеренно упрощены по ассортименту и весу):

- продукты (имеются 10 поддонов весом по 100 кг);
- инструменты для оборудования лагеря (3 ящика по 80 кг каждый);
- теплые спальные мешки и одеяла (6 упаковок по 30 кг);
- питьевая вода (10 канистр по 40 кг);

– медицинские препараты и припасы (2 комплекта по 20 кг);
– средства для обеззараживания воды (5 упаковок по 20 кг);
– дизель-генератор (2 штуки по 80 кг каждый);
– топливо для дизель-генератора (10 канистр по 24 кг).

Создайте модель загрузки для разрабатываемой автоматизированной системы поддержки логистики. Предложите вариант загрузки.

Решение. На первом этапе, выделим существенные элементы модели. Конкретные наименования доставляемых грузов могут нести избыточную информацию, и их лучше заменить на номера. Количество груза превышает располагаемую грузоподъемность. Для формирования плана загрузки необходимо определить важность (условную «стоимость») каждого наименования для снабжаемой группировки средств. Указанную стоимость будем назначать субъективно и выражать числовым значением от 0 до 10.

Например, может быть представлена такая модель (табл. 1.1):

Таблица 1.1

№	количество мест, шт	вес одного места, кг	условная стоимость	словесное наименование
1	10	100	**4**	продукты
2	3	80	**6**	инструменты
3	6	30	**2**	спальные мешки и одеяла
4	10	40	**1**	питьевая вода
5	2	20	**10**	медицинские препараты
6	5	20	**9**	средства для обеззараживания воды
7	2	80	**5**	дизель-генератор
8	10	24	**5**	топливо

Примечание. Жирным шрифтом в таблице выделены субъективно назначенные условные «стоимости».

Обозначим через x_i – количество загружаемых мест i-го груза. Тогда в качестве плана загрузки можно предложить следующий набор:

$x_1 = 1$; $x_2 = 2$; $x_3 = 1$; $x_4 = 1$; $x_5 = 4$; $x_6 = 2$; $x_7 = 1$; $x_8 = 0$.

Суммарный вес загрузки:

$$W = 1 \cdot 100 + 2 \cdot 80 + 1 \cdot 100 + 1 \cdot 40 + 4 \cdot 20 + 2 \cdot 20 + 1 \cdot 80 + 0 \cdot 24 =$$

$$= 100 + 160 + 100 + 40 + 80 + 40 + 80 + 0 =$$

$= 600$ кг.

Суммарная «стоимость» загрузки равна:

$$C = 1{\cdot}4 + 2{\cdot}6 + 1{\cdot}2 + 1{\cdot}1 + 4{\cdot}10 + 2{\cdot}9 + 1{\cdot}5 + 0{\cdot}5 =$$

$$= 4 + 12 + 2 + 1 + 40 + 18 + 5 + 0 + 0$$

$$= 82.$$

Замечание. Можно было бы предложить план загрузки для данной задачи с большей чем 82 «стоимостью». Более того, как будет показано далее, существует алгоритм, построенный на основе принципов динамического программирования, который может найти оптимальный (с наибольшей «стоимостью») план загрузки для текущей задачи.

В ходе построения модели для поставленной задачи можно наглядно наблюдать следующие общие особенности процесса моделирования:

1. Существенное упрощение в представлении явления. (Информация про ЧС, группировку сил и средств, особенности маршрута, тип транспортного средства, наименование груза, тару и объем тары хранения, отброшена, как несущественная для процесса загрузки.)

2. Часть отбрасываемой информации не имеет значения только в условиях поставленной задачи и может оказаться важной в других условиях. Так, например, знание того, что груз направляется в зону ЧС, указывает на срочность загрузки и т.п. Наименования груза могут нести информацию о важности того или иного груза для управляющего загрузкой человека, но не несут этой информации для планируемой автоматизированной системы поддержки логистики.

3. Более абстрактный характер модели по сравнению с моделируемым явлением. (Словесные наименования заменены на числовые идентификаторы, физические действия по загрузке заменены на арифметическую операцию сложения и т.д.)

Задачи для самостоятельного решения.

В следующих задачах, модифицируйте модель для системы поддержки логистики (Пример 1.4), так чтобы выполнялись указанные дополнительные условия, предложите план загрузки.

1. Дополнительные условия:

а) должна быть обеспечена доставка как минимум одного дизель-генератора и двух медицинских комплектов;

б) одновременно может перевозиться не более двух канистр с горючим.

<u>Указание.</u> В пункте а) выделить один дизель-генератор и два необходимых медицинских комплекта в отдельные группы с заведомо завышенными

«стоимостями» объектов. Тогда алгоритм оптимального планирования загрузки заведомо включит их в план.

2. Дополнительные условия:

а) должна быть обеспечена доставка как минимум одного дизель-генератора и два медицинских комплекта;

б) одновременно не могут перевозиться более двух канистр с горючим.

3. Дополнительные условия:

а) должна быть обеспечена доставка как минимум одного места для каждого наименования;

б) если перевозиться дизель-генератор, то обязательно должна быть доставлена, как минимум, одна канистра с горючим.

В заключение, обратим внимание на ряд дополнительных трудностей, сопутствующих системному анализу и системному синтезу процессов и явлений в таких сложных объектах, как рассматриваемые человеко-машинные системы, не говоря уже о техносфере в целом:

1. Большое число факторов, реально влияющих на человеко-машинную систему. С некоторым преувеличением можно утверждать, что на процесс ее функционирования влияет буквально все или почти все, тем самым с трудом выделяются главные взаимосвязи, что приводит к высокой сложности модели, которая может быть сопоставима со сложностью моделируемого явления.

2. Дефицит или низкое качество имеющейся информации, что делает ее зачастую непригодной для моделирования. Указанные причины обусловлены дефицитом моделей, позволяющих сформулировать требования к составу и параметрам оперируемых ими исходных данных.

Контрольные вопросы

1. Что такое система и из чего она состоит?
2. Существуют ли в природе системы как таковые?
3. Что называют структурой системы?
4. Какие основные признаки используются для классификации систем?
5. Приведите пример закрытой и изолированной системы?
6. В чем состоят принципиальные отличия между сложными и простыми системами?
7. Можно ли по внешнему виду судить о предназначении системы?
8. Какие выводы следует сделать из принципа, утверждающего о том, что причиной большинства проблем является сама система?
9. Укажите связи между системным анализом, системным синтезом и оптимизацией?
10. Как соотносятся между собой системный анализ и моделирование?

ГЛАВА 2. ПРИНЦИПЫ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Динамическое программирование

Постановка задачи динамического программирования

Термин «динамическое программирование» (далее ДП) ввёл американский математик Ричард Беллман в 1940-х гг. «Программирование» в этом сочетании имеет значение «оптимизация», также как в выражениях «математическое программирование», «линейное программирование», «выпуклое» и «целочисленное программирование». В этом значении термин «программирование» связан с организационным планированием, как, например, в высказывании «Программа переоснащения вооруженных сил».

Область динамического программирования также называлась «системный анализ и инжиниринг» и с самого начала имела практическую направленность.

Динамическое программирование – способ решения задач оптимизации путем разбиения их на более простые подзадачи, схожие по структуре с исходной, с последующим использованием полученных решений для построения решения исходной задачи.

ДП применимо к задачам, имеющем ***оптимальную подструктуру***, то есть таким, в которых оптимальные решения ***перекрывающихся*** подзадач могу быть использованы для синтеза оптимального решения исходной задачи.

Смысл введенных понятий проиллюстрируем на примерах.

Пример 2.1 Найти кратчайший путь во взвешенном ориентированном графе между двумя вершинами A и B (A и B – пункты выезда и назначения, v_2, …, v_{14} – перекрестки, двигаться можно только вправо или вниз). Граф может представлять условную схему проезда автомобиля по городу в час пик с указанием затрат времени при проезде вдоль участков дорог (рис. 2.1).

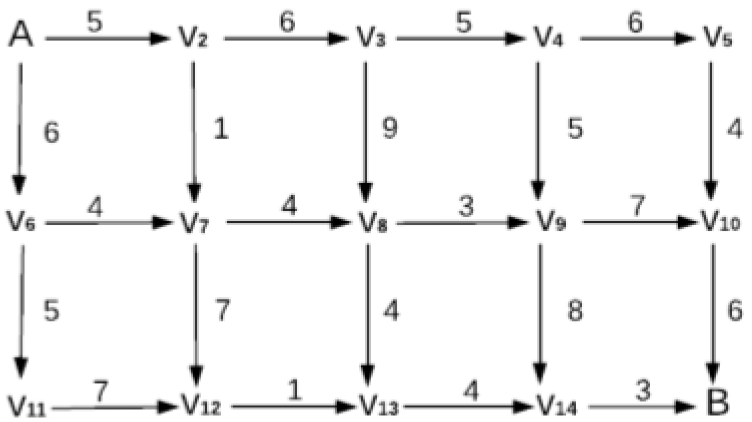


Рис. 2.1. Схема проезда из пункта A в пункт B с указанием затрат времени.

В общем случае граф может быть какого угодно вида, например, граф дорожной сети города непрямоугольной планировки, с многоуровневыми развязками, переулками, дворовыми проездами и т.п. Помимо затрат времени, веса на таком графе могут означать: расстояния, расход топлива на прохождение участка, затраты на строительство, риски и так далее. Решение этой задачи используется в навигационных программах, логистических расчетах, компьютерных играх и т.п.

Математическое понятие графа подходит для описания всех этих и многих других ситуаций и поэтому является базовым для соответствующего вида задач динамического программирования.

Пример 2.2 ***Задача о загрузке транспортного средства*** (*«задача о загрузке корабля», «задача о рюкзаке» или «задача линейного раскроя»*).

В общей постановке задача формулируется следующим образом: пусть имеется транспортное средство грузоподъёмности W, которое может быть загружено предметами n типов. Если обозначить через w_k и c_k вес и ценность предмета k-го типа (от английских “weight” – вес и “cost” – цена), через x_k –количество таких предметов, то можно поставить задачу загрузки транспортного средства грузом максимальной ценности.

Алгоритмы решения этой задачи используются в логистике, при планировании эффективного раскроя материалов и т.п.

Пример 2.3 Пример ***задачи о распределении ресурсов***.

Городское управление получило N пожарных автоцистерн, которые следует распределить между K пожарными частями. Каждая из пожарных частей F_i, $i=1, \dots, K$, при поступлении в неё n автоцистерн повышает уровень технической готовности машинного парка до какого-то значения, зависящего от n, т.е. представляющего собой функцию $f_i(n)$. Все функции

$$f_i(n), \quad i=1,\dots,K, \quad n=0,\dots,N,$$

заданы. Как следует распределить закупленную технику, чтобы в сумме это давало максимальное значение технической готовности по управлению?

Алгоритмы решения этой задачи используется при организационном и производственном планировании.

Изучение методов динамического программирования начнем с формулировки общей схемы решения.

Общая 3-шаговая схема решения задач с оптимальной подструктурой

1. Разбиение задачи на подзадачи меньшего размера.
2. Рекурсивное нахождение оптимальных решений подзадач, применяя эту же схему.
3. Синтез общего решения из полученных оптимальных решений подзадач.

Выводы. Итак, метод динамического программирования опирается на следующие необходимые свойства задачи:

1) наличие перекрывающихся схожих подзадач;

2) наличие оптимальной подструктуры;

3) возможность запоминания решений часто встречающихся подзадач.

Понимание этих свойств позволяет определить какие задачи можно решать с помощью ДП и какие вычислительные средства и ресурсы нужны для этого.

Принцип поэтапного построения оптимального уравнения

Принцип поэтапного построения оптимального уравнения рассмотрим на примере детерминированной задачи динамического программирования в дискретном времени.

Общая постановка задачи

***Дискретное время**: $t = 0, 1, \ldots, n$ (шаги исполнения).*

В качестве шагов исполнения могут выступать, например, количество пройденных кварталов, видов загруженных товаров, уже укомплектованных пожарных частей.

Обозначим:

x_t – состояние системы, наступившее на t-ом шаге *(положение на карте, оставшаяся грузоподъемность, количество нераспределенных автомобилей на текущем шаге)*;

a_t – действие (управление, решение) предпринятое на t-ом шаге *(двигаться вниз или вправо, и т.д.)*.

Для состояния x_t обозначим:

$\Gamma(x_t)$ – множество управляющих действий допустимых в состоянии x_t;

$V(x_t)$ – оптимальный выигрыш, который может быть получен, начиная с состояния x_t.

Если на шаге t в состоянии x_t предпринято действие a_t, то обозначим:

$T(x_t, a_t)$ – результирующее состояние системы x_{t+1},

$f(x_t, a_t)$ – полученный в результате выигрыш.

Тогда (абсолютный) оптимальный выигрыш $V(x_0)$ для всей задачи может быть записан как

$$V(x_0) = \max_{\bar{a}} \sum_{t=0}^{n} f(x_t, a_t),$$

где $\overline{a} = (a_0, \ldots, a_n)$ – обозначает последовательность управлений, для которых выполнены условия

$$a_t \in \Gamma(x_t), \qquad x_{t+1} = T(x_t, a_t).$$

Замечание. Простейший, так называемый «наивный», способ решения этой задачи состоит в переборе всех возможных векторов управлений $\overline{a}$, сначала проверяя их на допустимость, затем вычисляя выигрыш; в конце выбираются управления с наибольшим выигрышем.

Уравнение Беллмана

«Наивное» решение задачи ДП простым перебором является чрезвычайно трудоемким, поскольку количество перебираемых случаев управлений экспоненциально зависит от числа шагов ($|\{\overline{a}\}| \approx |\Gamma(x_t)|^n$). Кардинально сократить трудоемкость позволяет использование принципа оптимальности.

Принцип оптимальности Беллмана для дискретных систем. Оптимальное управление имеет свойство, что каковы бы ни были первоначальное состояние и управление, последующие действия должны составлять оптимальное управление относительно состояния, наступившего после первого действия.

Согласно принципу оптимальности, можно отделить первый шаг от последующих:

$$V(x_0) = \max_{a_0} \left\{ f(x_0, a_0) + \max_{a_1, \ldots, a_n} \left[\sum_{t=1}^{n} f(x_t, a_t) \right] \right\}, \;\; a_t \in \Gamma(x_t)$$

откуда видно, что второе слагаемое представляет собой оптимальный выигрыш $V(x_1)$, который можно получить, начиная с состояния x_1, таким образом:

$$V(x_0) = \max_{a_0} \{ f(x_0, a_0) + V(x_1) \}$$

Окончательно, после итерирования, приходим к ***уравнению Беллмана***.

Уравнение Беллмана:

$$V(x) = \max_{a \in \Gamma(x)} \{ f(x, a) + V(T(x, a)) \}.$$

Величины $V(x_t)$, $t = 1, \ldots, n$ называются ***условными оптимальным выигрышами***, в отличии от ***абсолютного оптимального выигрыша*** $V(x_0)$ – полного решения задачи, поскольку они дают оптимальное значение, как будто бы задача начиналась с состояния x_t (то есть при дополнительном ***условии***).

Замечание. При использовании уравнения Беллмана на каждом шаге t тоже проводится перебор: для каждого выбора $a \in \Gamma(x_t)$ высчитывается сумма $f(x_t, a) + V(T(x_t, a))$ (предполагается, что $V(T(x_t, a))$ уже известны), после чего берется максимум результатов по a.

Рассматривают две схемы реализации алгоритмов динамического программирования: ***восходящую*** (***снизу***) и ***нисходящую*** (***сверху***). При нисходящей схеме высчитываются все условные оптимальные стоимости, независимо от того понадобятся их значения при расчете абсолютной оптимальной стоимости (окончательного решения задачи) или нет. При восходящей схеме, вычисляются только те условные оптимальные стоимости, которые применяются при расчете окончательного решения. В общем случае восходящая схема эффективней нисходящей. Более подробно различия между восходящей и нисходящей схемой можно увидеть в последующих примерах.

Рассмотрим реализацию принципов динамического программирования на примере простой модельной задачи.

Пример. (Сад расходящихся тропок.) Пусть требуется в саду пройти из от входа A к выходу D (рис. 2.2) за кратчайшее время. Двигаться можно только вдоль тропинок (стрелок). Числа у стрелок показывают затраты на прохождение соответствующего участка в минутах.

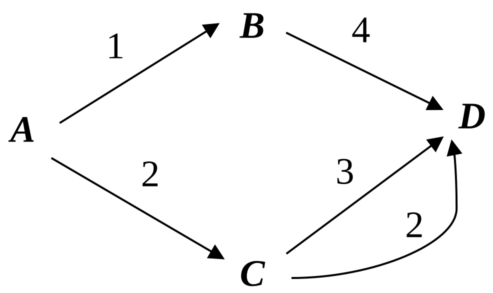


Рис. 2.2. План сада расходящихся тропок.

<u>Решение.</u> Если посмотреть на план, то сразу можно найти быстрейший путь. Но этот подход уже не работает, если дорожек десятки, а на плане современного города их могут быть сотни тысяч, а то и миллионы. Обоснуем формальный алгоритм нахождения кратчайшего пути.

Если смотреть от входа A, то более короткая тропинка лежит налево. Но она ведет в такой участок сада, откуда к выходу быстро попасть нельзя. Поступим по-другому, начнем с конца.

Находясь ровно за один переход от выхода (а это развилки B и C), мы можем рассчитать минимально необходимое время до выхода. Для B это четыре минуты, а для C – две ($2 = \min\{3; 2\}$). Обозначим, найденное минимальное время, за которое достижим выход за V. Другими словами, $V(B) = 4$, $V(C) = 2$.

Будем также обозначать расстояние между развилками, достижимыми с помощью прямых тропинок, с помощью функции f, таким образом, например, $f(A, B) = 1$, $f(A, C) = 2$, $f(C, B) = \infty$, $f(B, D) = 4$.

Находясь за два перехода от выхода, а в нашем маленьком саду это уже вход A, очевидно, нам нужно посчитать минимум величины

«время движения до следующей точки, плюс минимальное время прохождения из той следующей точки до выхода»,

то есть надо выбрать минимум из значений:

$$f(A, B) + V(B) = 1 + 4 = 5, \quad f(A, C) + V(C) = 2 + 2 = 4,$$

то есть 4. Это и есть минимальное время $V(A)$, за которое из A можно добраться до выхода. Сравните с формулой Беллмана, которая в наших обозначениях выглядит так:

$$V(A) = \min\{ f(A, X) + V(X)\},$$

для всех X, достижимых из A. □

Рассмотренный в примере подход, по сути совпадающий с подходом Беллмана, лежит в основе всех приложений[1] динамического программирования. В более общем случае, развилки соответствуют промежуточным состояниям управляемой системы, а «тропинки», исходящие из этой развилки, – вариантам управления, возможным в этих ситуациях. Числа (веса) при стрелках, соответственно, эта затраты, которые необходимо понести при выбранном управлении-тропинке. Мы увидим реализацию этой интерпретации в примере 2.4.

Контрольные вопросы

1. К задачам какого вида применим метод динамического программирования?
2. Чем характеризуется задачи с оптимальной подструктурой?
3. В чем заключается детерминированность рассмотренной постановки задачи динамического программирования?
4. Чем нисходящая схема решения задачи динамического программирования отличается от восходящей? Приведите примеры.
5. Сформулируйте принцип оптимальности для дискретных систем.
6. Запишите уравнение Беллмана.

[1] Правда, в самом общем случае надо учесть, что переходы могут иметь отрицательные веса. Это усложняет рассуждения, в частности, могут появиться циклы с отрицательными «затратами» на прохождения, и тогда задача на минимум вообще не имеет конечного решения.

Задачи динамического программирования

Рассмотрим применение динамического программирования для решения задач, поставленных ранее в примерах 2.1-2.3.

Алгоритм поиска кратчайшего пути

Рассматриваемый далее алгоритм имеет широчайшие практические приложения и помимо расчета кратчайшего пути. В зависимости от смысла, придаваемого весовым коэффициентам он может использовать для нахождения наискорейшего пути, пути с наименьшими затратами топлива или вообще каких-либо затрат для пожарной автоцистерны или транспортного конвоя, может рассчитывать план оптимальной прокладки электро-, энерго- и сигнальных сетей, маршрутизации сообщений и т.д. В частности, вариации этого алгоритма используются в навигаторах и навигационных программах.

Общее описание работы. В процессе выполнения этого алгоритма вершинам v графа приписываются временные пометки $l(v)$, каждая из которых даёт верхнюю границы для длины пути из s в эту вершину. Эти пометки постепенно уменьшаются с помощью некоторой итерационной процедуры, и на каждом шаге ровно одна пометка становится постоянной, что значит, что найдена длина минимального пути из s в эту вершину. Алгоритм останавливается, когда постоянной становится пометка терминальной вершины.

Алгоритм.

1. а) Положить $l(s) = 0$ и считать эту пометку постоянной.

   б) Для каждого $x_i \neq s$, полагаем $l(x_i) = \infty$ и считаем эту пометку временной.

   в) Положить текущую вершину $p = s$.

2. Для всех x достижимых из p, пометки которых временные, положить:

$$l(x) := \min\{l(x),\ l(p) + w(p, x)\}\ .$$

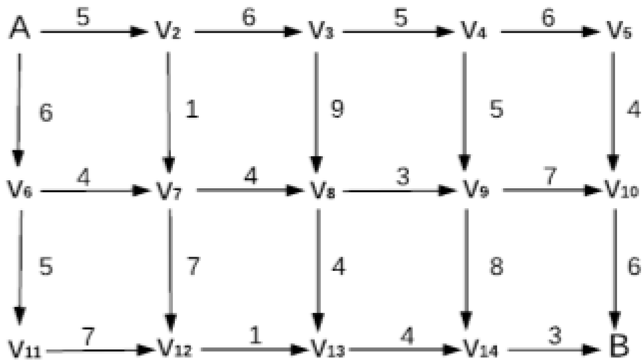
3. а) Среди всех вершин с временными пометками найти вершину x^*, пометка которой минимальна.

   б) Считать пометку этой вершины постоянной и положить $p = x^*$.

   в) Если $p = t$, то $l(p)$ является длиной кратчайшего пути, иначе перейти к пункту 2 алгоритма.

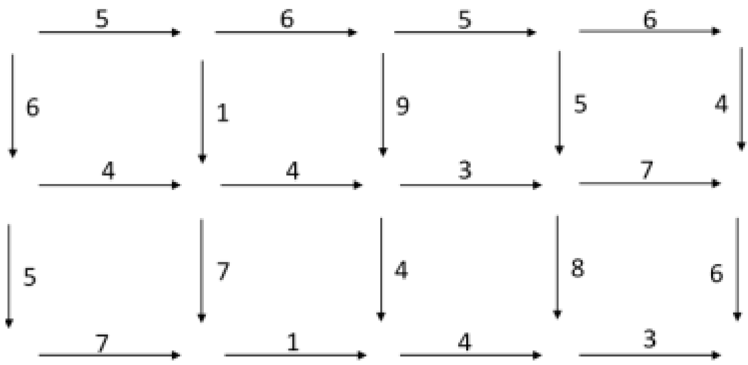
Пример 2.4 Пожарная автоцистерна движется по участку города со схемой движения, заданной таблицей 2.1. Найти кратчайший путь из вершины A в вершину B, при условии, что переход от вершины к вершине возможен только слева-направо и сверху-вниз:

Таблица 2.1



Решение. Решение будем искать согласно алгоритму нахождения кратчайшего пути. Ход решения будем записывать в следующую вспомогательную таблицу 2.2, которая получена из таблицы 1 условия, заменой обозначений вершин пробелами:

Таблица 2.2



Чтобы узнать номер вершины, стоящей в заданной позиции таблицы 2.2, нужно обратиться к исходной таблице 2.1. Хотя для записи решения задачи достаточно наличия только одного экземпляра таблицы 2.2, в целях наглядности будут приведены несколько частично заполненных вариантов таблицы 2.2 для последовательных шагов алгоритма.

Будем использовать следующие соглашения и обозначения:

а) значения временных пометок вершин будут записываться числами на месте соответствующих вершин, постоянные пометки будут выделяться жирным шрифтом;

б) веса дуг графа (длины соответствующих путей) записываем наклонным шрифтом;

в) вместо бесконечностей оставляем пустые места;

г) положение текущей вершины p обозначаем подчеркиванием соответствующей пометки.

Начало исполнения алгоритма:

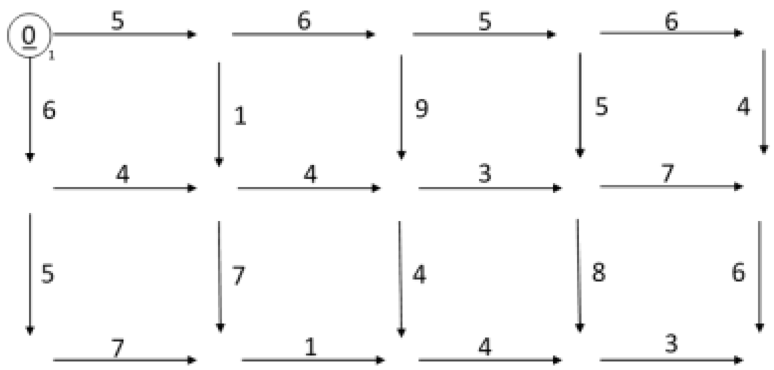
1. а) Ставим пометку для вершины s и делаем её постоянной.

б) Вместо бесконечностей для всех остальных вершин оставляем пустые места.

в) Полагаем $p = s$.

Записываем результаты этих шагов в таблицу 2.3:

Таблица 2.3

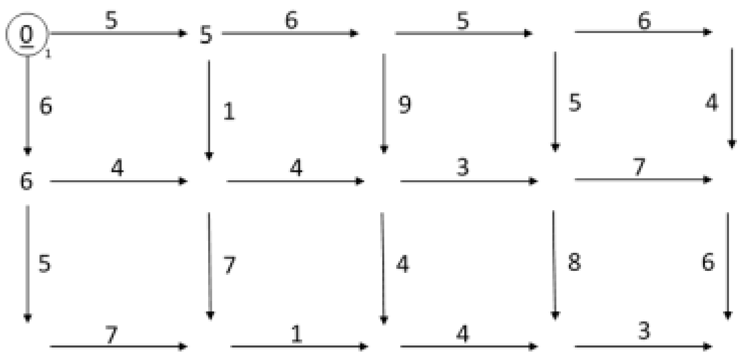


2. Достижимыми из $p = s$ являются вершины v_2 и v_6. Вычисляем для них новые значения временных пометок:

$$l(v_2) = \min\{\infty, 0+5\} = 5, \quad l(v_6) = \min\{\infty, 0+6\} = 6$$

Подставляем эти значения в таблицу (табл. 2.4):

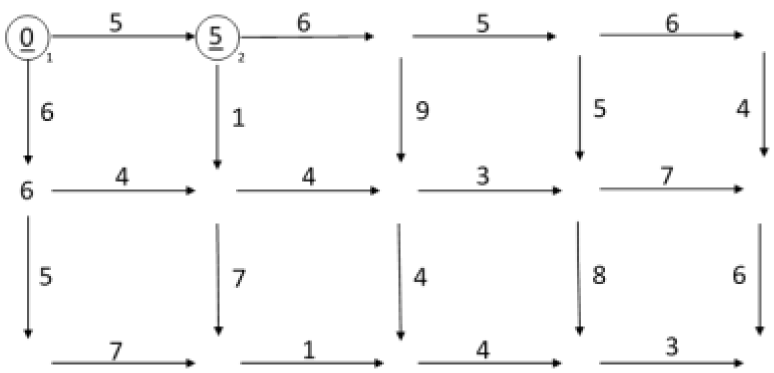
Таблица 2.4



3. а) Среди вершин с временными пометками (а это все вершины, кроме тех где стоит число жирным шрифтом) находим вершину с минимальной пометкой, это вершина $x^* = v_2$, где $l(v_2) = 5$.

б) Полагаем $p = x^* = v_2$ и делаем временную пометку $l(v_2) = 5$ постоянной (табл. 2.5):

Таблица 2.5



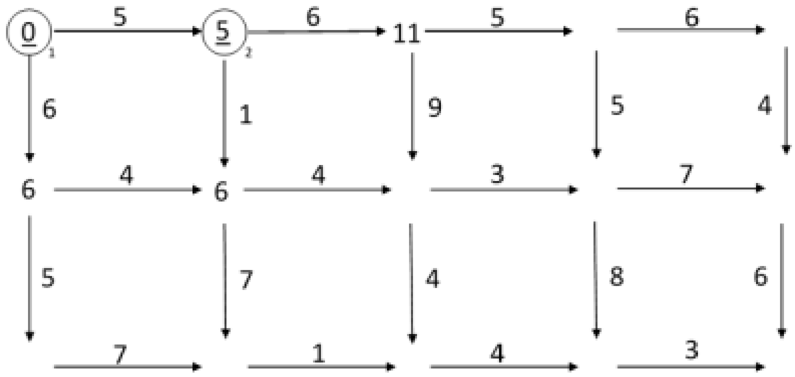
в) Так как $p = v_2 \neq t$, переходим к пункту 2 алгоритма.

2. Достижимыми из $p = v_2$ являются вершины v_3 и v_7. Вычисляем для них новые значения временных пометок:

$$l(v_3) = \min\{\infty, 5+6\} = 11, \quad l(v_7) = \min\{\infty, 5+1\} = 6,$$

поскольку постоянная пометка в v_2 равна 5.
Подставляем эти значения в таблицу 2.6:

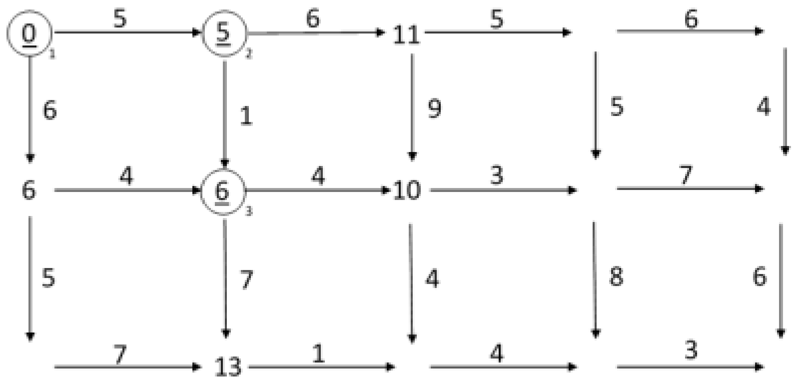
Таблица 2.6



3. а) Опять просматриваем все вершины с временными пометками и находим вершину x^* с минимальной пометкой, таких вершин две: v_6 и v_7, где $l(v) = 6$.

б) Полагаем $p = v_7$ (можно выбрать любую из двух найденных вершин, но мы будем выбирать ту, которая ближе к t) и делаем временную пометку $l(v_7) = 6$ постоянной (табл. 2.7):

Таблица 2.7

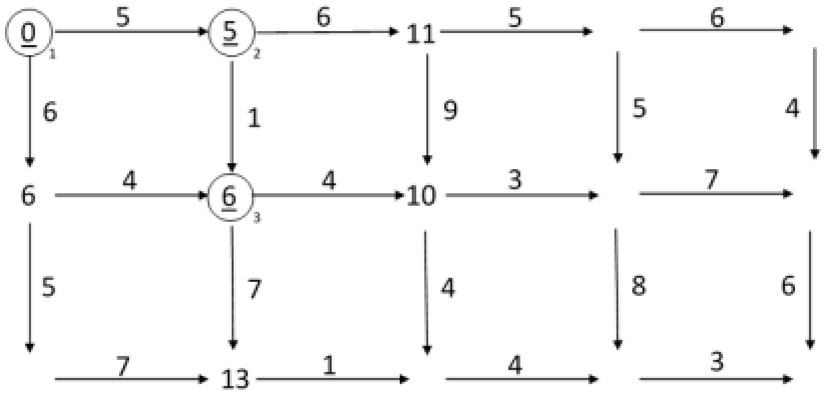


в) Так как $p = v_7 \neq t$, переходим к пункту 2 алгоритма.
2. Достижимыми из $p = v_7$ являются вершины v_8 и v_{12}. Вычисляем для них новые значения временных пометок:

$$l(v_8) = \min\{\infty, 6+4\} = 10, \quad l(v_{12}) = \min\{\infty, 6+7\} = 13,$$

поскольку $l(v_7) = 6$.
Подставляем эти значения в таблицу (табл. 2.8):

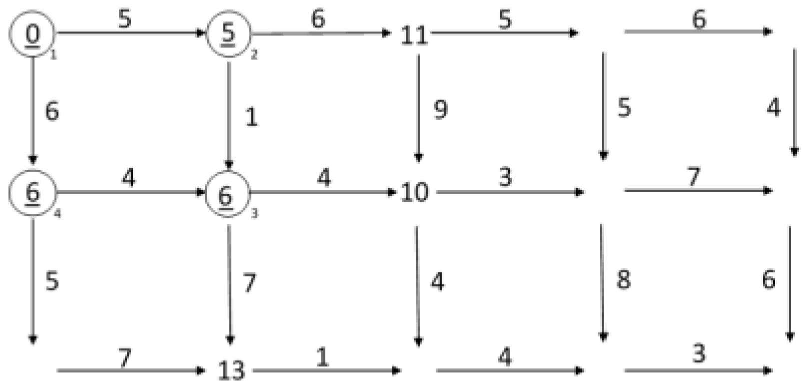
Таблица 2.8



3. а) Опять просматриваем все вершины с временными пометками и находим вершину x^* с минимальной пометкой, эта вершина v_6, где $l(v_6) = 6$.

б) Полагаем $p = v_6$ и делаем временную пометку $l(v_6) = 6$ постоянной (табл. 2.9):

Таблица 2.9



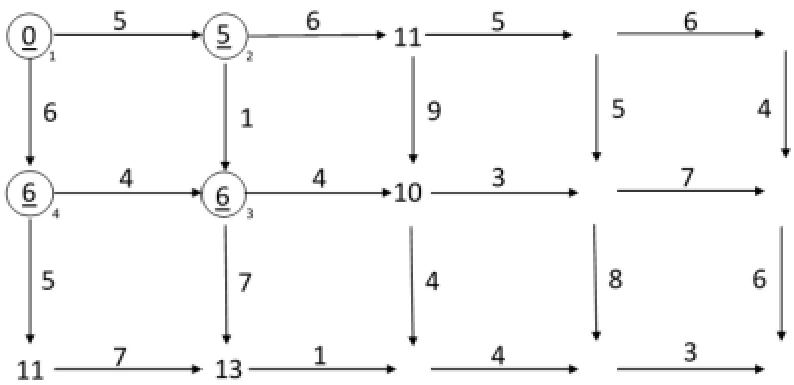
в) Так как $p = v_6 \neq t$, переходим к пункту 2 алгоритма.

2. Достижимыми из $p = v_6$ являются вершины v_{11} и v_7. Но вершина v_7 уже получила постоянную пометку, которая не пересчитывается. Вычисляем для v_{11} новое значение временной пометки:

$$l(v_{11}) = \min\{\infty, 6 + 5\} = 11.$$

Подставляем это значение в таблицу (табл. 2.10):

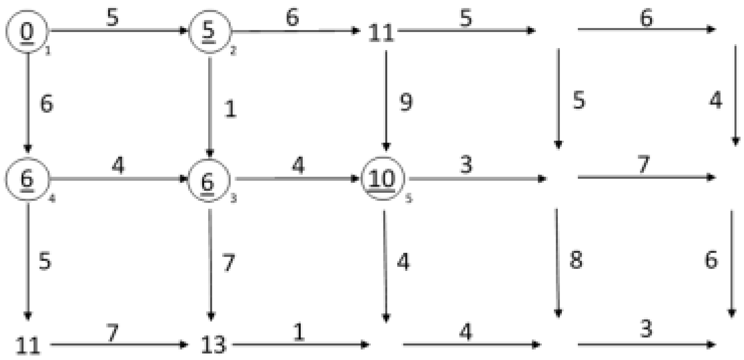
Таблица 2.10



3. а) Опять просматриваем все вершины с временными пометками (11, 13, 10, 11 и бесконечности) и находим вершину x с минимальной пометкой, эта вершина v_8 с $l(v_8) = 10$.

б) Полагаем $p = v_8$ и делаем временную пометку $l(v_8) = 10$ постоянной (табл. 2.11):

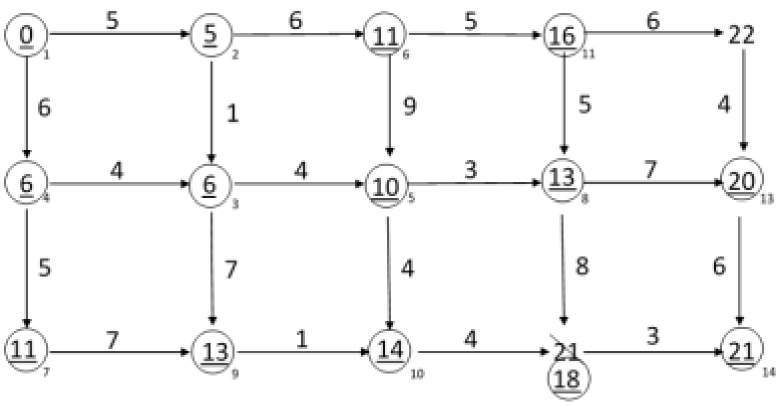
Таблица 2.11



в) Так как $p = v_8 \neq t$, переходим к пункту 2 алгоритма.

Продолжая выполнение алгоритма, окончательно приходим к таблице 2.12:

Таблица 2.12



На этом шаге $p = t = B$ и значит, по пункту 3. в), алгоритм завершается.

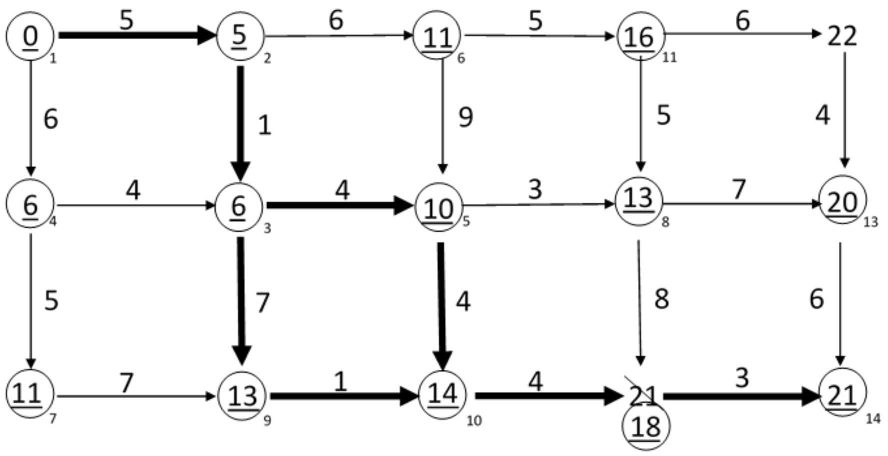
Заметим, что в вершине v_{14} стоит зачеркнутая временная пометка ~~21~~. Это значит, что временная пометка пересчитывалась (дважды: сначала с ∞ на 21, потом с 21 на 18). Заметим также, что в вершине v_5 временная пометка 22 так и не стала постоянной.

Осталось выписать решение. Значение $l(B) = 21$ дает длину наименьшего пути. Чтобы определить этот путь, действуют следующим образом: начиная с вершины B двигаются по графу против «движения», переходя к вершинам пометки которых отличаются от пометки текущей вершины в точности на вес дуги соединяющих эти вершины. Так в вершину B можно прийти из вершин v_{10} и v_{14}, но только для v_{14} имеем, что

$$l(B) - l(v_{14}) = 21 - 18 = 3 = w(v_{14}, B),$$

значит оптимальный путь проходил через v_{14}. Далее делаем шаг назад из v_{14} и так далее. В итоге приходим к следующему результату (табл. 2.13):

Таблица 2.13



Веса дуг, входящих в оптимальный путь, выделены жирным наклонным шрифтом с подчеркиванием. Заметим, что найдено два оптимальных пути:

$$A, v_2, v_7, v_8, v_{13}, v_{14}, B \quad \text{и} \quad A, v_2, v_7, v_{12}, v_{13}, v_{14}, B.$$

<u>Ответ:</u> Имеется два оптимальных пути:

$$A, v_2, v_7, v_8, v_{13}, v_{14}, B \quad \text{и} \quad A, v_2, v_7, v_{12}, v_{13}, v_{14}, B$$

длиной 21.

Задача для самостоятельного решения. Найти кратчайший путь из вершины A в вершину H (рис. 2.3). (Решение оформлять строго согласно образцу.)

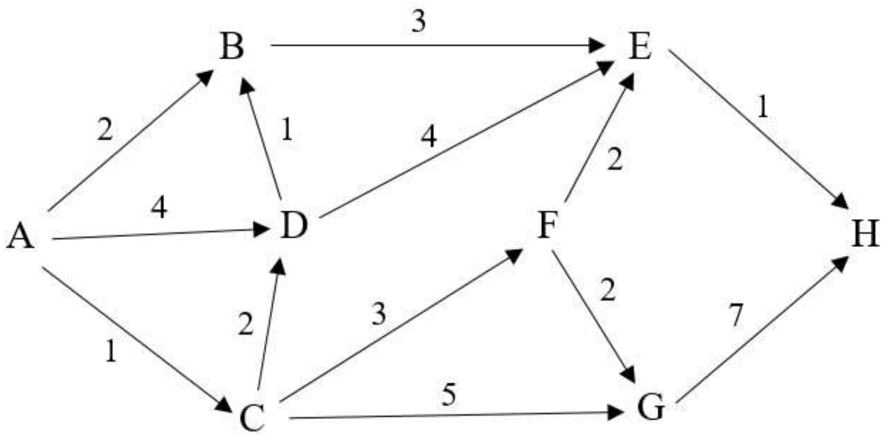


Рис. 2.3. Граф дорожной сети.

Задача о загрузке транспортного средства

Рассмотрим задачу о загрузке транспортного средства, также известную как «задача о загрузке корабля», «задача о рюкзаке» или «задача линейного раскроя».

В общей постановке задача формулируется следующим образом: пусть имеется транспортное средство грузоподъёмности W, которое может быть загружено предметами n типов. Если обозначить через w_k и c_k вес и ценность предмета k -го типа (от английских “weight” – вес и “cost” – цена), через x_k – количество таких предметов, то можно поставить задачу загрузки транспортного средства грузом максимальной ценности в виде:

$$\text{максимизировать} \quad \sum_{k=1}^{n} c_k \cdot x_k ; \tag{2.1}$$

$$\text{при условиях} \quad \sum_{k=1}^{n} w_k \cdot x_k,, \ W, \ x_k \geq 0 \text{ при всех } k. \tag{2.2}$$

Если потребовать неделимости предметов, т.е. целочисленности x_k, то возникшую задачу можно решать методами целочисленного линейного программирования, такими как, например, методом ветвей и границ или методом Гомори, получая решение для конкретного значения W. В этом случае, появление дополнительных условий обычно приводит к усложнению процесса решения. Применим вместо этого принцип динамического программирования.

Поскольку, в общем случае, безразлично, в каком порядке грузить предметы, то загрузку можно представить n-шаговым процессом, на первом шаге которого загружается x_n предметов n-го типа (на втором x_{n-1} – $n-1$-го, затем $n-2$ -го и на последнем – первого типа). Тогда, обозначив через $f_n(w)$ суммарную ценность загрузки в n-шаговом процессе при заданной грузоподъемности w и при использовании оптимальной загрузки, на основе принципа оптимальности сводим задачу к системе рекуррентных соотношений:

$$\begin{aligned} &f_k(w) = \max\{c_k \cdot x_k + f_{k-1}(w - w_k \cdot x_k)\}, \ k = 2,3,\ldots,n; \\ &f_1(w) = c_1 \cdot x_1, \end{aligned} \tag{2.3}$$

где область допустимых значений решения определяется целыми значениями x_k в диапазоне от нуля до целой части отношения w/w_k. Значения $f_k(w)$, как и прежде, называются ***условными оптимальными стоимостями***, ввиду того что они определяют оптимальную стоимость, как *если бы* нам пришлось производить загрузку только w тонн груза и только предметами первых k типов – то есть при выполнении некоторого дополнительного условия, отсюда *условные* в названии.

Пример 2.5 Найти наиболее выгодный вариант загрузки автомобиля максимальной грузоподъемностью W = 5 тонн, то есть определить максимальную стоимость, количество и номера загружаемых предметов.

Предмет (i)	Вес, тонн (w_i)	Стоимость, тыс. руб. (c_i)
1	3	25
2	2	30
3	1	15

Решение. В качестве предварительного (нулевого) этапа решения, необходимо провести анализ задачи, для этого следует попробовать подобрать несколько вариантов загрузки автомобиля с расчетом получаемой суммарной стоимости. Например, можно полностью загрузить автомобиль, выбрав по одной единице груза 1-го и 2-го типа (3 тонны + 2 тонны = 5 тонн), при этом суммарная стоимость груза будет

$$1 \cdot 25 + 1 \cdot 30 = 55 \text{ тыс. рублей.}$$

Однако, если загрузить автомобиль только предметами 3-типа, то суммарная стоимость груза будет больше:

$$5 \cdot 15 = 75 \text{ тыс. рублей.}$$

Полученные на нулевом этапе оценки стоимости загрузки, могут оказаться полезны для самопроверки, как в ходе решения задачи, так и для проверки полученного ответа.

Решение задачи сводится к заполнению следующей таблицы:

Таблица 2.14

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0						
1						
2						
3						
4						
5						

которую будем заполнять по два столбца одновременно.

1) Согласно второй формуле из (2.3)

$$f_1(w) = c_1 \cdot x_1(w),$$

заполнение двух первых свободных столбцов проводится достаточно просто: обратите внимание, что эти столбцы описывают заполнение веса w = 0,1,2,3,4,5 количеством $x_1(w)$ предметов 1-го типа, вес каждого из которых, согласно условию, равен w_1 = 3 тоннам, а цена c_1 = 25 тыс. рублей. Очевидно, что при w = 5, 4, 3, можно загрузить один и только один предмет 1-го типа, то есть $x_1(5) = x_1(4) = x_1(3) = 1$, а при w = 2, 1, 0 – ни одного предмета, то есть $x_1(2) = x_1(1) = x_1(0) = 0$, при соответствующей суммарной стоимости загруженных предметов. Таким образом, после первого шага (здесь сначала заполняем второй столбец, потом рассчитываем стоимость $f_1(w) = c_1 \cdot x_1(w) = 25 \cdot x_1(w)$ для заполнения первого столбца), получаем:

Таблица 2.15

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0				
2	0	0				
3	25	1				
4	25	1				
5	25	1				

2) Заполнение двух последующих незаполненных столбцов проводится с использованием первой формулы из (2.3):

$$f_k(w) = \max\{c_k \cdot x_k + f_{k-1}(w - w_k \cdot x_k)\} \qquad (2.4)$$

Заполнение столбцов будем проводить с конца, причем для каждого значения w = 5,4,3,2,1,0; сначала находится значение $f_2(w)$, а в процессе его нахождения, определяется значение $x_2(w)$ (или значения – их может быть несколько). Во всех вычислениях этого шага, значение c_2 определяется из условия задачи и равно 30 тыс. рублей.

- $\underline{f_2(5)}$ Поскольку w = 5 тонн, а вес предметов 2-го типа равен w_2 = 2 тонны, то $x_2(3)$= 0, 1 или 2. Нам предстоит определить, какое именно из этих значений соответствует оптимальной загрузке предметами 1-го и 2-го типа. Согласно (4), оптимальная условная стоимость:

$$f_2(5) = \max\{c_2 \cdot 0 + f_1(5),\ c_2 \cdot 1 + f_1(5 - w_2 \cdot 1),\ c_2 \cdot 2 + f_1(5 - w_2 \cdot 2)\}$$

$$= \max\{c_2 \cdot 0 + f_1(5),\ c_2 \cdot 1 + f_1(5 - 2 \cdot 1),\ c_2 \cdot 2 + f_1(5 - 2 \cdot 2)\}$$

$$= \max\{30 \cdot 0 + f_1(5),\ 30 \cdot 1 + f_1(3),\ 30 \cdot 2 + f_1(1)\}.$$

Значения стоимостей $f_1(5)$, $f_1(3)$, $f_1(1)$ определяем из уже заполненной части таблицы:

$$= \max\{30 \cdot 0 + 25,\ 30 \cdot 1 + 25,\ 30 \cdot 2 + 0\}$$

$$= \max\{25, 55, 60\} = 60.$$

Отсюда, $x_2(5)$, поскольку именно при таком количестве предметов 2-типа достигается максимум оптимальной условной стоимости $f_2(5)$.

Внесем найденные значения в таблицу 2.16:

Таблица 2.16

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0				
2	0	0				
3	25	1				
4	25	1				
5	25	1	**60**	**2**		

- $f_2(4)$ Поскольку w = 4 тонн, а вес предметов 2-го типа равен $w_2 = 2$ тонны, то $x_2(4)$ равно опять 0, 1 или 2. Тогда:

$$f_2(4) = \max\{c_2 \cdot 0 + f_1(4),\ c_2 \cdot 1 + f_1(4 - w_2 \cdot 1),\ c_2 \cdot 2 + f_1(4 - w_2 \cdot 2)\}$$

$$= \max\{c_2 \cdot 0 + f_1(4),\ c_2 \cdot 1 + f_1(4 - 2 \cdot 1),\ c_2 \cdot 2 + f_1(4 - 2 \cdot 2)\}$$

$$= \max\{30 \cdot 0 + f_1(4),\ 30 \cdot 1 + f_1(2),\ 30 \cdot 2 + f_1(0)\}$$

Значения стоимостей $f_1(4)$, $f_1(2)$, $f_1(0)$ определяем из уже заполненной части таблицы:

$$= \max\{30 \cdot 0 + 25,\ 30 \cdot 1 + 0,\ 30 \cdot 2 + 0\}$$

$$= \max\{25,\ 30,\ 60\} = 60.$$

Отсюда, $x_2(4) = 2$. Внесем найденные значения в таблицу 2.17:

Таблица 2.17

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0				
2	0	0				
3	25	1				
4	25	1	**60**	**2**		
5	25	1	60	2		

- $f_2(3)$ Поскольку w = 3 тонн, а вес предметов 2-го типа равен w_2 = 2 тонны, то $x_2(3)$ = 0 или 1. Тогда:

$$f_2(3) = \max\{c_2 \cdot 0 + f_1(3),\ c_2 \cdot 1 + f_1(3 - w_2 \cdot 1)\}$$

$$= \max\{c_2 \cdot 0 + f_1(3),\ c_2 \cdot 1 + f_1(3 - 2 \cdot 1)\}$$

$$= \max\{30 \cdot 0 + f_1(3),\ 30 \cdot 1 + f_1(1)\}$$

Значения стоимостей $f_1(3) = 25$ и $f_1(1) = 0$ определяем из уже заполненной части таблицы:

$$= \max\{30 \cdot 0 + 25,\ 30 \cdot 1 + 0\}$$

$$= \max\{25,\ 30\} = 30.$$

Отсюда, $x_2(3) = 1$. Внесем найденные значения в таблицу 2.18:

Таблица 2.18

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0				
2	0	0				
3	25	1	**30**	**1**		
4	25	1	60	2		
5	25	1	60	2		

- $\underline{f_2(2)}$Тогда $x_2(2) = 0$ или 1.

$$f_2(2) = \max\{c_2 \cdot 0 + f_1(2),\ c_2 \cdot 1 + f_1(2 - w_2 \cdot 1)\}$$

$$= \max\{c_2 \cdot 0 + f_1(2),\ c_2 \cdot 1 + f_1(2 - 2 \cdot 1)\}$$

$$= \max\{30 \cdot 0 + f_1(2),\ 30 \cdot 1 + f_1(0)\}$$

Значения стоимостей $f_1(2) = 0$ и $f_1(0) = 0$ определяем из уже заполненной части таблицы:

$$= \max\{30 \cdot 0 + 0,\ 30 \cdot 1 + 0\}$$

$$= \max\{0,\ 30\} = 30.$$

Отсюда, $x_2(2) = 1$. Внесем найденные значения в таблицу 2.19:

Таблица 2.19

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0				
2	0	0	**30**	**1**		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2		

- $\underline{f_2(1)}$ Тогда $x_2(1) = 0$ и других вариантов нет. Значит это и есть количество предметов 2-го типа дающих максимум условной оптимальной стоимости. Осталось определить этот максимум:

$$f_2(1) = \max\{c_2 \cdot 0 + f_1(1)\}$$

$$= \max\{c_2 \cdot 0 + f_1(1)\}$$

$$= \max\{30 \cdot 0 + f_1(1)\}$$

Значение стоимости $f_1(1)$, как и прежде, определяем из уже заполненной части таблицы:

$$= \max\{30 \cdot 0 + 0\}$$

$$= \max\{0\} = 0.$$

Внесем найденные значения в таблицу 2.20:

Таблица 2.20

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0				
1	0	0	**0**	**0**		
2	0	0	30	1		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2		

- $f_2(0)$ Аналогично предыдущему пункту, $x_2(0) = 0$ и других вариантов нет. Значит это и есть количество предметов 2-го типа дающих максимум условной оптимальной стоимости. Осталось определить этот максимум:

$$f_2(0) = \max\{c_2 \cdot 0 + f_1(0)\} = \max\{30 \cdot 0 + 0\} = 0.$$

Внесем найденные значения в таблицу 2.21:

Таблица 2.21

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0	**0**	**0**		
1	0	0	0	0		
2	0	0	30	1		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2		

3) Поскольку $f_3(5) = f_3(W)$ определяет оптимальную загрузку всего автомобиля, то в двух последних столбцах достаточно заполнить только нижнюю строку. Заполнение производится опять по формуле (4):

$$f_k(w) = \max\{c_k \cdot x_k + f_{k-1}(w - w_k \cdot x_k)\}.$$

- $f_3(5)$ Поскольку $w = 5$ тонн, а вес предметов 3-го типа равен $w_3 = 1$ тонны, то $x_3(5) = 0, 1, 2, 3, 4$ или 5. Согласно (4), оптимальная условная стоимость:

$f_3(5) = \max\{ c_3 \cdot 0 + f_2(5), c_3 \cdot 1 + f_2(5 - w_3 \cdot 1), c_3 \cdot 2 + f_2(5 - w_3 \cdot 2),$

$$c_3 \cdot 3 + f_2(5 - w_3 \cdot 3),\ \ c_3 \cdot 4 + f_2(5 - w_3 \cdot 4), c_3 \cdot 5 + f_2(5 - w_3 \cdot 5)\}$$

$$= \max\{ c_3 \cdot 0 + f_2(5), c_3 \cdot 1 + f_2(5 - 1 \cdot 1), c_3 \cdot 2 + f_2(5 - 1 \cdot 2), c_3 \cdot 3 + f_2(5 - 1 \cdot 3),$$

$$c_3 \cdot 4 + f_2(5 - 1 \cdot 4), c_3 \cdot 5 + f_2(5 - 1 \cdot 5)\}$$

$$= \max\{15 \cdot 0 + f_2(5), 15 \cdot 1 + f_2(4), 15 \cdot 2 + f_2(3), 15 \cdot 3 + f_2(2),$$

$$15 \cdot 4 + f_2(1), 15 \cdot 5 + f_2(0)\}$$

$$= \max\{15 \cdot 0 + 60, 15 \cdot 1 + 60, 15 \cdot 2 + 30, 15 \cdot 3 + 30, 15 \cdot 4 + 0, 15 \cdot 5 + 0\}$$

$$= \max\{60, 75, 60, 75, 60, 75\}.$$

Отсюда, $x_3(5)$ =1, 3 или 5, поскольку при всех трех значениях достигается максимум условной оптимальной стоимости. Внесем найденные значения в таблицу 2.22:

Таблица 2.22

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	0	0	0	0		
1	0	0	0	0		
2	0	0	30	1		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2	**75**	**1;3;5**

4) Осталось определить ответ по Таблице 2.22.

Максимальная стоимость груза равна 75 тыс. рублей и достигается при загрузке 1-ого, 3-х или 5-ти предметов 3-го типа.

а) Если загружен 1 предмет 3-го типа, то на предметы 1-го и 2-го типа осталось 4 тонны вместимости и следующее значение надо смотреть в ячейке для $x_2(4)$, рядом с $f_2(4)$ (Таблица 2.23). Таким образом, оптимальная по стоимости загрузка достигается при 2-х предметах 2-го типа и одном предмете 3-го типа, суммарно весящих $2 \cdot 2 + 1 \cdot 1 = 5$ тонн. Для предметов 1-типа места не остается.

Таблица 2.23

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	**0**	**0**	0	0		
1	0	0	0	0		
2	0	0	30	1		
3	25	1	30	1		
4	25	1	**60**	**2**		
5	25	1	60	2	**75**	**1**

Первый ответ: $x_1 = 0$, $x_2 = 2$, $x_3 = 1$.

б) Если загружены 3 предмета 3-го типа, то на предметы 1-го и 2-го типа осталось 2 тонны вместимости и следующее значение надо смотреть в ячейке для $x_2(2)$, рядом с $f_2(2)$. Таким образом, оптимальная по стоимости загрузка достигается при 1-ом предмете 2-го типа и 3-х предмета 3-го типа, суммарно весящих 2·1+3·1=5 тонн. Для предметов 1-типа места также не остается.

Таблица 2.24

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	**0**	**0**	0	0		
1	0	0	0	0		
2	0	0	**30**	**1**		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2	**75**	**3**

Второй ответ: $x_1 = 0$, $x_2 = 1$, $x_3 = 3$.

в) Если загружены 5 предметов 3-го типа, то на предметы 1-го и 2-го типа места не остается.

Таблица 2.25

w	$f_1(w)$	$x_1(w)$	$f_2(w)$	$x_2(w)$	$f_3(w)$	$x_3(w)$
0	**0**	**0**	**0**	**0**		
1	0	0	0	0		
2	0	0	30	1		
3	25	1	30	1		
4	25	1	60	2		
5	25	1	60	2	**75**	**5**

Третий ответ: $x_1 = 0$, $x_2 = 0$, $x_3 = 5$.

Проверка: Для проверки посчитаем стоимости для каждой найденной загрузки:

первая: $c = 0 \cdot 25 + 2 \cdot 30 + 1 \cdot 15 = 75$ тыс. рублей,

вторая: $c = 0 \cdot 25 + 1 \cdot 30 + 3 \cdot 15 = 75$ тыс. рублей,

третья: $c = 0 \cdot 25 + 0 \cdot 30 + 7 \cdot 15 = 75$ тыс. рублей,

как и ожидалось. Заметьте также, что при предварительном анализе задачи, нам не встречались загрузки с большей стоимостью.

Ответ: При управлениях (0,2,1), (0,1,3), (0,0,5) достигается максимальная стоимость загрузки в 75 тыс. рублей.

Задача распределения ресурсов

Пример 2.6 Решить следующую задачу: городское управление получило N = 4 пожарных автоцистерны, которые следует распределить между 3 пожарными частями. Каждая из пожарных частей F_i, i = 1,2,3, при поступлении в неё n автоцистерн повышает уровень технической готовности машинного парка, зависящий от n, т.е. представляющий собой какую-то функцию $f_i(n)$. Все функции $f_i(n)$, i = 1,2,3, n = 1,2,3,4, заданы. Как следует распределить закупленную технику, чтобы в сумме это давало максимальное значение технической готовности по управлению?

Таблица 2.26

n	$f_1(n)$	$f_2(n)$	$f_3(n)$
0	0,94	0,93	0,92
1	0,95	0,93	0,93
2	0,96	0,94	0,94
3	0,96	0,94	0,95
4	0,97	0,96	0,95

Решение. Используем принцип оптимальности. В приложении к этой задаче, он приводит с следующей системе формул:

$$u_1(n) = f_1(n), \quad n = N, \ldots, 0; \tag{1}$$

$$u_k(n) = \max\{f_k(j) + u_{k-1}(n-j)\}_{j=0,\ldots,n}, \ k = 2, \ldots, N; \ \ n = N, \ldots, 0; \tag{2}$$

где $u_k(n)$ – условные оптимальные выигрыши.

Решение задачи сводится к заполнению следующей таблицы (табл. 2.27), которую будем заполнять, как и раньше, по два столбца одновременно:

Таблица 2.27

n	$u_1(n)$	$x_1(n)$	$u_2(n)$	$x_2(n)$	$u_3(n)$	$x_3(n)$
0						
1						
2						
3						
4						

$\underline{u_1(4)}$, $\underline{u_1(3)}$, $\underline{u_1(2)}$, $\underline{u_1(1)}$, $\underline{u_1(0)}$

Поскольку в данном случае условные оптимальные выигрыши $u_i(n)$ рассчитываются исходя из того, что автоцистерны направляются только в первую пожарную часть в количестве n = 4,3,2,1,0, соответственно, то значения $u_1(n)$ = $f_1(n)$, как и следует из формулы (1).

$$\underline{u_2(4)} = \max\{ f_2(0)+u_1(4),\ \ f_2(1)+u_1(3),\ \ f_2(2)+u_1(2),\ \ f_2(3)+u_1(1),\ \ f_2(4)+u_1(0)\}$$
$$= \max\{0,93+0,97,\ \ 0,93+0,96,\ 0,94+0,96,\ \ 0,94+0,95,\ 0,96+0,94\}$$
$$= \max\{1,90;\ \ 1,89;\ \ 1,90;\ \ 1,89;\ \ 1,90\}$$
$=1,90$, при $x_2(4)$ = 0; 2; 4.

$$\underline{u_2(3)} = \max\{ f_2(0)+u_1(3),\ \ f_2(1)+u_1(2),\ \ f_2(2)+u_1(1),\ \ f_2(3)+u_1(0)\}$$
$$= \max\{0,93+0,96,\ \ 0,93+0,96,\ 0,94+0,95,\ \ 0,94+0,94\}$$
$$= \max\{1,89;\ 1,89;\ 1,89;\ 1,88\}$$
$=1,89$, при $x_2(3)$ = 0; 1; 2.

$$\underline{u_2(2)} = \max\{ f_2(0)+u_1(2),\ \ f_2(1)+u_1(1),\ \ f_2(2)+u_1(0)\}$$
$$= \max\{0,93+0,96,\ \ 0,93+0,95,\ 0,94+0,94\}$$
$$= \max\{1,89;\ 1,88;\ \ 1,88\}$$
$=1,89$, при $x_2(2)$ = 0.

$$\underline{u_2(1)} = \max\{ f_2(0)+u_1(1),\ \ f_2(1)+u_1(0)\}$$
$$= \max\{0,93+0,95,\ \ 0,93+0,94\}$$
$$= \max\{1,88;\ 1,87\}$$
$=1,88$, при $x_2(1)$ = 0.

$$\underline{u_2(0)} = \max\{ f_2(0)+u_1(0)\}$$
$$= \max\{0,94+0,93\}$$
$$= \max\{1,87\}$$
$=1,87$, при $x_2(0)=0$.

$$\underline{u_3(4)} = \max\{ f_3(0)+u_2(4),\ \ f_3(1)+u_2(3),\ \ f_3(2)+u_2(2),\ \ f_3(3)+u_2(1),\ \ f_3(4)+u_2(0)\}$$
$$= \max\{0,92+1,90,\ \ 0,93+1,89,\ 0,94+1,89,\ \ 0,95+1,88,\ 0,95+1,87\}$$
$$= \max\{2,82;\ \ 2,82;\ \ 2,83;\ \ 2,83;\ \ 2,82\}$$
$=2,83$, при $x_3(4)$ = 2; 3.

В результате заполнения получаем таблицу 2.28:

Таблица 2.28

n	$u_1(n)$	$x_1(n)$	$u_2(n)$	$x_2(n)$	$u_3(n)$	$x_3(n)$
0	0,94	0	1,87	0		
1	0,95	1	1,88	0		
2	0,96	2	1,88	0		
3	0,96	3	1,89	0;1;2		
4	0,97	4	1,90	0;2;4	**2,83**	**2;3**

Из таблицы 2.28 следует, что для достижения максимальной технической готовности 2,83 (значение $u_3(4)$) необходимо отправить 2 или 3 пожарных автоцистерны в третью часть (значение $x_3(4)$).

Разберём оба случая.

Если отправить 2 автоцистерны, то на 1-ую и 2-ую пожарную части останется 2 автоцистерны, и тогда максимальная техническая готовность по двум первым пожарным частям 1,88 ($u_2(2)$) достигается при $x_2(2) = 0$, то есть при отправке во вторую часть 0 автоцистерн. Таким образом, на 1-ую часть остается 2 автоцистерны.

Если в третью часть отправить 3 автоцистерны, то на 1-ую и 2-ую пожарную части останется 1 автоцистерна, и тогда максимальная техническая готовность по двум первым пожарным частям 1,88 ($u_2(1)$) достигается при $x_2(1) = 0$, то есть при отправке во вторую часть 0 автоцистерн. Таким образом, на 1-ую часть остается 1 автоцистерна.

Ответ: наибольшая суммарная техническая готовность по управлению равна 2,83 и достигается при двух следующих распределениях пожарных автоцистерн:

$$x_1 = 2,\ x_2 = 0,\ x_3 = 2 \quad \text{или} \quad x_1 = 1,\ x_2 = 0,\ x_3 = 3.$$

Дискретизация задач оптимального управления

Существует огромный класс задач, связанных с оптимальным управлением системами с непрерывными состояниями – ***задачи оптимального управления***. К таким системам относятся системы водо- и теплоснабжения, газопроводы, атомные и химические реакторы, объекты механотроники и робототехнические комплексы, и многие другие. Как правило, такие системы описываются с помощью дифференциальных уравнений, описывающих подлежащие физические процессы функционирования. Сложность получения точных аналитических (формульных) решений для таких систем, а также удобство применения и мощность современных цифровых компьютеров, мотивируют переход к рассмотрению дискретных моделей непрерывных систем – дискретизации. Это дает широчайший спектр применений динамического программирования. Тем не менее переход к дискретной модели управления помимо преимуществ, имеет и определенные недостатки, которые мы обсудим в этом вопросе.

Рассмотрим задачу оптимального быстродействия, как пример задачи оптимального управления, подходящей для дискретизации.

Пример.

$$\begin{cases} T \to \inf, \\ \dot{\bar{x}}(t) = \bar{f}(\bar{x}, u), \quad \bar{x} \in X, \quad u \in U \\ \bar{x}(0) = \bar{x}^0, \quad \bar{x}(T) = \bar{0}, \end{cases}$$

где $\bar{x} = (x_1, x_2)$ – вектор фазовых переменных положения x_1 и скорости x_2 материальной точки.

В частности, в этой задаче, в отличии от простейшей, могут использоваться более общие уравнения движения

$$\begin{cases} \dot{x}_1 = f_1(\bar{x}, u), \\ \dot{x}_2 = f_2(\bar{x}, u). \end{cases}$$

Предположим, что

1) для любой начальной точки $\bar{x}^0 \in X$ существует оптимальная траектория, приводящая в точку (0,0); обозначим через $T(\bar{x}^0)$ – полное время движения по этой траектории;

2) $T(x) \in C^1(\mathbb{R}^2 \setminus \{(0,0)\})$.

Несмотря на то, что уравнение Беллмана имеет более узкие условия применимости, чем принципы, используемые в теории оптимального управления, как уже говорилось, динамическое программирование имеет преимущества, когда точное аналитическое решение задачи управления трудно или невозможно найти. Помимо этого, динамическое программирование, как мы увидим, может дать ***синтез оптимального управления***, а не просто ***программное управления*** для конкретного начального условия. Недостатком же является огрубление модели задачи, со всеми потенциально вытекающими проблемами: неверифицируемость полученной модели, неустойчивость по начальным данным и так далее. Эти проблемы характерны для любых моделей. Например, типичной задачей оптимального управления можно считать задачу автоматической стыковки корабля снабжения с космической станцией. Здесь оптимальность по времени не является существенной, основным является требование минимизации расхода топлива. Решение этой задачи не обязано приводить в *точку* фазового пространства, поскольку допустимы вариации по конечной скорости, скажем с заданной погрешностью ± 0,15 *м/с*. Допущение, что двигатели коррекции могут выдать сколь угодно малую тягу, приводит к тому, что управляемый объект, пытаясь удовлетворить слишком жесткие ограничения, быстро растрачивает топливо, за счет постоянно включаемых двигателей. Для модели динамического программирования задачи оптимального управления те же проблемы умножаются тем, что встречаются дважды, поскольку это модель второго уровня, то есть неточная модель от неточной модели.

Предположим, что существует оптимальное управление ($\hat{x}(t)$, $\hat{u}(t)$) для нестационарной задачи оптимального управления:

$$\begin{cases} J(x,u) = \int_{t_0}^{T} F(t, x(t), u(t))\,dt \to \text{extr}, \\ \dot{x} = f(t, x, u), \\ x(t_0) = x_0, \quad u \in U. \end{cases}$$

Пусть $\tau \in [t_0, T]$, $x = \hat{x}(\tau)$. Обозначим

$$\hat{J}(\tau, x) = \int_{\tau}^{T} f(t, \hat{x}(t), \hat{u}(t))\,dt\,.$$

Исследуем теперь поведение функции $\hat{J}(t, x)$ с течением времени вдоль оптимальной траектории $\hat{x}(t)$. Для этого рассмотрим малое приращение времени dt. За время dt, система перейдет в новое состояние

$$\hat{x}(t + dt) \approx \hat{x}(t) + d\hat{x}(t)$$

– ***приближенное вычисление с помощью дифференциала***.

Имеем

$$d\hat{x}(t) = f(t, \hat{x}(t), \hat{u}(t))\,dt,$$

поскольку

1. $d\hat{x}(t) = \dot{\hat{x}}(t)\,dt$ – ***дифференциал функции равен произведению её производной на дифференциал переменной***.

2. $\dot{\hat{x}}(t) = f(t, \hat{x}(t), \hat{u}(t))$ – получается из уравнений движения задачи ОУ.

Изменение значения функционала за время dt приближенно равно

$$d\hat{J}(t, x) \approx F(t, \hat{x}, \hat{u})\,dt\,,$$

(поскольку в общем случае $dF = f dx$, если F – первообразная для f; в нашем случае, J есть первообразная для F), а оставшаяся часть, согласно принципу оптимальности Беллмана, будет равна $\hat{J}(t + dt, \hat{x}(t + dt))$. Таким образом, получено следующее рекуррентное соотношение:

$$\hat{J}(t, \hat{x}(t)) \approx F\big(t, \hat{x}(t), \hat{u}(t)\big)dt + \hat{J}\big(t + dt, \hat{x}(t + dt)\big)\,.$$

Которое, пользуясь оптимальностью $\hat{u}(t)$, можно переписать как:

$$\hat{J}\big(t,\ x(t)\big) = \max_{u(t) \in U}\Big\{F\big(t, x(t), u(t)\big)dt + \hat{J}\big(t + dt, x(t + dt)\big)\Big\}\,.$$

В ***предположении дифференцируемости функции*** $\hat{J}(t, x)$ ***по всем своим аргументам***, переходя к пределу при $dt \to 0$ и учитывая уравнения движения задачи оптимального управления, получим ***уравнение Беллмана в дифференциальной форме***:

$$-\frac{d\hat{J}(t,\ x(t))}{dt}=\max_{u(t)\in U}\left\{F\left(t,x(t),u(t)\right)+\frac{d\hat{J}(t,x)}{dx}\,f\left(t,x(t),u(t)\right)\right\}$$

Это уравнение представляет собой дифференциальное уравнение ***в частных производных*** первого порядка для определения функции $\hat{J}(t, x)$ с нулевым краевым условием.

Как правило, аналитическое решение дифференциального уравнения Беллмана найти сложно или вовсе невозможно. Поэтому прибегают к дискретизации задачи с последующим ее численным решением. Дискретная задача формулируется следующим образом:

$$\begin{cases} J(x,u)=\sum_{i=0}^{N-1}F(t_i,x_i,u_i)\cdot\Delta t_i, \\ x_{i+1}=x_i+f(t_i,x_i,u_i)\Delta t_i, \\ x(t_0)=x_0,\quad u_i\in U_i. \end{cases}$$

Замечание. В дискретной задаче динамического программирования состояние системы задается вектором фазовых переменных $\bar{x}=(x_0, x_1, \dots, x_N)\in \mathbb{R}^{N+1}$, а управление – вектором управляющих переменных $\bar{u}=(u_0, u_1, \dots, u_N)\in \mathbb{R}^{N+1}$.

Уравнение Беллмана принимает следующий вид:

$$\hat{J}_i(x_i)=\max_{u_i\in U_i}\left\{F\left(t_i,x_i,u_i\right)\Delta t_i+\hat{J}_{i+1}(f(x_i,u_i))\right\}$$

с краевым условием $\hat{J}_N(x_N)=0$.

Решение задачи производится последовательным решением уравнения Беллмана для шагов $i=N-1,\ N-2,\ \dots,0$ (обратный (восходящий) ход метода Беллмана). При этом на каждом шаге получается оптимальное управление $\hat{u}_i$ как функция от текущего состояния системы x_i.

На втором этапе по полученным функциям $\hat{u}_i(t)$ производится синтез оптимального управления для задачи с конкретным начальным условием x_0.

Таким образом, метод динамического программирования, когда он применим, в отличие от решений, дающих оптимальное управление как функцию времени $\hat{u}(t)$ (***программное управление***), позволяет определять оптимальное управление как функцию состояния системы $\hat{u}(t, x)$ (***синтезированное управление***[2]), что дает возможность отыскивать решение сразу для целого класса задач с различными начальными условиями, со всеми вытекающими преимуществами, как, например, возможностью оптимальной коррекции фазовой траектории.

[2] Для получения синтезированного управления, нужно использовать прямой (нисходящий) ход (схему) метода Беллмана.

ГЛАВА 3. ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Задачи с линейными целевыми функциям и ограничениями

История линейного программирования

Основоположником линейного программирования (далее ЛП) считается советский ученый Л.В. Канторович (получил Нобелевскую премию по экономике в 1975 году «за вклад в теорию оптимального распределения ресурсов»). Консультируя в 1938 году фанерный тест по проблеме эффективного использования лущильных станков, он понял, что эта задача («задача о линейном раскрое»), как и многие другие в экономике, сводится к оптимизации линейной функции при наличии линейных ограничений-неравенств – ***линейному программированию***.

Более подробно рассмотрим другую знаменитую задачу линейного программирования, которая не требует целочисленности решения, – «*задачу о диете*».

Задача о диете. В 1945 году Джордж Стиглер (будущий Нобелевский лауреат по экономике 1982 г.) в статье «Стоимость проживания» поставил вопрос о минимальной стоимости армейского рациона:

Постановка задачи. Из списка 77 продуктов, таких как:

> *Хлеб пшеничный, ржаной, цельный; кукурузные мука, хлопья; рис, макароны, горох, бобы, …, свиная отбивная, говяжья печень, бифштекс.*

выбрать набор минимальной стоимости, удовлетворяющий физиологическим потребностям человека – содержанию 9 компонентов:

калорий, белка, кальция, железа, витаминов: A, B1, B2, B3, C.

Решение, полученное Стиглером, давало стоимость \$39.93/год или 10.87 центов/день, в ценах 1939 года и включало 5 продуктов – обогащенная пшеничная мука, сухое молоко, капуста, шпинат, флотские бобы.

Формализованная постановка задачи Стиглера

Обозначим ($i = 1, \ldots, 77$; $j = 1, \ldots, 9$):

x_i – количество i-го продукта в рационе (неизвестные);
c_i – стоимость i-го продукта;
a_{ji} – долевое содержание j-го компонента в i-ом продукте;
b_j – минимум дневной потребности в j-ом компоненте.

Тогда задача имеет вид:

$$
\begin{cases}
\sum_{i=1}^{77} c_i x_i \to \min \\
\sum_{i=1}^{77} a_{ji} x_i \ge b_j,\ j = 1,\dots,9; \\
x_i \ge 0,\ i = 1,\dots,77.
\end{cases}
$$

В 1947 году задача была пересчитана с использованием разработанного в том же году ***симплекс-метода***. Решение, найденное симплекс-методом, давало оптимальную стоимость \$39.67/год (в ценах 1939 года) и включало 9 продуктов:

пшеничная мука, кукурузная мука, сухое молоко, говяжья печень, лярд (смалец), арахисовое масло, капуста, картофель, шпинат.

На нахождение этого решения 9 клерков затратили 120 человеко-дней. В наше время персональный компьютер решает её за доли секунды. Указанное ускорение связано как с развитием вычислительной техники, так и улучшениями (на порядки) в алгоритме.

Контрольные вопросы

1. В чем состоит задача «о диете»?
2. Напишите формализованную постановку задачи «о диете».

Постановка задачи линейного программирования

Перейдем к рассмотрению общей задачи линейного программирования.

Задачей линейного программирования называется оптимальная задача, в которой и целевая функция, и ограничения задаются линейными функциями.

Общая постановка:

$$
\begin{cases}
\sum_{i=1}^{n} c_i x_i \to \max \\
\sum_{i=1}^{n} a_{ji} x_i \ [\le | \ge | =]\ b_j,\ j = 1,\dots,m.
\end{cases}
$$

где [≤|≥|=]означает одно из отношений ≤, ≥ или =, причем знаки могут разными для различных выражений $j = 1, \dots, m$.

Замечания.

1. При поиске минимума рассматривают целевую функцию

$$f(\overline{x}) = -\sum_{i=1}^{n} c_i x_i .$$

2. Ввиду того, что линейные равенства и неравенства задают выпуклые множества, а линейная функция является одновременно и выпуклой и вогнутой функцией, задача линейного программирования относится к более общей категории задач ***выпуклого программирования***.

3. Допустимое множество задачи линейного программирования имеет вид ***конечного*** (т.е. ограниченного) или ***бесконечного*** (т.е. неограниченного) n-мерного многогранника с внутренностью – ***полиэдра***.

4. Для задач линейного программирования существуют специализированные, более эффективные методы решения, чем, например, для более общих задач выпуклого программирования. Хотя полиэдр является **выпуклым множеством**.

Полезно помнить две следующих содержательных интерпретации для коэффициентов в задаче линейного программирования.

Возможный смысл коэффициентов

$$a_{ji},\ i=1,\ldots,n;\ j=1,\ldots,m:$$

- содержание j-го питательного вещества в i-ом продовольственном продукте (как в задаче о диете), причем ограничения накладывается на минимум содержания питательных веществ

$$a_{j1}x_1 + \ldots + a_{jn}x_n \geq b_j;$$

- затраты j-го материала (ресурса) при производстве x_i единиц i-ого продукта (в задаче о производстве), причем возможности производства ограничены запасами этих ресурсов

$$a_{j1}x_1 + \ldots + a_{jn}x_n \leq b_j.$$

Для классификации задач линейного программирования удобно использовать матричную запись.

В развернутом виде запись задачи линейного программирования выглядит следующем виде:

$$\begin{cases} c_1x_1 + \ldots + c_nx_n \to \max \\ a_{11}x_1 + \ldots + a_{1n}x_n\ [\leq | \geq | =]\ b_1, \\ \qquad\qquad \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n\ [\leq | \geq | =]\ b_m. \end{cases}$$

Соответственно, в более компактном, матричном виде задача записывается так:

$$\begin{cases} \overline{c} \cdot \overline{x}^T \to \max, \\ A\overline{x}^T \ \overline{*}^T \ \overline{b}^T, \end{cases}$$

где $\overline{c} = (c_1, \ldots, c_n)$, $\overline{x} = (x_1, \ldots, x_n)$ – вектора длины n;

$\overline{b} = (b_1, \ldots, b_m)$ – вектор длины m;

$A = (a_{ji})_{m\times n}$ – матрица размерности $m\times n$;

$\overline{*} = (*_1, \ldots, *_m)$ – вектор, для которого $\overline{*}_j \in \{\le, \ge, =\}$;

операция $(\ldots)^T$ означает транспонирование вектора.

Одну из основных особенностей задачи линейного программирования, на которой будут базироваться рассматриваемые далее методы, дает следующая теорема.

Теорема. Экстремум задачи линейного программирования, если он существует, достигается в ***угловой точке*** допустимого множества.

Замечание. Теорема ничего не говорит об единственности, и, в общем случае, экстремум может достигаться не только в вершине, но и на ребре допустимого n-мерного многогранника, грани и так далее.

Контрольные вопросы

1. Какая задача называется задачей линейного программирования?
2. Что такое угловая, граничная точка выпуклого множества?
3. Что представляет из себя допустимое множество задачи ЛП?
4. Что такое неограниченный многогранник? Приведите пример.
5. Где достигается экстремум задачи ЛП, если он существует?
6. Может ли экстремум ЗЛП достигаться более чем в одной точке?

Различные постановки задачи линейного программирования

В зависимости от характера линейных ограничений различают следующие 4 вида постановки задачи ЛП:

$$\overline{c} \cdot \overline{x}^T \to \max .$$

Общая: $A\overline{x}^T \ \overline{*}^T \ \overline{b}^T$.

Основная: $A\overline{x}^T \ \overline{*}^T \ \overline{b}^T, \ \ \overline{x} \ge 0$.

Стандартная: $A\overline{x}^T \le \overline{b}^T, \ \ \overline{x} \ge 0$.

Каноническая: $A\overline{x}^T = \overline{b}^T, \ \ \overline{x} \ge 0$.

Здесь $\bar{*}$ обозначает вектор с компонентами $\bar{*}_j \in \{\le, \ge, =\}$.

Пример. Приведем примеры различных постановок задачи ЛП.

1. ***Общая постановка***:

$$\begin{cases} 2x - 3y + z \to \max \\ 2x - y - 4z \le 6 \\ 3x + y - 3z = 14 \\ x + 3y - 2z \le -3 \end{cases}$$

Обратите внимание, что в общей постановке в ограничениях могут присутствовать, как оба типа нестрогих неравенств: «больше, либо равно» и «меньше, либо равно», так и равенства, кроме того отсутствует требование неотрицательности переменных.

2. ***Основная постановка*** – как общая, но добавляется требование неотрицательности переменных (последняя строка):

$$\begin{cases} 4x - 2y + 3z \to \max \\ -x - 2y + 4z \le 8 \\ 3x - 2y - 6z = 12 \\ x + 3y - 2z \ge -5 \\ x, y, z \ge 0 \end{cases}$$

3. ***Стандартная постановка*** **–** как основная, но неравенства только вида «меньше, либо равно»:

$$\begin{cases} 4x - 2y + 3z \to \max \\ -x - 2y + 4z \le 8 \\ 3x - 2y - 6z \le 12 \\ x + 3y - 2z \le -5 \\ x, y, z \ge 0 \end{cases}$$

4. ***Каноническая постановка*** **–** как стандартная, но вместо неравенств только равенства:

$$\begin{cases} 4x - 2y + 3z \to \max \\ -x - 2y + 4z = 8 \\ 3x - 2y - 6z = 12 \\ x, y, z \ge 0 \end{cases}$$

В целом во всех постановках кроме общей, требуется неотрицательность переменных. Кроме того, в канонической постановке ограничения – только равенства, в стандартной – только неравенства «меньше, либо равно», в основной – любые неравенства и равенства.

Все 4 вида задачи ЛП эквивалентны между собой и сводимы друг к другу следующими преобразованиями:

- если переменная x_i не ограничена, то есть система не содержит неравенства $x_i \ge 0$, то делаем замену $x_i = u_i - v_i$ и добавляем к системе неравенства $u_i, v_i \ge 0$, где u_i, v_i – новые переменные;
- ограничение-равенство заменяется на пару неравенств

$$\bar{a}_j \cdot \bar{x}^T = b_j \quad \mapsto \quad \bar{a}_j \cdot \bar{x}^T \ge b_j; \quad -\bar{a}_j \cdot \bar{x}^T \ge b_j;$$

- ограничение-неравенство заменяется на равенство и простое неравенство

$$\bar{a}_j \cdot \bar{x}^T \le b_j \quad \mapsto \quad \bar{a}_j \cdot \bar{x}^T + w_j = b_j, \quad w_j \ge 0,$$

$$\bar{a}_j \cdot \bar{x}^T \ge b_j \quad \mapsto \quad \bar{a}_j \cdot \bar{x}^T - w_j = b_j, \quad w_j \ge 0,$$

где w_j – новая переменная.

Пример. Привести следующую задачу ЛП к каноническому виду:

$$\begin{cases} 2x - 3y \to \max \\ 2x - y \le 6, \\ x \ge 0. \end{cases}$$

Решение. Задача дана в стандартной постановке.

1. Поскольку переменная y не ограничена, делаем замену $y = u - v$:

$$\begin{cases} 2x - 3(u - v) \to \max \\ 2x - (u - v) \le 6, \\ x, u, v \ge 0, \end{cases}$$

получаем задачу в ***стандартной*** постановке.

2. Единственное сложное неравенство заменяем на равенство и простое неравенство $w \ge 0$:

$$\begin{cases} 2x - 3u + 3v \to \max \\ 2x - u + v + w = 6 \\ x, u, v, w \ge 0 \end{cases}$$

получаем задачу в ***канонической*** постановке. □

В целом, любая задача в канонической и стандартной постановке является задачей в основной постановке, а та, в свою очередь является задачей в общей постановке.

Контрольные вопросы

1. Рассматривается оптимальная задача минимизации стоимости закупки трех видов материально-технических средств. Целевая функция выглядит следующим образом (ограничения в этой задаче не важны, но известно, что они линейны):

   $$z_{min} = c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_3,$$

   где

   x_i – количество закупаемых единиц i-ресурса;

   c_i – стоимость закупки единицы i-ресурса.

   Какие предположения на условия закупок позволяют использовать здесь методы линейного программирования? Как могут повлиять на коэффициенты c_i и выбор метода: оптовые скидки, дефицит необходимого ресурса, его отсутствие?
2. Чем отличаются общая, основная, стандартная и каноническая постановки задачи линейного программирования
3. Какие постановки задач линейного программирования сводятся к каким? Какими приемами?
4. Как расписывается матричная запись задачи линейного программирования

   $$\begin{cases} \overline{c} \cdot \overline{x}^T \to \max, \\ A\overline{x}^T = \overline{b}^T, \\ \overline{x} \ge \overline{0}. \end{cases} \qquad ?$$

Перейдем к рассмотрению способов решения задачи линейного программирования.

Графическое решение задачи линейного программирования

Для иллюстрации основных особенностей задачи линейного программирования рассмотрим решение примера ***графическим методом***. ***Графический*** (или как ещё говорят ***геометрический***) метод применим только к задачам линейного программирования, где количество неизвестных не превышает двух.

Пример. Решить графически задачу линейного программирования для заданной целевой функции и ограничений:

$$z_{max} = 3x_1 + 4x_2, \text{ при } \begin{cases} 3x_1 + 2x_2 \le 6 \\ x_1 + 2x_2 \le 4 \\ x_1 - x_2 \le 1 \end{cases}, \text{ где } x_1 \ge 0,\ x_2 \ge 0. \qquad (3.1)$$

Решение. Определим подобласть плоскости, определяемую заданным набором ограничений. Для этого, для каждого из пяти линейных неравенств

$$3x_1 + 2x_2 \le 6;\quad x_1 + 2x_2 \le 4;\quad x_1 - x_2 \le 1;\quad x_1 \ge 0;\quad x_2 \ge 0;$$

построим полуплоскость, определяемую этим неравенством.

Начнём с простых неравенств.

$x_1 \ge 0$ Полуплоскость справа от оси Ox_2 (рис. 3.1).

$x_2 \ge 0$ Верхняя полуплоскость (рис. 3.2).

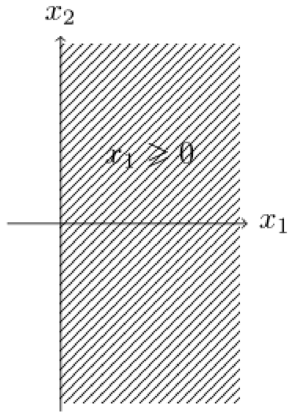


Рис. 3.1

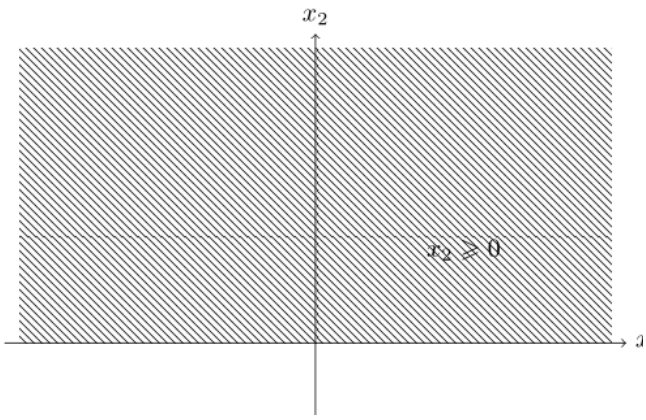


Рис. 3.2

$3x_1 + 2x_2 \le 6$ Для определения полуплоскости, задаваемой неравенством

$$3x_1 + 2x_2 \le 6, \tag{3.2}$$

построим сначала её границу, задаваемую уравнением

$$3x_1 + 2x_2 = 6. \tag{3.3}$$

Линейно уравнение определяет прямую, а для построения прямой достаточно знать две точки плоскости ей принадлежащие. Для нахождения первой точки, положим $x_1 = 0$, и из (3.3) находим, что $x_2 = 3$. Для нахождения второй точки, положим $x_2 =0$, и из (3.3) находим, что $x_1 = 2$. Таким образом, имеются две точки (0; 3) и (2; 0), которые отметим на плоскости и проведём через них искомую прямую $3x_1 + 2x_2 = 6$. Для определения нужной полуплоскости из двух возможных, возьмем координаты какой-нибудь точки, например, (0; 0), и подставим в неравенство (3.2), если неравенство выполняется, как в данном случае: $3\cdot0 + 2\cdot0 = 0 \le 6$, то эта точка и вся содержащая её полуплоскость соответствуют неравенству (рис. 3.3).

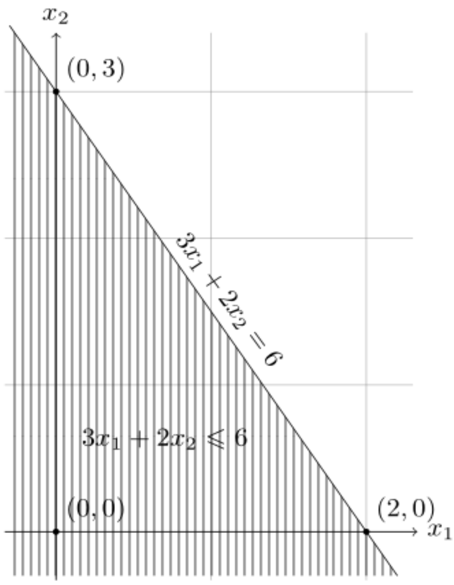


Рис. 3.3

$\underline{x_1 + 2x_2 \le 4}$ Аналогично, для определения полуплоскости, задаваемой неравенством

$$x_1 + 2x_2 \le 4, \tag{3.4}$$

построим сначала её границу, задаваемую уравнением

$$x_1 + 2x_2 = 4. \tag{3.5}$$

Как и раньше, для нахождения первой точки, положим $x_1 = 0$, и из (3.5) получаем, что $x_2 = 2$. Для второй точки, положим $x_2 = 0$, и из (3.5) находим, что $x_1 = 4$. Отметим на плоскости две найденные точки (0; 2) и (4; 0), и проведём через них прямую $x_1 + 2x_2 = 4$. Для определения нужной полуплоскости, подставляем координаты точки (0; 0) в неравенство (3.4), поскольку неравенство выполняется: $0 + 2 \cdot 0 = 0 \le 4$, то эта точка и соответствующая полуплоскость есть решение неравенства (рис. 3.4).

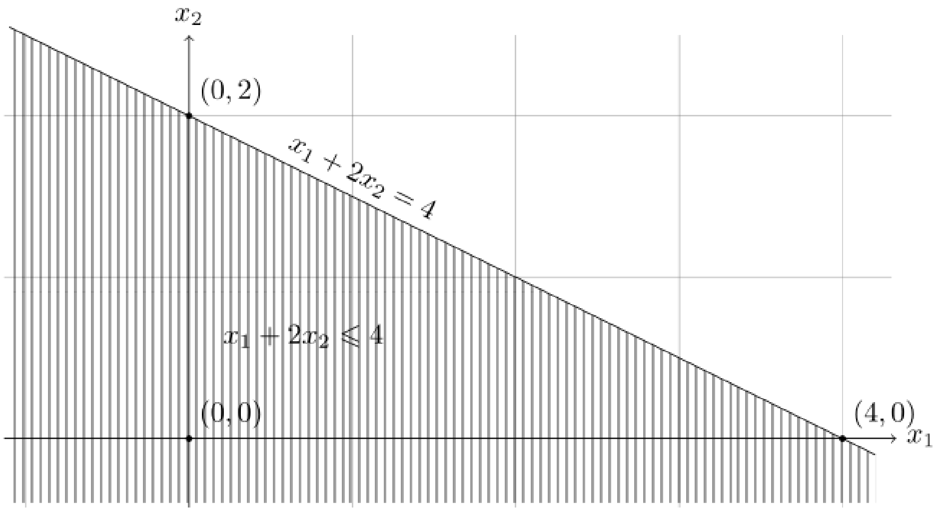


Рис. 3.4

$x_1 - x_2 \leqslant 1$ Следуя тому же плану, находим полуплоскость, соответствующую неравенству $x_1 - x_2 \leqslant 1$ (рис. 3.5).

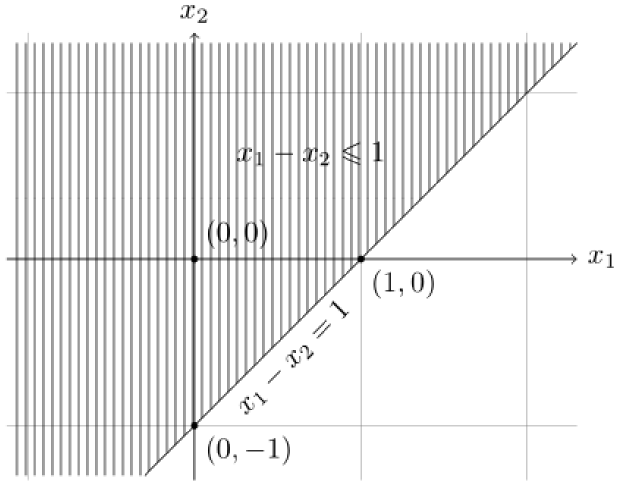


Рис. 3.5

Объединим результаты для всех неравенств и получим многоугольник, состоящий из точек, координаты которых удовлетворяют условиям (3.1) задачи (рис. 3.6).

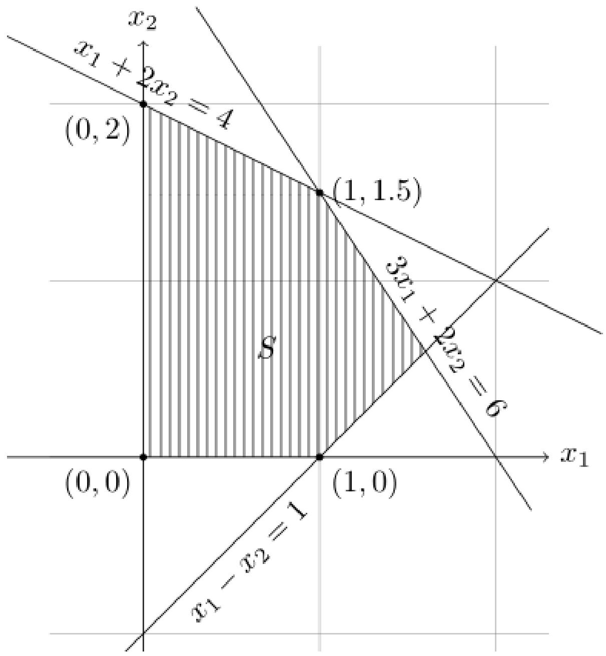


Рис. 3.6.

Далее определим направление, вдоль которого возрастает целевая функция $z = 3x_1 + 4x_2$. Для этого отложим вектор (3; 4) градиента этой функции, компоненты которого для линейной функции определяются как коэффициенты при соответствующих неизвестных. Отложим также несколько линий уровня:

$$3x_1 + 4x_2 = 0,\ 3x_1 + 4x_2 = 4,\ 3x_1 + 4x_2 = 8,\ 3x_1 + 4x_2 = 12.$$

Таким образом, при движении в направлении вектора (3; 4) значение целевой функции $z = 3x_1 + 4x_2$ возрастает и постоянно вдоль перпендикуляров к вектору (3; 4) (рис. 3.7).

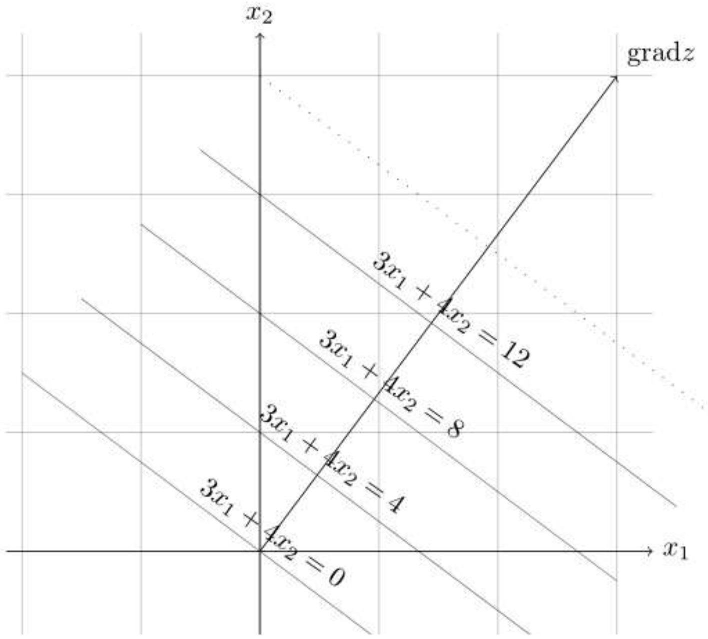


Рис. 3.7

Совмещая информацию с рис. 3.6 и 3.7, получаем рис. 3.8, из которого следует, что крайняя точка многоугольника, которой касается линия уровня целевой функции, есть точка (1; 1,5), и значение целевой функции в этой точке

$$z = 3 \cdot 1 + 4 \cdot 1{,}5 = 9 .$$

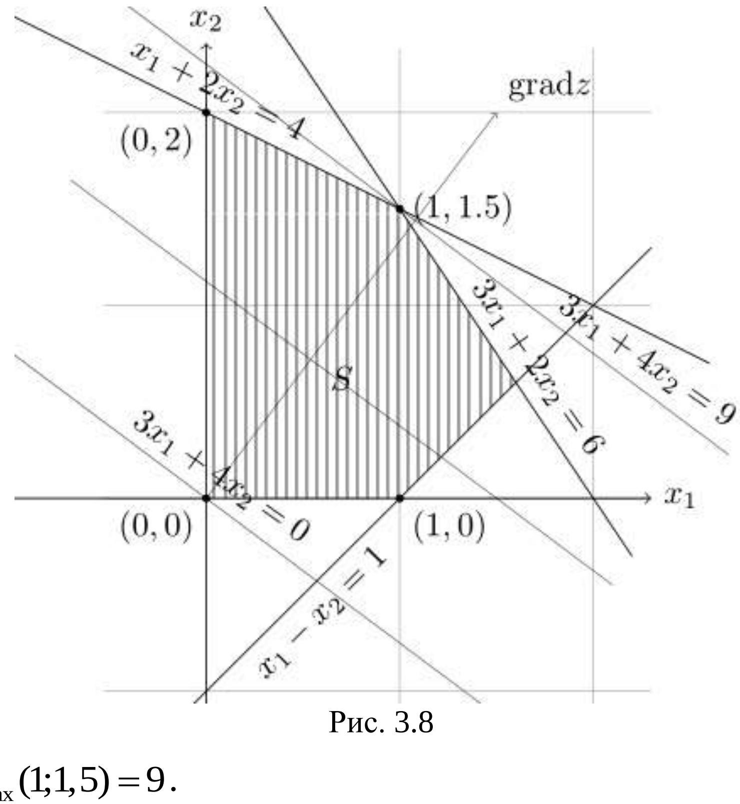


Рис. 3.8

<u>Ответ.</u> $z_{\max}(1;1{,}5) = 9$.

Симплекс-метод

Напомним, что оптимальное решение задачи линейного программирования, если оно существует, достигается в угловой точке. На основе этого свойства в 1947 году Дж. Данциг разработал специальный метод решения задач ЛП – ***симплекс-метод***. Модифицированные вариации этого метода до сих пор являются основным средством решения задач линейного программирования.

Замечание. В настоящее время в практических целях начали использоваться также ***методы внутренней точки***, которые при больших размерностях задач начинают превосходить по эффективности симплекс-метод.

Общая геометрическая схема работы симплекс-метода

1. Находим какую-нибудь вершину допустимого многоугольника.

2. Начиная с найденной вершины, двигаемся вдоль ребер допустимого многогранника от вершины к вершине, в сторону возрастания целевой функции.

3. При размерности задачи 2 и выше, из любой вершины всегда выходит несколько ребер, причем в сторону возрастанию целевой функции могут быть направлены несколько из них. За то, какое из них выбрать, отвечает, так называемое, ***правило поворота*** **(***англ.* **pivot rule)**.

Замечание. Существует множество вариантов правила поворота.

Алгебраически, геометрическая схема работы симплекс-метода реализуется следующим образом.

Алгебраическая схема работы симплекс-метода

Пусть n – количество переменных задачи.

1. Каждая вершина задается некоторой системой из n ***базисных*** ограничений-равенств (координаты вершины есть решение этой ***базисной*** системы линейных уравнений-ограничений).

2. Переход от одной вершины к другой заключается в замене одного из ограничений-равенств на другое – говорят, что одно из ограничений ***выводится*** из списка базисных, а взамен него ***вводится*** другое.

Особенности и проблемы алгебраической реализации

1. ***Зацикливание.*** Для задания вершины в пространстве $\mathbb{R}^n$ достаточно n линейно-независимых линейных ограничений-равенств. Однако, в так называемом ***вырожденном случае,*** координаты вершины удовлетворяют системе из большего количества ограничений-равенств (хотя можно выделить эквивалентную подсистему из n базовых ограничений, задающих эту же вершину, причем в общем случае, можно выделить не единственную такую подсистему). На рис. 3.9 вершина D соответствует невырожденному случаю, а все остальные вершины A, B, C, E и F вдоль отмеченного на рисунке пути соответствуют вырожденным случаям. Если при замене базисных ограничений вводятся только ограничения из исходной расширенной системы, то и происходит зацикливание, причем при этом текущая вершина реально не меняется. Имеются специальные эффективные методы борьбы с зацикливанием симплекс-метода.

2. *Высокая теоретическая сложность.* Практически симплекс-метод показывает очень хорошую эффективность: в среднем для решения задачи требуется от m до $3m+n$ итераций (смен вершин), где m – число ограничений задачи, n – число переменных.

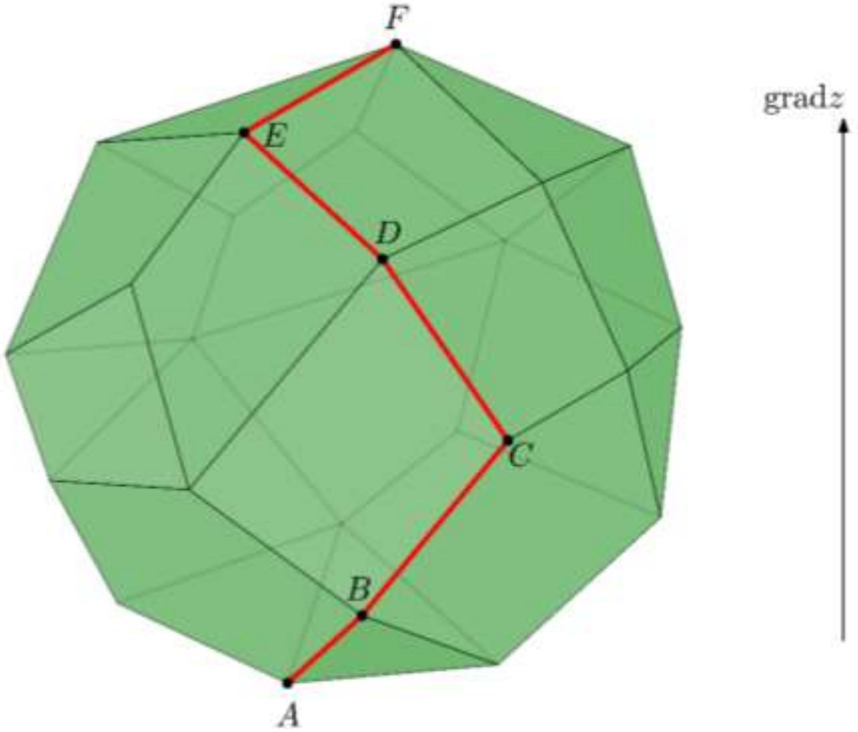


Рис. 3.9. Движение вдоль ребер допустимого многогранника при реализации симплекс-метода.

Однако на допустимых многогранниках особого вида (их называют кубы Кли-Минти, по именам авторов первого такого примера; кубы Кли-Минти выглядят как деформированные кубы, см. рис. 3.10), симплекс-метод показывает экспоненциальную сложность (обходит все вершины допустимого многогранника). Причем для каждого из основных известных правил поворота, построен свой куб Кли-Минти. Проблема существования полиномиально быстрой реализации симплекс-метода на сегодняшний день остается открытой.

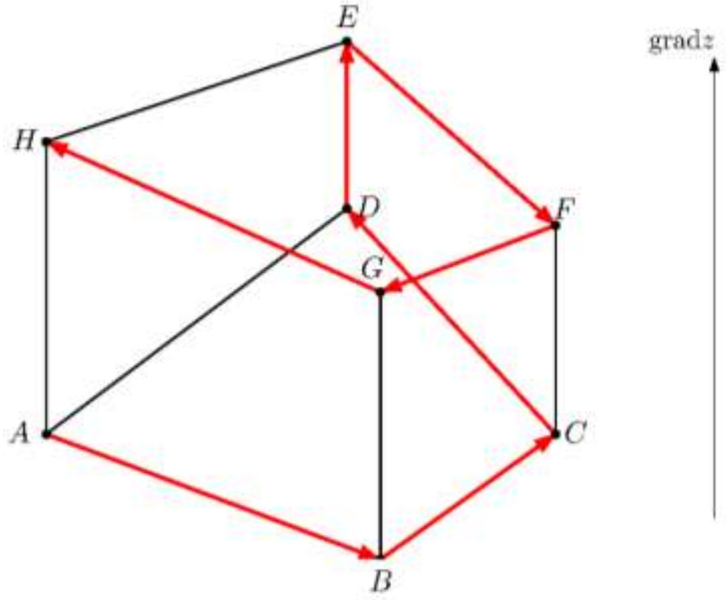


Рис. 3.10. Куб Кли-Минти для простого правила поворота.

Разберем работу симплекс-метода на примере.

Пример. Решить задачу ЛП заданную в стандартном виде:

$$\begin{cases} z = 5x_1 + 4x_2 + 3x_3 \to \max \\ 2x_1 + 3x_2 + x_3 \le 5 \\ 4x_1 + x_2 + 2x_3 \le 11 \\ 3x_1 + 4x_2 + 2x_3 \le 8 \\ x_1, x_2, x_3 \ge 0 \end{cases}$$

Решение. Добавим ***дополнительные*** переменные w_i, по одной на каждое из ограничений-равенств (тем самым переходя к каноническому виду):

$$\begin{cases} z = 5x_1 + 4x_2 + 3x_3 \to \max \\ w_1 = 5 - 2x_1 - 3x_2 - x_3 \\ w_2 = 11 - 4x_1 - x_2 - 2x_3 \\ w_3 = 8 - 3x_1 - 4x_2 - 2x_3 \\ x_1, x_2, x_3, w_1, w_2, w_3 \ge 0 \end{cases}$$

Будем переходить от одного допустимого решения к другому, улучшая значение целевой функции между итерациями.

В данном случае легко найти допустимое решение, например, полагаем что $x_1 = x_2 = x_3 = 0$, получим

$$w_1 = 5,\ w_2 = 11,\ w_3 = 8.$$

При этом решении значение допустимой функции равно 0.

По виду целевой функции видно, что при увеличении с 0 значений переменных x_1, x_2, x_3, значение целевой функции будет только увеличиваться, поскольку эти переменные входят в целевую функцию с положительными коэффициентами.

Проверим, на сколько их можно независимо увеличить, чтобы не нарушить условия w_1, w_2, $w_3 \geq 0$.

$$x_1: 5-2x_1 \geq 0,\ 11-4x_1 \geq 0,\ 8-3x_1 \geq 0,\ \ x_2 = x_3 = 0$$

$$\Rightarrow \quad x_1 \leq \min\left\{\frac{5}{2}, \frac{11}{4}, \frac{8}{3}\right\} \quad \Rightarrow \quad x_1 \leq \frac{5}{2}$$

$$x_2: 5-3x_2 \geq 0, 11-4x_2 \geq 0, 8-4x_2 \geq 0,\ \ x_1 = x_3 = 0$$

$$\Rightarrow \quad x_2 \leq \min\left\{\frac{5}{3}, \frac{11}{2}, 2\right\} \quad \Rightarrow \quad x_2 \leq \frac{5}{3}$$

$$x_3: 5-x_3 \geq 0, 11-2x_3 \geq 0, 8-2x_3 \geq 0,\ \ x_1 = x_2 = 0$$

$$\Rightarrow \quad x_3 \leq \min\left\{5, \frac{11}{2}, 4\right\} \quad \Rightarrow \quad x_3 \leq 4$$

С учетом коэффициентов 5, 4, 3 целевой функции, видно, что наибольшее увеличение целевой функции принесет изменение x_1 с 0 до 5/2, причем w_1 изменится с 5 до 0, то есть займёт место x_1 в списке базисных переменных. Для этого, из 1-го ограничения $w_1 = 5 - 2x_1 - 3x_2 - x_3$ исходной системы, выразим переменную x_1:

$$x_1 = \frac{5}{2} - \frac{1}{2}w_1 - \frac{3}{2}x_2 - \frac{1}{2}x_3.$$

Новое базисное решение

$$(x_1, x_2, x_3, w_1, w_2, w_3) = (5/2; 0; 0; 0; 1; 1/2).$$

Новое значение целевой функции

$$z = 5 \cdot 5/2 + 4 \cdot 0 + 3 \cdot 0 = 12.5.$$

Новая система получается из исходной подстановкой вместо всех вхождений переменной x_1 её выражения $5/2 - 1/2\,w_1 - 3/2\,x_2 - 1/2\,x_3$ через переменные w_1, x_2, x_3. После арифметических преобразований и приведения подобных получаем систему:

$$\begin{cases} z = \frac{25}{2} - \frac{5}{1}w_1 - \frac{7}{2}x_2 + \frac{1}{2}x_3 \to \max \\ x_1 = \frac{5}{2} - \frac{1}{2}w_1 - \frac{3}{2}x_2 - \frac{1}{2}x_3 \\ w_2 = 1 + 2w_1 + 5x_2 \\ w_3 = \frac{1}{2} + \frac{3}{2}w_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 \\ x_1, x_2, x_3, w_1, w_2, w_3 \ge 0 \end{cases}$$

По новому виду целевой функции видно, что увеличить можно только переменную x_3. Найдем промежутки изменению x_3, сохраняющие неотрицательность переменных x_1, w_2, w_3:

$$x_1 = \frac{5}{2} - \frac{1}{2}\cdot 0 - \frac{3}{2}\cdot 0 - \frac{1}{2}\cdot 0 - \frac{1}{2}x_3 \ge 0 \quad \Rightarrow \quad x_3 \le 5;$$

$w_2 = 1 + 2w_1 + 5x_1 \ge 0$ – не зависит от x_3,

следовательно $x_3 < \infty$;

$$w_3 = \frac{1}{2} + \frac{3}{2}\cdot 0 + \frac{1}{2}\cdot 0 - \frac{1}{2}x_3 \ge 0 \quad \Rightarrow \quad x_3 \le 1.$$

Таким образом: $x_3 \le \min\{5, \infty, 1\} \quad \Rightarrow \quad x_3 \le 1.$

Новое базисное решение

$$(x_1, x_2, x_3, w_1, w_2, w_3) = (2, 0, 1, 0, 1, 0).$$

Новое значение целевой функции: z = 3.

Новая система

$$\begin{cases} z = 13 - 2w_1 - 3x_2 - w_3 \\ x_1 = 2 - 2w_1 - 2x_2 + w_3 \\ w_2 = 1 + 2w_1 + 5x_2 \\ w_3 = 1 + 3w_1 + x_2 - 2w_3 \\ x_1, x_2, x_3, w_1, w_2, w_3 \ge 0 \end{cases}$$

Больше увеличить целевую функцию нельзя, поскольку все коэффициенты в ней отрицательны – найден максимум.

Ответ: $z_{\max}(2;0;1) = 13.$

Замечание. В рассмотренном примере изначально имелось допустимое базисное решение, благодаря тому, что все b_i были положительные. В общем случае это не выполняется. Поэтому применение симплекс-метода разбивается на две фазы: на первой фазе находится какое-нибудь допустимое базисное решение, на второй – решается собственно задача линейного программирования.

Контрольные вопросы

1. Опишите общую схему действия симплекс-метода.
2. Что такое «правило поворота»?
3. В чем заключается «зацикливание» симплекс-метода?
4. Что называется «базисным решением» задачи ЛП?
5. Что такое куб Кли-Минти?
6. Каков наихудший сценарий использования симплекс-метода на допустимом многограннике?

Двойственная задача линейного программирования

С понятием ***двойственной*** задачи удобно познакомиться на примере экономической задачи оптимального распределения ресурсов.

Пример (***Задача о выпуске продукции***). Пусть предприятие выпускает 4 вида продукции, стоимостью 2, 4, 1 и 5 тыс. рублей за единицу продукции каждого вида. Для выпуска продукции каждого из 4 видов продукции используется 3 вида ресурсов, имеющихся в количестве 30, 10 и 40 единиц.

Для выпуска единицы продукции используется следующее количество каждого из трех видов ресурсов:

для 1-ого вида: 4, 0 и 1 единиц;

для 2-ого вида: 1, 4 и 0 единиц;

для 3-ого вида: 2, 1 и 1 единиц;

для 4-ого вида: 1, 2 и 2 единиц.

Задача. С учетом ресурсных ограничений спланировать производство (составить *программу* выпуска), так чтобы максимизировать стоимость выпущенной продукции.

Введем переменные x_1, x_2, x_3, x_4 для обозначения количества каждого из четырех видов выпускаемой продукции.

Тогда задача о выпуске продукции сводится к задаче линейного программирования:

$$\begin{cases} 2x_1 + 4x_2 + x_3 + 5x_4 \to \max \\ 4x_1 + x_2 + 2x_3 + x_4 \le 30 \\ \qquad 4x_2 + x_3 + 2x_4 \le 10 \\ x_1 + \ \ 0 \ + x_3 + 2x_4 \le 40 \\ x_1, x_2, x_3, x_4 \ge 0 \end{cases}$$

Задача дана в стандартной постановке.

При решении стандартной задачи симплекс-методом, для каждого сложного ограничения-неравенства

$$a_{j1}x_1 + \ldots + a_{jn}x_n \le b_j$$

вводится новая ***дополнительная*** переменная

$$w_j = b_j - (a_{j1}x_1 + \ldots + a_{jn}x_n)\,.$$

В задаче о выпуске продукции значение переменной w_i соответствует остаткам j-ого вида сырья при выбранном плане производства $x_1, \ldots, x_n$.

Поставим теперь ***двойственную задачу*** (Д) к исходной задаче выпуска продукции (И).

Задача (Д). Предположим, что сырье можно распродавать, например, чтобы не держать на складе. Найти минимальную стоимость продажи сырья, чтобы выручка не упала по сравнению с продажей готовой продукции.

Решение. Пусть неизвестные y_1, y_2, y_3 обозначают ***условную*** (***продажную***) ***цену*** на сырье каждого вида.

Если i-ый вид продукции продается по цене c_i, то, чтобы выручка не упала, цены y_1, y_2, y_3 на продаваемое сырье должны удовлетворять неравенству

$$a_{1i}y_1 + a_{2i}y_2 + a_{3i}y_3 \ge c_i\,,$$

то есть условная (продажная) стоимость сырья, использованного в производстве i-ой продукции, должна быть не меньше чем стоимость этой продукции.

Кроме того, очевидно, что если j-ый вид сырья был в избытке ($w_j > 0$), то можно продавать его хоть по нулевой цене, выручка от этого не упадет. В этом случае $y_j = 0$.

Таким образом задача нахождения минимальной приемлемой стоимости сырья приобретает вид (члены с нулевыми коэффициентами пропущены):

$$\begin{cases} 30y_1 + 10y_2 + 40y_3 \to \min \\ 4y_1 + \quad 0 \;+ \quad y_3 \ge 2 \\ \;y_1 \;+ 4y_2 + \quad 0 \;\ge 4 \\ 2y_1 + \quad y_2 + \quad y_3 \ge 1 \\ \;y_1 + 2y_2 + 2y_3 \ge 5 \\ y_1, y_2, y_3 \ge 0 \end{cases}$$

Если $\hat{x}$, $\hat{y}$ оптимальные решения для исходной и двойственной задачи, то оптимальные значение целевых функции этих задач равны.

Экономическая интерпретация двойственности

(И) Исходная задача имеет следующий экономический смысл:

- основные переменные x_i обозначают количество произведенной продукции i-ого вида;
- j-ое ограничение-неравенство обозначает затраты j-ого сырья в сравнении с общим запасом этого сырья;
- дополнительные переменные w_i обозначают количество излишков каждого вида сырья при выбранном плане производства;
- целевая функция дает стоимость произведенной продукции (выручку).

(Д) Двойственная задача имеет следующий экономический смысл:

- основные переменные y_j обозначают цены при которых можно производить из сырья продукцию, без потери выручки по сравнению с продажей сырья;
- i-ое ограничение-неравенство обозначает выручку от продажи сырья, использованного при производстве единицы i-ой продукции, по сравнению со стоимостью этой продукции;
- целевая функция дает условную суммарную стоимость имеющегося сырья.

В общем случае задача двойственная стандартной задачи ЛП строится по следующим правилам.

1. Количество переменных в двойственной задаче равно количеству ограничений в исходной.
2. Матрица коэффициентов двойственной задачи равна транспонированной матрице коэффициентов исходной.
3. Столбец свободных членов исходной задачи является строкой коэффициентов целевой функции двойственной.
4. Целевая функция в одной задаче минимизируется, в другой максимизируется.

5. Условиям неотрицательности переменных исходной задачи соответствуют неравенства-ограничения двойственной направленные в другую сторону.
6. Неравенствам-ограничениям в исходной задаче соответствуют условия неотрицательности в двойственной.

В матричной записи эти правила реализуются следующим образом:

$$\text{(И)} \qquad\qquad\qquad\qquad \text{(Д)}$$

$$\begin{cases} \overline{c} \cdot \overline{x}^T \to \max \\ A\overline{x}^T \le \overline{b} \\ \overline{x} \ge \overline{0} \end{cases} \quad \Leftrightarrow \quad \begin{cases} \overline{b} \cdot \overline{y}^T \to \min \\ A^T \overline{y}^T \ge \overline{c} \\ \overline{y} \ge \overline{0} \end{cases}$$

Видно, что для задачи в стандартной постановке, задача двойственная к двойственной к стандартной совпадает с исходной.

Для задачи из примера двойственная задача будет выглядеть следующим образом:

$$\begin{array}{lcl} \text{(И)} & & \text{(Д)} \\ 2x_1 + 4x_2 + x_3 + 5x_4 \to \max & & 30y_1 + 10y_2 + 40y_3 \to \min \\ 4x_1 + x_2 + 2x_3 + x_4 \le 30 & \longleftrightarrow & y_1 \ge 0 \\ 4x_2 + x_3 + 2x_4 \le 10 & \longleftrightarrow & y_2 \ge 0 \\ x_1 + \ 0 \ + x_3 + 2x_4 \le 40 & \longleftrightarrow & y_3 \ge 0 \\ x_1 \ge 0 & \longleftrightarrow & 4y_1 + \ 0 \ + y_3 \ge 2 \\ x_2 \ge 0 & \longleftrightarrow & y_1 + 4y_2 \ge 4 \\ x_3 \ge 0 & \longleftrightarrow & 2y_1 + y_2 + y_3 \ge 1 \\ x_4 \ge 0 & \longleftrightarrow & y_1 + 2y_2 + 2y_3 \ge 5 \end{array}$$

Неравенства, соединенные стрелками, называются ***сопряженными***.

Фундаментальная роль двойственности в линейном программировании связана со следующими фактами.

Теорема 1 (*Первая теорема двойственности*.) Если одна из пары двойственных задач (И) и (Д) имеет оптимальный план $\hat{x}$, то и двойственная задача имеет оптимальный план $\hat{y}$, при этом оптимальные значения целевых функций равны: $\overline{c}\hat{x}^T = \overline{b}\hat{y}^T$.

Если целевая функция одной задачи неограничена, то вторая не имеет решения.

Теорема 2 (*Вторая теорема двойственности*.) Планы $\hat{x}$ и $\hat{y}$ оптимальны в задачах (И) и (Д) тогда и только тогда, когда при подстановке их в ограничения соответствующих задач, для каждой пары сопряженных неравенств, хотя бы одно из них обращается в равенство.

Условия в Теореме 2 называют ***условиями дополняющей нежесткости***.

Между переменными в исходной и двойственной задаче существует тесная взаимосвязь, а именно: ***основные и дополнительные переменные взаимодвойственных задач находятся во взаимно-однозначном соответствии***.

Контрольные вопросы

1. Как формулируется двойственная задача линейного программирования?
2. Каков экономический смысл переменных двойственной задачи?
3. Как оптимальное решение двойственной задачи связано с оптимальным решением исходной задачи?
4. Сформулируйте первую и вторую теорему двойственности.

Транспортная задача

Постановка транспортной задачи

Пусть требуется развезти с наименьшими затратами груз с m складов в n пунктов распределения, при следующих заданных условиях:

A_i – имеющиеся запасы на i-ом складе;

B_j – потребность на j-ом пункте распределения;

c_{ij} – стоимость перевозки единицы груза с i-го склада на j-ый пункт распределения.

План перевозок задается матрицей неизвестных $X = (x_{ij})_{m\times n}$, где x_{ij} – количество перевезенного груза с i-го склада на j-ый пункт распределения.

Решением задачи будет ***оптимальный план перевозок***, то есть план перевозок с минимальной стоимостью.

Условия транспортной задачи и ход решения будем записывать в таблицу следующего вида, где в ячейках мелким шрифтом будем записывать соответствующие тарифы c_{ij}:

$A_i \backslash B_j$	B_1	$\ldots$	B_n	$\sum_{j=1}^{n} B_j$
A_1	c_{11}	$\ldots$	c_{1n}	
$\vdots$	$\vdots$	$\ddots$	$\vdots$	
A_n	c_{m1}	$\ldots$	c_{mn}	
$\sum_{i=1}^{m} A_i$				$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\cdot x_{ij}$

Транспортную задачу можно сформулировать как задачу линейного программирования:

$$\begin{cases} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\cdot x_{ij} \to \min \\ \sum_{i=1}^{n} x_{ij} = B_j, \quad j = 1,\ldots,m; & (3.6) \\ \sum_{j=1}^{m} x_{ij} = A_i, \quad i = 1,\ldots,n; & (3.7) \\ x_{ij} \ge 0, \quad i = 1,\ldots,m;\ j = 1,\ldots,n; & (3.8) \end{cases}$$

где равенство (3.6) означает, что при перевозках будут удовлетворены потребности на всех пунктах распределениях, равенство (3.7) означает, что все запасы со складов будут вывезены.

Определение. ***Допустимый план перевозок*** – это любой план перевозок $X = (x_{ij})_{m\times n}$, удовлетворяющий условиям (3.6)-(3.8).

Для совместности транспортной задачи требуется чтобы:

$$\sum_{i=1}^{m} A_i = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} x_{ij}\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} x_{ij}\right) = \sum_{j=1}^{n} B_j .$$

другими словами, требуется чтобы суммарные запасы на складах были равны суммарной потребности в пунктах распределения.

Определение. Транспортная задача, для которой

$$\sum_{i=1}^{m} A_i = \sum_{j=1}^{n} B_j ,$$

называется ***транспортной задачей закрытого*** типа.

Любую задачу ***открытого типа*** $\left(\sum_{i=1}^{m} A_i \neq \sum_{j=1}^{n} B_j\right)$ можно свести к задаче закрытого типа одним из следующих способов.

- Если $\sum_{i=1}^{m} A_i < \sum_{j=1}^{n} B_j$, то вводим $m+1$-ый фиктивный склад с запасом $A_{m+1} = \sum_{j=1}^{n} B_j - \sum_{i=1}^{m} A_i$, полагая $c_{m+1,j} = 0,\ j = 1,\ldots,n$.
- Если $\sum_{i=1}^{m} A_i > \sum_{j=1}^{n} B_j$, то вводим $n+1$-ый фиктивный пункт распределения («свалку») с потребностью $B_{n+1} = \sum_{i=1}^{m} A_i - \sum_{j=1}^{n} B_j$, полагая $c_{i,n+1} = 0,\ i = 1,\ldots,m$.

Вырожденная транспортная задача

Определение. Если какой-то допустимый план транспортной задачи содержит меньше чем $m + n - 1$ перевозок (клеток) с $x_{ij} > 0$, то и план и транспортная задача называются ***вырожденными***. Если такого допустимого плана нет, то задача называется ***невырожденной***.

Критерий вырожденности транспортной задачи дается следующей теоремой.

Теорема. Закрытая транспортная задача вырождена тогда и только тогда, когда существует неполная группа складов, которая полностью удовлетворяет потребности некоторой группы пунктов распределения.

Замечания.

1. Транспортную задачу можно решать, как задачу линейного программирования с $m \times n$ переменными на которые накладываются $m + n$ ограничений.

2. Более эффективными являются специально разработанные методы вроде ***метода потенциалов*** или ***венгерского метода***.

3. Метод потенциалов, который будет рассмотрен далее, начинает работу с некоторого ***опорного*** плана перевозок (допустимого плана перевозок), необязательно оптимального, а затем последовательно улучшает его, до достижения оптимальности.

4. В следующем вопросе будут рассмотрены два метода получения опорного плана.

Контрольные вопросы

1. Сформулируйте транспортную задачу как задачу линейного программирования. Поясните обозначения и смысл ограничений. Почему задача линейная?
2. Что является решением транспортной задачи?
3. Какая транспортная задача называется закрытой? Открытой?
4. Что называется допустимым планом перевозок? Опорным планом?
5. Какая транспортная задача называется вырожденной, невырожденной?
6. Может ли для вырожденной задачи существовать невырожденный план перевозок?
7. Может ли для невырожденной задачи существовать вырожденный план перевозок?

Методы построения первоначального опорного плана

Рассмотрим два метода построения первоначального опорного плана для закрытой модели транспортной задачи: ***метод северо-западного угла*** (СЗУ) и ***метод минимального элемента*** (ММЭ).

Метод северо-западного угла

Таблицу перевозок начинаем заполнять с левой верхней клетки (***северо-западного угла***), двигаясь по строкам слева-направо.

Величину ***перевозки*** x_{ij} в ij-ой клетке выбираем как минимум значения из двух величин: нераспределенного еще запаса на i-ом шаге или неудовлетворенной еще потребности в j-ом пункте распределения.

Если потребность в данном пункте уже удовлетворена, то переходим к следующей клетке в этой строке.

Если ресурс в данной строке (складе) исчерпан, то переходим к следующей строке.

Пример. Условия перевозок заданы в таблице:

$A_i \backslash B_j$	40	70	40	100	
100	3	4	1	5	
40	5	3	2	4	
60	6	1	7	5	
50	1	2	3	3	

<u>Решение.</u> Проверим, что задача закрытого типа

$$\sum_{i=1}^{m} A_i = 100 + 40 + 60 + 50 = 250 = \sum_{j=1}^{n} B_j = 40 + 70 + 40 + 100.$$

Далее заполним таблицу согласно описанию метода.

$A_i \backslash B_j$	40	70	40	100	250
100	40 ^{3}	60 ^{4}	— ^{1}	— ^{5}	
40	| ^{5}	10 ^{3}	30 ^{2}	— ^{4}	
60	| ^{6}	| ^{1}	10 ^{7}	50 ^{5}	
50	| ^{1}	| ^{2}	| ^{3}	50 ^{3}	
250					820

Стоимость найденного невырожденного опорного плана перевозок равна:

$$Q(X) = 3 \cdot 40 + 4 \cdot 60 + 3 \cdot 10 + 2 \cdot 30 + 7 \cdot 10 + 5 \cdot 50 + 3 \cdot 50 =$$
$$= 120 + 240 + 30 + 60 + 70 + 250 + 150 = 820.$$

Метод минимального элемента

В таблице ищется клетка с минимальным значением c_{ij} и заполняется значением $x_{ij} = \min\{A_i, B_j\}$.

- Если $A_i < B_j$, то вычеркивается остаток i-ой строки.
- Если $A_i > B_j$, то вычеркивается остаток j-ого столбца.
- Вместо A_i, B_j записываются нераспределенный остаток на i-ом складе и неудовлетворенная потребность на j-ом пункте распределения, соответственно.
- Процедура повторяется сначала.

Пример. Построить начальный опорный план методом минимального элемента, если условия перевозок заданы в следующей таблице:

$A_i \backslash B_j$	40	70	40	100	
100	3	4	1	5	
40	5	3	2	4	
60	6	1	7	5	
50	1	2	3	3	

<u>Решение.</u> Мы уже проверили, что это задача закрытого типа.

Далее заполним таблицу согласно описанию метода ММЭ.

$A_i \backslash B_j$	40	70	40	100	250
100	$\vert$ ^{3}	$\vert$ ^{4}	40 ^{1}	60 ^{5}	
40	$\vert$ ^{5}	10 ^{3}	$\vert$ ^{2}	30 ^{4}	
60	— ^{6}	60 ^{1}	$\vert$ ^{7}	— ^{5}	
50	40 ^{1}	$\vert$ ^{2}	$\vert$ ^{3}	10 ^{3}	
250					620

Стоимость найденного опорного плана перевозок равна:

$$Q(X) = 1\cdot 40 + 3\cdot 10 + 1\cdot 60 + 1\cdot 40 + 5\cdot 60 + 4\cdot 30 + 3\cdot 10 =$$
$$= 40 + 30 + 60 + 40 + 300 + 120 + 30 = 620.$$

Контрольные вопросы

1. Как расшифровывается аббревиатура СЗУ? ММЭ?
2. Какую задачу выполняют методы северо-западного угла и метод минимального элемента?
3. Что вычисляют метод потенциалов и венгерский метод?
4. Кратко опишите метод северо-западного угла.
5. Кратко опишите метод минимального элемента.

Метод потенциалов

Рассмотрим один из методов решения транспортной задачи – ***метод потенциалов***, который работает если уже имеется опорный план перевозок.

Для решения транспортной задачи методом потенциалов, для каждого склада и пункта распределения вводятся дополнительные числовые величины – ***потенциалы***:

- для каждого склада A_i – соответствующий потенциал обозначим
$$u_i,\ i = 1,\ldots,m;$$
- для пункта распределения B_j – потенциал обозначим
$$v_j,\ \ j = 1,\ldots,n.$$

Получаем набор $u_1,\ldots,u_m,v_1,\ldots,v_n$ из $m+n$ величин.

Определение. Допустимый план перевозок $X = (x_{ij})_{m\times n}$ называется ***потенциальным***, если существует набор потенциалов $u_1, \ldots, u_m, v_1, \ldots, v_n$ такой, что выполняются следующие ***условия потенциальности***:

$$v_j - u_i \le c_{ij}\text{, для всех } i = 1, \ldots, m;\ j = 1, \ldots, n; \tag{3.9}$$

$$v_j - u_i = c_{ij}\text{, если } x_{ij} = 0. \tag{3.10}$$

В основе метода потенциалов лежит следующая теорема.

Теорема. Допустимый план перевозок транспортной задачи является оптимальным тогда и только тогда, когда он потенциален.

При описании алгоритма метода потенциалов используются следующие понятия.

1. ***Перевозка*** – любое значение x_{ij} или клетка таблицы, содержащая это значение.

2. ***Цепь*** – любой упорядоченный набор перевозок в котором каждая пара соседних клеток расположены либо в одном столбце, либо в одной строке (необязательно в соседних клетках), причем переходы по строку и столбцу чередуются.

3. ***Цикл*** – цепь, первый и последний элементы которой, совпадают, причем первый и последний переход тоже чередуются.

Определить цикл не всегда просто, как показывают примеры на рис. 4.1.

а)

$A_i \backslash B_j$	…	…	…
…	**2**	…	**7**
…	…	…	…
…	**3**	…	**5**

б)

$A_i \backslash B_j$	…	…	…
…	**2**	…	**7**
…	**1**	**4**	…
…	…	**3**	**5**

в)

$A_i \backslash B_j$	…	…	…
…	**2**	**1**	…
…	**4**	…	**6**
…	…	**3**	**5**

г)

$A_i \backslash B_j$	…	…	…
…	**2**	**4**	**7**
…	…	…	…
…	**3**	…	**5**

д)

$A_i \backslash B_j$	…	…	…
…	**2**	**1**	…
…	**5**	…	…
…	**3**	**4**	**6**

е)

$A_i \backslash B_j$	…	…	…
…	**2**	**1**	**7**
…	**5**	…	**6**
…	**3**	**4**	…

Рис 4.1. Примеры циклов.

На рис. 4.1 (многоточия обозначают одну или несколько пропущенных нулевых строк или столбцов) в таблице:

а) имеется единственный цикл 2-7-3-5-2 (он же 7-5-3-2-7 или 5-3-2-7-5 или 3-2-7-5-3 – будем придерживаться порядка обхода циклов по часовой стрелке, при этом не имеет значения с какого элемента начинается перечисление цикла, при условии, что сохраняется правильный порядок перечисления);

б) имеется единственный цикл 2-7-5-3-4-1-2;

в) имеется единственный цикл 2-1-3-5-6-4-2;

г) имеется единственный цикл 2-7-5-3-2;

д) имеется единственный цикл 2-1-4-3-2;

е) имеется три цикла: 2-1-4-3-2, 2-7-6-5-2 и 1-7-6-5-3-4-1. Цепь 2-7-6-5-3-4-1-2 циклом не являются поскольку первый и последний переходы 2-7 и 1-2 оба делаются по строке, то есть без чередования.

Для понимания работы метода, сначала рассмотрим его общую схему, которую разберем затем на примере.

Общая схема метода потенциалов

I. *Предварительный этап.*

Любым доступным методом (например, СЗУ или ММЭ) строится первоначальный опорный план.

II. *Основной этап.*

1. Проверяем имеющийся допустимый план на оптимальность, рассчитав для него потенциалы.
2. Если план потенциален переходим к пункту 4, если нет, то к пункту 3.
3. Улучшаем план перевозок по специальному алгоритму, возвращаемся к пункту 1.
4. Имеющийся план оптимален, рассчитываем стоимость перевозки и записываем ответ.

Покажем расчет потенциалов на примере.

Пример. Условия задачи заданы в таблице $A_i \backslash B_j$

$A_i \backslash B_j$	40	70	50	
20	**4**	**2**	**1**	
80	**7**	**6**	**5**	
60	**3**	**1**	**7**	

<u>Решение.</u> Проверим задачу на закрытость и построим опорный план перевозок методом северо-западного угла.

$A_i \backslash B_j$	40	70	50	160
20	20 **3**	— **4**	— **1**	
80	20 **5**	60 **3**	— **2**	
60	| **6**	10 **1**	50 **7**	
160				940

Рассчитаем потенциалы для полученного опорного плана.

$$u_1 := 0$$
$$x_{11} > 0:\ v_1 - u_1 = 4 \ \Rightarrow\ v_1 = u_1 + 4 = 4;$$
$$x_{21} > 0:\ v_1 - u_2 = 7 \ \Rightarrow\ u_2 = v_1 - 7 = -3;$$
$$x_{22} > 0:\ v_2 - u_2 = 6 \ \Rightarrow\ v_2 = u_2 + 6 = -3;$$
$$x_{32} > 0:\ v_1 - u_2 = 7 \ \Rightarrow\ u_2 = v_1 - 7 = -3;$$
$$x_{33} > 0:\ v_3 - u_3 = 7 \ \Rightarrow\ v_3 = u_3 + 7 = 9.$$

Получаем

$A_i \backslash B_j$	40	70	50	160	
20	20 [3]	— [4]	— [1]	0	u_1
80	20 [5]	60 [3]	— [2]	-3	u_2
60	| [6]	10 [1]	50 [7]	2	u_3
160	4	3	9	940	
	v_1	v_2	v_3		

Контрольные вопросы

1. Сколько потенциалов вводится для задачи с m складами и n пунктами распределения?
2. Сформулируйте условия потенциальности.
3. Как оптимальность допустимого плана перевозок можно определить через условия потенциальности?

Теперь рассмотрим полное решение транспортной задачи на примере.

Пример. Решить транспортную задачу тремя способами

Склады	Пункты назначения				Запасы
	B1	B2	B3	B4	
A1	2	4	7	1	10
A2	3	6	5	8	10
A3	6	1	6	3	14
Потребности	7	6	9	12	

I. Используя СЗУ и метод потенциалов.

II. Используя ММЭ и метод потенциалов.

III. Используя подходящие программные средства.

Решение. Проведем предварительный анализ задачи.

Количество неизвестных: $m \cdot n$ = 3· 4 =12.

Задача является задачей закрытого типа, поскольку суммарно

«Потребности» = 7 + 6 + 9 + 12 = 34 = 10 + 10 +14 = «Запасы»

Задача является невырожденной, поскольку никакое меньшее чем три количество складов не удовлетворяет полностью потребности никакой группы пунктов назначения.

В силу невырожденности задачи, каждый опорный план будет иметь ровно $m + n - 1 = 3 + 4 - 1 = 6$ ненулевых значений перевозок.

I. СЗУ и метод потенциалов.

Замечание. Подробно разберем использование метода потенциалов на примере первой итерации. Расчеты по дальнейшим итерациям записываться не будут, а результаты будут сразу записываться в таблицу.

1. Получим начальный опорный план перевозок, используя метод северо-западного угла.

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	— ^{7}	— ^{1}	
10	$\vert$ ^{3}	3 ^{6}	7 ^{5}	— ^{8}	
14	$\vert$ ^{6}	$\vert$ ^{1}	2 ^{6}	12 ^{3}	
v_j					**127**

Высчитаем стоимость полученного плана перевозок:

$$7 \cdot 2 + 3 \cdot 4 + 3 \cdot 6 + 7 \cdot 5 + 2 \cdot 6 + 12 \cdot 3 = 14 + 12 + 18 + 35 + 12 + 36 = 127.$$

Замечание. Для проверки рекомендуется на каждой итерации суммировать полученные перевозки по каждой строке и столбце – результат должен совпадать с запасом и потребностью в соответствующем складе и пункте распределения.

Полагаем $u_1 = 0$.

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	— ^{7}	— ^{1}	**0**

10	| ^{3}	3 ^{6}	7 ^{5}	— ^{8}	
14	| ^{6}	| ^{1}	2 ^{6}	12 ^{3}	
v_j	**2**				127

Тогда из условия $v_1 - u_1 = c_{11}$ для базисной клетки $x_{11} = 7 > 0$, получаем $v_1 = c_{11} + u_1 = 2 + 0 = 2$:

Из условия $v_2 - u_1 = c_{12}$ для базисной клетки $x_{12} = 3 > 0$, получаем $v_2 = c_{12} + u_1 = 4 + 0 = 4$:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	— ^{7}	— ^{1}	0
10	| ^{3}	3 ^{6}	7 ^{5}	— ^{8}	
14	| ^{6}	| ^{1}	2 ^{6}	12 ^{3}	
v_j	2	**4**			127

Из условия $v_2 - u_2 = c_{22}$ для базисной клетки $x_{22} = 3 > 0$, получаем $u_2 = v_2 - c_{22} = 4 - 6 = -2$ и так далее

$$v_3 - u_2 = c_{23}\ (x_{23} = 7 > 0) \ \Rightarrow\ v_3 = c_{23} + u_2 = 5 + (-2) = 3;$$

$$v_3 - u_2 = c_{23}\ (x_{23} = 7 > 0) \ \Rightarrow\ v_3 = c_{23} + u_2 = 5 + (-2) = 3;$$

$$v_3 - u_3 = c_{33}\ (x_{33} = 2 > 0) \ \Rightarrow\ u_3 = v_3 - c_{33} = 3 - 6 = -3;$$

$$v_4 - u_3 = c_{34}\ (x_{34} = 12 > 0) \ \Rightarrow\ v_4 = c_{34} + u_3 = 3 + (-3) = 0.$$

Запишем полученные значения в таблицу:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	— ^{7}	— ^{1}	0
10	| ^{3}	3 ^{6}	7 ^{5}	— ^{8}	**-2**
14	| ^{6}	| ^{1}	2 ^{6}	12 ^{3}	**-3**
v_j	2	4	**3**	**0**	127

Замечания.

1. В дальнейшем расчет потенциалов не будет подробно обсуждаться, а все результаты буду сразу записываться в таблицу.
2. Количества базисных клеток ($m + n - 1$) и их расположения достаточно, чтобы вычислить значения всех потенциалов.
3. В случае вырожденной задачи, базисная клетка может иметь нулевое значение ($x_{ij} = 0$). Этот нуль явно записывается в таблице.

Далее, для всех клеток рассчитаем значения

$$\Delta_{ij} = v_j - u_i - c_{ij}$$

и запишем их в правом нижнем углу каждой небазисной клетки $x_{ij} = 0$, для базисных клеток эти значения равны 0, по построению, и не требуют вычисления.

Напомним, что для выполнения условий потенциальности все значения Δ_{ij} должны быть меньше либо равны 0. Если есть положительные значения, то есть возможность улучшения стоимости, и, значит, имеющийся план перевозок не оптимален.

$$x_{21}: \ \Delta_{21} = v_1 - u_2 - c_{21} = 2 - (-2) - 3 = 1;$$

$$x_{31}: \ \Delta_{31} = v_1 - u_3 - c_{31} = 2 - (-3) - 6 = -1;$$

$$x_{32}: \ \Delta_{32} = v_2 - u_3 - c_{32} = 4 - (-3) - 1 = 6;$$

$$x_{13}: \ \Delta_{13} = v_3 - u_1 - c_{13} = 3 - 0 - 7 = -4;$$

$$x_{14}: \ \Delta_{14} = v_4 - u_1 - c_{14} = 0 - 0 - 1 = -1;$$

$$x_{24}: \ \Delta_{24} = v_4 - u_2 - c_{24} = 0 - (-2) - 8 = -6.$$

Два значения положительны и значит план не оптимален – потребуется как минимум еще одна итерация. Занесем, вычисленные значения в таблицу:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	2 7	4 3	7 — **-4**	1 — **-1**	0
10	3 | **1**	6 3	5 7	8 — **-6**	-2
14	6 | **-1**	1 | **6**	6 2	3 12	-3
v_j	2	4	3	0	127

Далее ищем клетку с наибольшим положительным значением Δ_{ij} – это x_{32} (в таблице значение подчеркнуто). Начиная с этой свободной клетки, строим

цикл все остальные клетки которого базисные. В общем случае, такой цикл определяется неоднозначно. В данном случае подойдет

$$x_{32} = 0, \quad x_{33} = 2, \quad x_{23} = 7, \quad x_{22} = 3.$$

Начиная с клетки x_{32}, в цикле расставим чередующиеся знаки «+» и «–», начиная с «+»:

$$x_{32}^{+} = 0, x_{33}^{-} = 2, x_{23}^{+} = 7, x_{22}^{-} = 3.$$

В таблице отметим это, поставив знаки в левых верхних углах соответствующих клеток цикла:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	— ^{7} $_{\mathbf{-4}}$	— ^{1} $_{\mathbf{-1}}$	0
10	$\vert$ ^{3} $_{\mathbf{1}}$	– 3 ^{6}	+ 7 ^{5}	— ^{8} $_{\mathbf{-6}}$	-2
14	$\vert$ ^{6} $_{\mathbf{-1}}$	+ $\vert$ ^{1} $_{\mathbf{6}}$	– **2** ^{6}	12 ^{3}	-3
v_j	2	4	3	0	127

Среди клеток цикла, помеченных минусами, выберем ту, что содержит минимальную перевозку. В нашем случае это клетка x_{32} = 2. Далее «сдвинем» эту перевозку по циклу, то есть для всех «–»-клеток вычитаем из имеющейся перевозки 2, а для всех «+»-клеток – прибавляем 2.

Замечания.

1. После сдвига по циклу, суммарная перевозка по клеткам цикла не должна измениться.

2. В дальнейших итерациях расчеты потенциалов, Δ_{ij} и циклов не будут расписываться, а результаты будут сразу записываться в таблицу.

3. После пересчета вдоль выбранного цикла получаем таблицу (по соглашению, в небазисных клетках нули не записываются):

Склады	Пункты назначения				u_i
	7	6	9	12	
10	7 ^{2}	3 ^{4}	^{7}	^{1}	
10	^{3}	1 ^{6}	9 ^{5}	^{8}	

14	6	1 2	6 ...	3 12	
v_j					**115**

Многоточие означает нулевое значение перевозки для вновь полученной свободной (небазисной) клетки, который мы, по соглашению, не записываем.

Стоимость полученного плана перевозок:

$$14 + 12 + 6 + 2 + 45 + 36 = 115.$$

После расчетов потенциалов и Δ_{ij}-х, проверки на потенциальность, получаем, что нам предстоит сдвинуть $x_{12} = 3$ вдоль цикла $x_{14}^{+} = 0$, $x_{12}^{-} = 3$, $x_{32}^{+} = 2$, $x_{34}^{-} = 12$:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	2 7	– 4 3	7 -4	+ 1 **5**	0
10	3 1	6 1	5 9	8 0	-2
14	6 -7	+ 1 2	6 -6	– 3 12	3
v_j	2	4	3	6	115

4. После сдвига по циклу, расчета новой стоимости перевозки, пересчета потенциалов и Δ_{ij}-х, проверки на потенциальность, получаем, что нам предстоит сдвинуть $x_{22} = 1$ вдоль цикла

$$x_{21}^{+} = 0, x_{22}^{-} = 1, x_{32}^{+} = 5, x_{34}^{-} = 9, x_{14}^{+} = 3, x_{11}^{-} = 7:$$

Склады	Пункты назначения				u_i
	7	6	9	12	
10	– 2 7	4 … -5	7 -9	+ 1 3	0
10	+ 3 **6**	– 6 1	5 9	8 0	-7
14	6 -2	+ 1 5	6 -6	– 3 9	-2
v_j	2	-1	-2	1	100

5. После сдвига по циклу, расчета новой стоимости перевозки и пересчета потенциалов и Δ_{ij}-х, получаем:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	2 6	4 -4	7 -9	1 4	0
10	3 1	6 … -4	5 9	8 -6	-1
14	6 -2	1 6	6 0	3 8	-2
v_j	2	1	4	1	94

Поскольку условия потенциальности

$$\Delta_{ij} \le 0, i = 1,\dots,n;\ j = 1,\dots,m$$

выполняются, полученный план перевозок является оптимальным.

Минимальная стоимость 94 достигается при следующем плане перевозок:

- с 1-го склада необходимо направить 6 ед. груза в 1-й пункт и 4 ед. в 4-й пункт;
- со 2-го склада необходимо направить 1 ед. груза в 1-й пункт и 9 ед. груза в 3-й пункт;
- с 3-го склада необходимо направить 6 ед. груза в 2-й пункт и 8 ед. в 4-й пункт.

II. ММЭ и метод потенциалов.

1. Получим начальный опорный план перевозок, используя метод минимального элемента.

Склады	Пункты назначения				u_i
	7	6	9	12	
10	2 —	4 —	7 —	1 **10**	
10	3 **7**	6 |	5 **3**	8 |	
14	6 |	1 **6**	6 **6**	3 **2**	
v_j					**94**

Высчитаем стоимость полученного плана перевозок:

$$21 + 6 + 15 + 36 + 10 + 6 = 94.$$

Рассчитаем потенциалы u_i, v_j и значения $\Delta_{ij} = v_j - u_i - c_{ij}$ для полученного плана перевозок:

Склады	Пункты назначения				u_i
	7	6	9	12	
10	2 — 0	4 — -5	7 — -3	1 10	0
10	3 7	6 | -6	5 3	8 | -6	-1
14	6 | -2	1 6	6 6	3 2	-2
v_i	2	-1	4	1	94

Поскольку для небазисных клеток ($x_{ij} = 0$) все значения $\Delta_{ij} \le 0$, а для базисных ($x_{ij} > 0$): $\Delta_{ij} = 0$, по построению, то условия потенциальности выполняются, значит начальный опорный план перевозок не может быть улучшен, то есть уже является оптимальным.

Ответ. Минимальная стоимость 94 достигается при следующих перевозках:

- с 1-го склада необходимо весь груз направить в 4-й пункт;
- со 2-го склада необходимо направить 7 ед. груза в 1-й пункт и 3 ед. груза в 3-й пункт;
- с 3-го склада необходимо направить 6 ед. груза в 2-й пункт, 6 ед. в 3-й пункт, 2 ед. в 4-й пункт.

Замечания.

1. Метод минимального элемента для задач малой размерности часто сразу дает оптимальное решение.

2. Сравнивая решения, полученные в I и II частях, видно, что методы могут давать различные (но эквивалентные, то есть имеющие одинаковую стоимость) оптимальные решения.

III. Найдем решение, используя онлайн-калькулятор для решения транспортной задачи. Ниже приведены выдержки из типичного компьютерного решения.

Стоимость доставки единицы груза из каждого пункта отправления в соответствующие пункты назначения задана матрицей тарифов

	1	2	3	4	Запасы
1	2	4	7	1	10
2	3	6	5	8	10
3	6	1	6	3	14
Потребности	7	6	9	12	

Проверим необходимое и достаточное условие разрешимости задачи.

$$\sum a = 10 + 10 + 14 = 34$$
$$\sum b = 7 + 6 + 9 + 12 = 34$$

Условие баланса соблюдается. Запасы равны потребностям. Следовательно, модель транспортной задачи является закрытой.

Минимальные затраты составят:

$$F(x) = 1\cdot10 + 3\cdot7 + 5\cdot3 + 1\cdot6 + 6\cdot6 + 3\cdot2 = 94.$$

Анализ оптимального плана.

- Из 1-го склада необходимо весь груз направить в 4-й магазин.
- Из 2-го склада необходимо груз направить в 1-й магазин (7), в 3-й магазин (3).
- Из 3-го склада необходимо груз направить в 2-й магазин (6), в 3-й магазин (6), в 4-й магазин (2).

ГЛАВА 4. ЭЛЕМЕНТЫ ТЕОРИИ ИГР

Основные понятия теории игр

Предметом изучения теории игр являются формализованные ситуации, для исхода которых определяющую роль играют конфликтующие и/или совместные действия.

Ситуация будет ***конфликтной***, если в ней действуют стороны, интересы которых полностью или частично противоположны. Стороны, принимающие участие в рассматриваемом конфликте, называются ***игроками***.

Примеры. Конфликтными будут, например, ситуации:

1) конкурентного рынка, где стратегиями поведения будут устанавливаемые продавцами цены;
2) военный конфликт, где стратегиями будут принимаемые сторонами решения;
3) обычные игровые ситуации некоторых карточных игр, в которых присутствует неопределенность.

Конфликт может проявиться не только в результате сознательных действий различных участников, но и как результат действия некоторых внешних сил. Такого рода ситуации называются ***«играми с природой»***.

Примеры. Играми с природой можно считать ситуации:

1) игры на биржевом рынке, где стратегией будет распределение инвестиций между различными ценными бумагами;
2) выбор площадей для посевов различных сельскохозяйственных культур, где другая «играющая» сторона – погодные условия.

Главное отличие «игр с природой» от антагонистических игр (где выигрыш одного из двух игроков означает ровно такой же по величине проигрыш другого) заключается в том, что при выборе стратегий можно не рассчитывать, что противник (природа) будет действовать так, чтобы причинить максимальный проигрыш, как в случаях игры с противником. Вместо этого считается, что природа действует просто случайным образом.

Теорией игр рассматриваются формализованные игры, где изначально фиксируются правила игры и выигрыши. Правила игры перечисляют допустимые действия (стратегии) игроков, направленные на достижение ими выигрыша.

Однозначное описание действий игрока в каждой из возможных ситуаций, в которой он должен сделать свой ход, называется ***стратегией игрока***.

Количественная оценка результатов хода называется ***платежом***, или ***выигрышем***, того или иного игрока. Соответствие между возможным набором возникающих в игре ситуаций, выбранной стратегией и соответствующими выигрышем игрока называется ***функцией выигрыша*** или ***платежной функцией***.

Замечание. Рассматриваемые формализованные игры исключают многие аспекты реальных конфликтов, включая психологические и ресурсные.

- Так, например, исключаются ситуации, когда один игрок намеренно играет слабее (в формализованной игре игрок всегда играет оптимально для себя), стараясь усыпить внимание противника, чтобы неожиданным ходом обернуть игру в свою пользу или спровоцировать противника на рискованные ходы (вариант «засады»).
- Также в реальных конфликтах, в частности, в военных или в бизнес-конкуренции, допустимые ходы противника не всегда известны заранее и составляют важный секрет, в то время как в формализованных играх, все допустимые ходы известны заранее.
- Частным случаем предыдущего пункта является случай, когда для осуществления хода необходимы ресурсы, которые могут быть истощены.

Классификация игр

Итак, формальное описание игры может включать:

1) перечень игроков (обозначаемых буквами или номерами);
2) список допустимых ходов каждого игрока (также обозначаемых буквами и/или номерами);
3) описание функции выигрыша каждого игрока.

Игры, соответственно, можно ***классифицировать*** по следующим видам:

1) по числу игроков (два, три, четыре, и так далее, или бесконечное число игроков),
2) по числу стратегий (***конечные*** или ***бесконечные игры***),
3) по свойствам функций выигрыша,
4) по возможности предварительных переговоров и взаимодействий между игроками в ходе игры.

Стратегии в конечных играх часто называют ***чистыми стратегиями*** (термин будет пояснен позже). В бесконечных играх игроки имеют бесконечное число возможных стратегий.

С точки зрения свойств платежных функций важным классом являются так называемые антагонистические игры, или игры с нулевой суммой:

Игра называется ***игрой с нулевой суммой***, если выигрыш одного игрока равен проигрышу другого.

Рассматриваются также игры, где возможны как конфликтные ходы, так и кооперативное поведение, или их сочетания.

Замечание. «Нулевая сумма» выигрыша (антагонистичность) считалась неотъемлемой чертой капиталистической конкуренции и вошли в идеологический шаблон «правильного» капитализма. Позднее пришедшее понимание, которое связывают с теоремой Нэша о равновесии, о том, что индивидуальные выигрыши в случае кооперативного поведения могут быть выше (например, в результате синергии), чем в случае чисто конфликтующего поведения, привели к определенному мировоззренческому перевороту в западном обществе и бизнес-сообществе, в частности. Так, например, основатель Microsoft Билл Гейтс называл открытое программное обеспечение противоречащим «американским ценностям» (надо полагать свободной конкуренции). В настоящее время Microsoft является активным участником разработки (contributor) открытой операционной системы Линукс, а также сотрудничает в разработке браузера Chromium со своим прямым конкурентом Google. Другим примером кооперативного поведения в конкурентной среде, на этот раз незаконным, могут считаться картельные соглашения между крупными игроками рынка об ограничении производства (по сути добровольном ограничении своей доли рынка) для создания искусственного дефицита и/или завышения цены продажи.

Матричные игры

Будем рассматривать конечные игры двух лиц с нулевой суммой.

Пусть есть два игрока, первый из которых (игрок A) имеет m возможных ***чистых стратегий***

$$A_1, A_2, \ldots, A_m,$$

а второй (игрок B) – n чистых стратегий

$$B_1, B_2, \ldots, B_n.$$

Пусть при выборе первым игроком A_i -й стратегии, а вторым игроком B_j -й стратегии первый игрок получает выигрыш, равный c_{ij}, а второй игрок эту величину проигрывает[3].

Величины c_{ij} удобно располагать в виде матрицы, строки которой соответствуют чистым стратегиям первого игрока, а столбцы – чистым стратегиям второго игрока.

[3] Заметим, что это не делает игру несимметричной для игроков. Если величина c_{ij} отрицательна, то первый игрок получает отрицательный выигрыш, т. е. проигрывает, а второй игрок, наоборот, при этом выигрывает.

$$\begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_m \end{array} \begin{array}{c} \begin{array}{cccc} B_1 & B_2 & \cdots & B_n \end{array} \\ \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \end{array}$$

Матрица $M = (c_{ij})$, $i = 1, \ldots m$; $j = 1, \ldots n$

$$M = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}$$

называется ***платежной матрицей***.

Основываясь на таком представлении конечных игр двух лиц с нулевой суммой, их называют ***матричными играми***.

Общую схему матричной игры поясним на следующем примере:

Пример. Два игрока A и B играют на деньги в матричную игру с платежной матрицей:

$$M = \begin{pmatrix} 2 & 4 & -1 \\ -3 & -5 & 2 \end{pmatrix}$$

Напомним, что строки соответствуют доступным стратегиям первого игрок, а столбцы – второго.

$$\begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 2 & 4 & -1 \\ -3 & -5 & 2 \end{pmatrix} \end{array}$$

Игра могла бы осуществляться по следующему протоколу (результат будет тот же, что и для матричной игры общего типа с этой же платежной матрицей).

Каждый из игроков имеет набор карточек с названиями допустимых для него ходов. Согласно имеющейся платежной матрице

у первого игрока имеются карточки $\boxed{A_1}$ и $\boxed{A_2}$

у второго игрока, соответственно, $\boxed{B_1}$, $\boxed{B_2}$, $\boxed{B_3}$.

Оба игрока независимо, не показывая друг другу, откладывают карточку, соответствующую выбираемому ходу. Эти карточки перевернутыми кладутся на середину и одновременно открываются (переворачиваются). Далее игроки совместно вычисляют по матрице свой выигрыш или проигрыш.

Например, если первый выложил карточку $\boxed{A_1}$, а второй – $\boxed{B_2}$, то выигрыш первого игрока составил 4 рубля, соответственно проигрыш второго тоже 4 рубля (в игре с нулевой суммой выигрыш одного игрока это проигрыш другого).

Если на столе оказались карточки $\boxed{A_2}$ и $\boxed{B_1}$, то первый игрок проиграл 3 рубля[4], а второй, соответственно, выиграл 3 рубля.

Далее карточки возвращаются игрокам и партия (раунд) повторяется. □

Вопрос о решении матричной игры заключается в том, чтобы предложить оптимальные стратегии для обоих игроков.

Оптимальной называется стратегия, которая при многократном повторении игры обеспечивает первому игроку максимально возможный выигрыш, а второму игроку минимально возможный проигрыш.

Итак, матричная игра характеризуется следующими позициями:

1) в ней участвует два игрока;
2) сколько выигрывает один, столько проигрывает другой игрок;
3) каждый игрок имеет конечное число ходов (чистых стратегий).

Максиминные и минимаксные стратегии

Естественно, для первого игрока выбирать стратегию, максимизирующую проигрыш второго игрока. Соответственно, для второго игрока, естественно, выбирать стратегию, минимизирущую выигрыш первого игрока.

На этом основаны принципы минимакса и максимина.

Обозначим

$$a_i = \min_j c_{ij}, \quad b_j = \max_i c_{ij}. \tag{4.1}$$

По смыслу, число a_i равно минимальному гарантированному выигрышу первого игрока при выборе стратегии A_i, какие бы действия не предпринимал второй игрок.

Тогда наибольший гарантированный выигрыш первого игрока равен

$$v_{\min} = \max_i a_i = \max_i \min_j c_{ij}. \tag{4.2}$$

[4] В платежной матрице записаны выигрыши первого игрока (проигрыши – второго), поэтому знак минус означает отрицательный выигрыш – проигрыш первого игрока.

Число $v_{\min}$ называется ***нижней ценой игры*** (или максимином), а стратегия, максимизирующая минимальный возможный выигрыш игрока A, т. е. стратегия A_i, для которой $v_{\min} = a_i$ – ***максиминной стратегией*** игрока A.

Аналогично, b_j представляет собой гарантированный уровень стратегии j, т. е. выбрав стратегию B_j, второй игрок имеет полную гарантию, что его проигрыш независимо от действий первого игрока будет не больше чем b_j. Тогда наименьший гарантированный проигрыш

$$v_{\max} = \min_j b_j = \min_j \max_i c_{ij}. \tag{4.3}$$

Стратегия B_j, которая минимизирует максимальные возможные потери игрока B, называется его минимаксной стратегией, а число $v_{\max}$ – ***верхней ценой*** игры (или ***минимаксом***).

Теорема. В любой матричной игре нижняя цена игры не превосходит верхней цены:

$$v_{\min} \le v_{\max}.$$

Доказательство. Действительно, при любых i, j выполняется условие

$$a_i = \min c_{ik} \le c_{ij} \le \max c_{kj} = b_j,$$

откуда

$$v_{\min} = \max a_i \le \min b_j = v_{\max}. \qquad \square$$

Если нижняя и верхняя цена игры совпадает, то число

$$v = v_{\min} = v_{\max}$$

называется ***ценой игры***.

Вычисление нижней и верхней цен игры рассмотрим на примере.

Пример 4.1. Найти нижнюю и верхнюю цену матричной игры заданной платежной матрицей:

а) $\begin{pmatrix} 2 & 7 & 1 \\ 6 & 2 & 4 \\ 5 & 6 & 7 \end{pmatrix}$ б) $\begin{pmatrix} 8 & 7 & 1 & 7 \\ 7 & 8 & 5 & 6 \\ 8 & 2 & 4 & 3 \end{pmatrix}$

Решение. а) Для удобства записи результатов вычислений? запишем платежную матрицу в виде таблицы, добавим столбец для строчных минимумов a_i и строку для столбцовых максимумов b_j. Наибольшее из чисел a_i, будет нижней ценой, а наименьшее из чисел b_j – верхней ценой.

	B_1	B_2	B_3	$a_i = \min c_{ij}$
A_1	2	7	1	1
A_2	6	2	4	2
A_3	5	6	7	5
$b_j = \max c_{ij}$	6	7	7	6 5

В данной игре $v_{\min} = 5 < v_{\max} = 6$.

б) Аналогично

	B_1	B_2	B_3	B_4	$a_i = \min c_{ij}$
A_1	8	7	1	7	1
A_2	7	8	5	6	5
A_3	8	2	4	3	2
$b_j = \max c_{ij}$	8	8	5	7	5 \ 5

В данной игре $v_{\min} = 5 = v_{\max} = 5$.

В этой игре выбор ситуации (2, 3), первой компонентой которой является максиминная стратегия игрока A, а второй – минимаксная стратегия игрока B, устойчив, так как ни одному из игроков невыгодно одностороннее отклонение от нее. □

Существование цены для матричной игры тесно связано с наличием устойчивых ситуаций в матричной игре.

Пусть матричная игра имеет платежную матрицу $M=(c_{ij})$. Комбинация стратегий (k, l) называется ***седловой точкой*** в этой игре, если при всех i, j выполняется двойное неравенство:

$$c_{il} \le c_{kl} \le c_{kj}.$$

Другими словами, комбинация (k, l) – ***седловая точка игры*** тогда и только тогда, когда элемент c_{kl}, является ***одновременно наименьшим элементом своей строки и наибольшим элементом своего столбца*** (см. рис. 4.2).

Ясно, что устойчивые в матричной игре ситуации – это и есть ее седловые точки: одностороннее отклонение от седловой точки невыгодно ни одному из игроков.

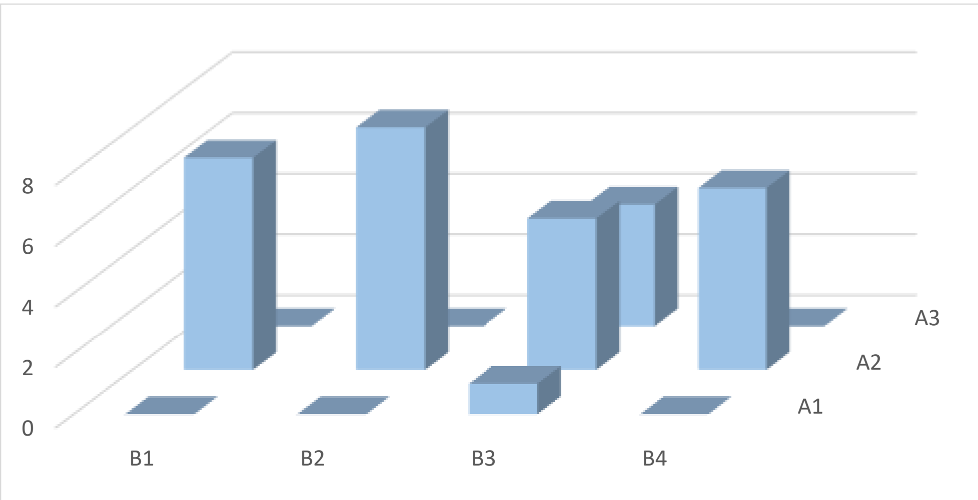


Рис. 4.2. Седловая точка примера 1 б). Выигрыш первого игрока в седловой точке не больше всех остальных выигрышей в той же строке A_2, и не меньше всех остальных выигрышей в этом же столбце B_3. Стоимости, не стоящие в этих строке и столбце, для наглядности не отображаются.

***Теорема (о седловой точке*).**

1. Пусть (k, l) – седловая точка матричной игры. Тогда:
   a) A_k является максиминной стратегией первого игрока;
   b) B_l является минимаксной стратегией второго игрока;
   c) игра имеет цену, причем равную исходу c_{kl} в седловой точке.
2. Пусть игра имеет цену v. Тогда любая ситуация (k, l), где A_k – максиминная стратегия игрока A, а B_l – минимаксная стратегия игрока B, является седловой точкой игры.

Теорема о седловой точке имеет два простых, но важных следствия.

Следствия.

1. В матричной игре наличие седловой точки эквивалентно наличию цены игры.
2. Исходы во всех седловых точках матричной игры совпадают между собой и совпадают с ценой игры.

Замечание. Таким образом, если матричная игра имеет цену, то в ней реализуются два принципа оптимальности:

- принцип оптимальности стратегий (в качестве оптимальных ситуаций выступают максиминные стратегии первого игрока и минимаксные стратегии второго игрока);
- принцип оптимальности ситуаций (в качестве оптимальных ситуаций выступают седловые точки).

При этом оба принципа оптимальности согласованы между собой:

если игроки выбирают оптимальные стратегии, то возникающая ситуация является оптимальной (седловой точкой);

и, наоборот,

компонентами любой седловой точки служат оптимальные стратегии игроков.

Кроме того, исход в любой седловой точке равен цене игры.

В общем случае, однако, не всякая матричная игра имеет цену и поэтому не может быть решена таким простым способом. Однако был найден подход, который позволяет в рамках матричной игры обеспечить цену игры в некотором обобщенном смысле. Этот метод, состоящий в переходе к ***смешанным стратегиям***, впервые был обоснован в теории игр фон Нейманом.

Смешанные расширения матричных игр

Если матричная игра не имеет седловой точки, то у нее нет решения в чистых стратегиях. В этом случае ее решение нужно искать в так называемых смешанных стратегиях, которые определяются следующим образом.

Предположим, что игра является многоходовой, т. е. состоит из многих партий. Тогда каждый игрок будет в разных партиях применять свои разные чистые стратегии, причем одни из них чаще, а другие реже.

Смешанную стратегию первого игрока описывает вектор вероятностей

$$\overline{p} = (p_1, \ldots, p_m),$$

где p_i это вероятность применения первым игроком его i-й чистой стратегии в многоходовой игре.

Из определения вероятности следует, что компоненты вектора p удовлетворяют следующим соотношениям:

$$p_1 + \ldots + p_m = 1,\ 0 \le p_1, \ldots, p_m \le 1$$

Аналогично определяется смешанная стратегия второго игрока – через вектор

$$\overline{q} = (q_1, \ldots, q_n),$$

где q_j это вероятность применения вторым игроком его j-й чистой стратегии в многоходовой игре. Как и в случае смешанной стратегии первого игрока, компоненты вектора $\overline{q}$ удовлетворяют соотношениям

$$q_1 + \ldots + q_n = 1,\ 0 \le q_1, \ldots\ q_n \le 1$$

Чистые стратегии игроков являются частными случаями смешанных стратегий. Например, i-я чистая стратегия игрока может быть представлена в виде такой смешанной стратегии, у которой i-я компонента равна единице, а остальные – нулю, например, A_1 соответствует вектор (1, 0, …, 0).

Если игроки используют свои смешанные стратегии, то функция

$$E(\overline{p}, \overline{q}) = \sum_{i,j} c_{ij} p_i q_j \tag{4.4}$$

называется ***платежной***[5]или ***функцией выигрышей*** смешанного расширения матричной игры.

Оптимальная смешанная стратегия первого игрока $\overline{p}^*$ – это решение следующей задачи:

[5] По сути, это математическое ожидание выигрыша первого игрока (или проигрыша второго). Даже для обозначения используется та же буква – *E*, от *expectancy* – ожидание (англ.)

$$a^* = \max_{\bar{p}} \min_{\bar{q}} E(\bar{p},\bar{q}) \qquad (4.5)$$

Оптимальная смешанная стратегия второго игрока $\bar{q}^*$ – это решение следующей задачи:

$$b^* = \min_{\bar{q}} \max_{\bar{p}} E(\bar{p},\bar{q}) \qquad (4.6)$$

Если $\bar{p}^* = (p_1^*, \ldots, p_m^*)$ и $\bar{q} = (q_1^*, \ldots, q_n^*)$ – оптимальные смешанные стратегии игроков, то число

$$v = E(\bar{p}^*,\bar{q}^*) = \sum_{i,j} c_{ij} p_i^* q_j^* \qquad (4.7)$$

называется ***ценой игры***.

Теорема (фон Неймана). Для любой матричной игры верно равенство:

$$\max_{\bar{p}} \min_{\bar{q}} E(\bar{p},\bar{q}) = \min_{\bar{q}} \max_{\bar{p}} E(\bar{p},\bar{q}). \qquad (4.8)$$

Другими словами, в смешанном расширении любой матричной игры нижняя и верхняя цена игры совпадает.

Замечание. Формула в теореме фон Неймана похожа на формулу цены игры, но здесь максимум и минимум берутся уже по ***бесконечному*** множеству возможных смешанных стратегий, в то время как в исходной постановке матричной игры максимум и минимум брался по конечному множеству пар чистых стратегий. Таким образом, исходно матричная игра относится к классу конечных игр, а её смешанное расширение уже к классу бесконечных игр.

Существует несколько простых правил, позволяющих упростить нахождение решения матричной игры. Одно из них – это правило отбрасывания ***мажорируемых (доминируемых) стратегий***:

При решении матричной игры в смешанных стратегиях доминируемые стратегии игроков могут быть отброшены.

Здесь, чистая стратегия ***первого*** игрока $A_i = (c_{i1}, c_{i2}, \ldots, c_{im})$ ***доминируема***, если существует другая его стратегия $A_j = (c_{j1}, c_{j2}, \ldots, c_{jm})$, такая что соответствующие выигрыши больше или равны, чем выигрыши при использовании A_i:

$$c_{i1} \le c_{j1},\ \ c_{i2} \le c_{j2},\ \ldots,\ c_{im} \le c_{jm}.$$

Понятно, что честно играющий (стремящийся к выигрышу) первый игрок не будет пользоваться стратегий, приносящей при тех же контрходах второго игрока меньше прибыли.

Поскольку платежная матрица сформулирована в терминах выигрыша первого игрока, для второго игрока определение доминируемости отличается:

Чистая стратегия ***второго*** игрока $B_i = (c_{1i}, c_{2i}, \ldots, c_{ni})$ ***доминируема***, если существует другая его стратегия $B_j = (c_{1j}, c_{2j}, \ldots, c_{nj})$, такая что соответствующие выигрыши меньше или равны, чем выигрыши при использовании стратегии B_i:

$$c_{1j} \le c_{1i},\ \ c_{2j} \le c_{2i},\ \ldots,\ c_{nj} \le c_{ni}.$$

Применение этого правила позволяет иногда существенно уменьшить размерность матрицы игры.

Пример 4.2. Уменьшить размерность матрицы игры, выбрасывая доминируемые стратегии

$$A = \begin{pmatrix} 3 & 1 & 3 & 4 \\ 4 & 2 & 3 & 5 \\ 3 & 1 & 2 & 3 \\ 2 & 1 & 1 & 6 \end{pmatrix}$$

Решение. Вычеркнем последовательно заведомо худшие стратегии у первого и второго игроков.

$$A = \left(\begin{array}{c|ccc|} 3 & \cancel{1} & \cancel{3} & \cancel{4} \\ 4 & 2 & 3 & 5 \\ 3 & 3 & 2 & 3 \\ 2 & 1 & \cancel{3} & 6 \end{array}\right).$$

Остается матрица A' размерности 2x2:

$$A' = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix},$$

которую уже можно решить ***графическим методом***.

Графическое решение матричных игр

Разберем графический метод решения матричных игр на следующем примере.

Пример 4.3. Найдем оптимальные стратегии игроков в игре, заданной платежной матрицей

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

Решение.

1. Сначала проверим, есть ли в данной игре седловая точка.

	B_1	B_2	B_3	B_4	a_i
A_1	6	5	7	3	3
A_2	3	2	1	4	1

b_j	6	5	7	4	4 3

Поскольку нижняя цена игры не совпадает с верхней:

$$v_{\min} = \max\{3, 1\} = 3 \leq \min\{4, 5, 7, 6\} = 4 = v_{\max},$$

то седловой точки у данной игры нет и решение существует только в смешанных стратегиях.

2. Графически можно решить только те матричные игры, в которых хотя бы у одного из игроков есть лишь две чистые стратегии. Задача именно этого игрока и решается графически.

В задаче у первого игрока две чистых стратегии, а у второго – четыре, поэтому будем решать графически задачу первого игрока.

Смешанная стратегия первого игрока задается вектором $\bar{p} = (p_1, p_2)$. Отложим на горизонтальной прямой отрезок единичной длины. Каждой точке x этого отрезка будем сопоставлять смешанную стратегия первого игрока по следующему правилу (рис. 4.3):

расстояние от точки до правого конца отрезка задает величину p_1, а расстояние до левого его конца – величину p_2.

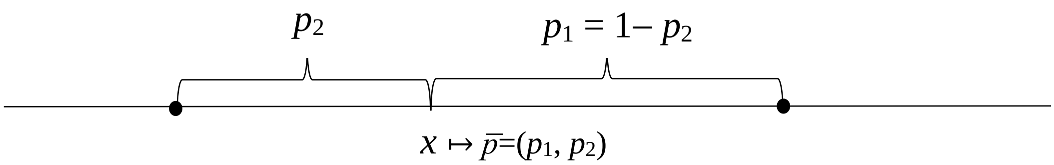


Рис. 4.3. Представление смешанных стратегий первого игрока точками отрезка [0, 1]

Для определенных таким образом величин p_1 и p_2 выполняются соотношения

$$p_1 + p_2 = 1;\ \ p_1\ , p_2 > 0,$$

поэтому такой вектор $p = (p_1, p_2)$ задает смешанные стратегии первого игрока.

Как видно из построения, точка 0, например, задает вектор (1, 0), т. е. первую чистую стратегию первого игрока, а точка 1 задает вектор (0, 1), т. е. вторую чистую стратегию первого игрока.

Далее, через концы единичного отрезка проводятся вертикальные линии. На этих линиях откладываются выигрыши первого игрока при применении вторым игроком его различных чистых стратегий. При этом выигрыши в случае применения первым игроком его первой чистой стратегии располагаются на левой вертикальной линии, а соответствующие второй чистой стратегии первого игрока – на правой вертикали. Точки левой и правой вертикали, соответствующие одной и той же чистой стратегии второго игрока, соединяются отрезками (рис. 4.4).

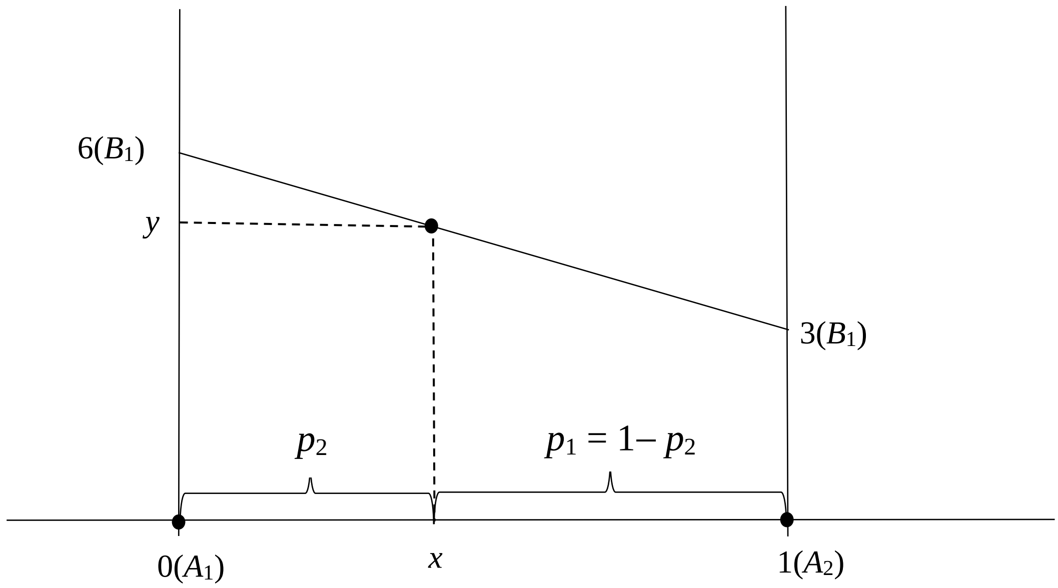


Рис. 4.4. Вспомогательные построения при графическом решении матричной игры

Таким образом, на рис. 4.4 изображены выигрыши первого игрока при применении вторым игроком первой чистой стратегии B_1.

Любая точка P этого отрезка с координатами (x, y) показывает, что если первый игрок будет применять свою смешанную стратегию (x, $1 - x$), а второй игрок – свою первую чистую стратегию, то средний выигрыш первого игрока будет равен y.

Аналогичные построения выполняются для остальных чистых стратегий второго игрока (рис. 4.5).

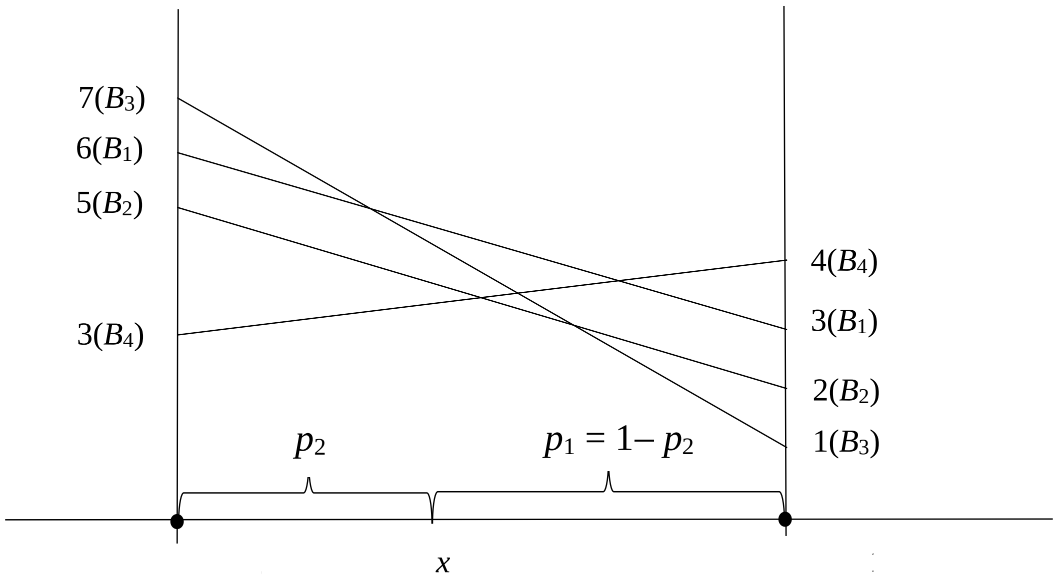


Рис. 4.5 Линии выигрышей первого игрока при применении вторым игроком чистых стратегий.

Выделим жирным ломаную линию, соответствующую нижней границе выигрыша первого игрока, т.е. дающую его средний гарантированный выигрыш. В точке P находится наибольший гарантированный выигрыш первого игрока (так как P – наивысшая точка ломаной). Эта точка является пересечением отрезков, соответствующих второй и четвертой чистым стратегиям второго игрока.

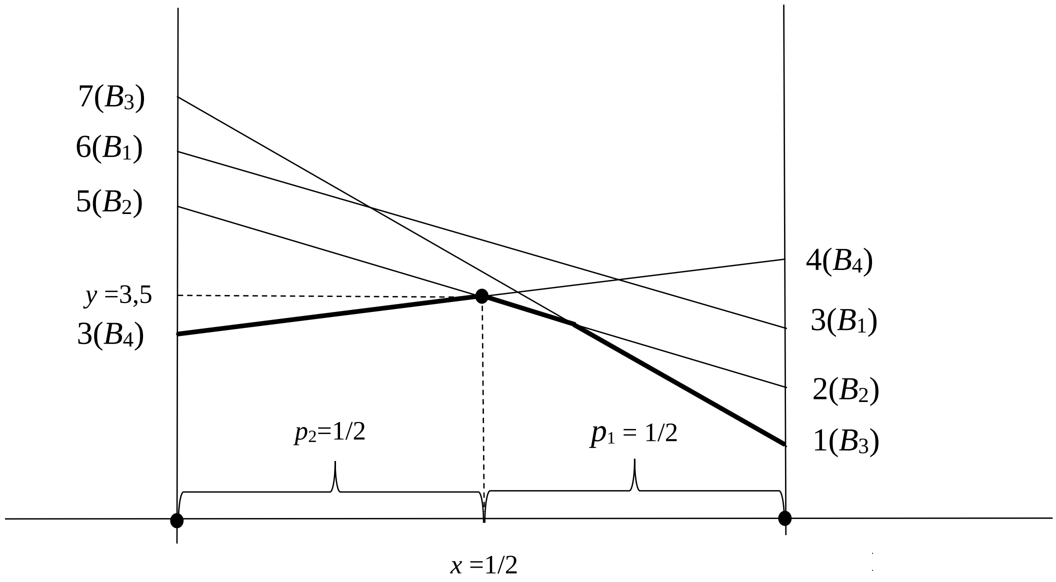


Рис. 4.6. Графическое решение задачи первого игрока.

Эти стратегии называются ***активными***. Второй игрок будет использовать их в своей оптимальной смешанной стратегии с ненулевой вероятностью. Отрезки, соответствующие первой и третьей чистым стратегиям второго игрока, не проходят через точку P, поэтому эти стратегии в оптимальную смешанную стратегию второго игрока войдут с нулевыми вероятностями, так как их реализация приведет к большему проигрышу второго игрока. Такие стратегии называют ***пассивными***.

Графическое решение задачи состоит в нахождении уравнений прямых, соответствующих активным стратегиям, и определении точки их пересечения.

Уравнения прямых в этом случае можно найти следующим способом:

$$y = y_0 \cdot (1- x) + y_1 \cdot x,$$

где y_0 – высота прямой над 0, а y_1 – высота над 1.

Найдем уравнение линии выигрышей первого игрока для второй стратегии второго игрока:

$$L_2:\ y = 5 \cdot (1- x) + 2 \cdot x = 5 - 3x.$$

Уравнение второй линии (для четвертой стратегии):

$$L_4:\ y = 3 \cdot (1-x) + 4 \cdot x = 3 + x.$$

Приравняем эти уравнения и решим:

$$5 - 3x = 3 + x \ \Rightarrow\ 4x = 2 \ \Rightarrow\ x = 1/2\ .$$

Таким образом, оптимальная смешанная стратегия первого игрока будет

$$\bar{p}^* = (1/2,\ 1/2)$$

или, в более развернутом виде:

$$S_A = \begin{pmatrix} A_1 & A_2 \\ 1/2 & 1/2 \end{pmatrix}$$

Для того чтобы найти цену игры, подставим найденное значение x в любое из уравнений L_2 или L_4:

$$y = 3 + 1/2 = 3{,}5.$$

Для нахождения оптимальной смешанной стратегии второго игрока, учитываем, что активными являются B_2 и B_4.

В общем случае, если активные стратегии игрока B имеют номера k и l соответственно, то общая формула расчета вероятности будет иметь вид

$$q_k = \frac{c_{2l} - c_{1l}}{(c_{2l} - c_{1l}) - (c_{2k} - c_{1k})}$$

В нашем случае,

$$q_2 = \frac{c_{24} - c_{14}}{(c_{24} - c_{14}) - (c_{22} - c_{12})} = \frac{2-5}{(2-5)-(4-3)} = \frac{-3}{-3-1} = \frac{3}{4}\,.$$

Следовательно, $q_1 = 1 - q_2 = 1 - 3/4 = 1/4$.

В итоге, оптимальная смешанная стратегия второго игрока имеет вид

$$\bar{q}^* = (0,\ 1/4,\ 0,\ 3/4),$$

или, если расписать подробнее:

$$S_B = \begin{pmatrix} B_1 & B_2 & B_3 & B_4 \\ 0 & 1/4 & 0 & 3/4 \end{pmatrix}.$$

Для проверки вычислим цену игры (она должна совпадать с выигрышами первого и второго игроков). Для этого запишем по матрице стоимости C общую функцию выигрыша для этой игры:

$$E(\bar{p},\bar{q}) = 6p_1q_1 + 5p_1q_2 + 7p_1q_3 + 3p_1q_4 + 3p_2q_1 + 2p_2q_2 + p_2q_3 + 4p_2q_4\,.$$

Тогда для найденных оптимальных смешанных стратегий:

$$v = E(\overline{p}^*, \overline{q}^*) = 5\cdot\frac{1}{2}\cdot\frac{1}{4} + 3\cdot\frac{1}{2}\cdot\frac{3}{4} + 2\cdot\frac{1}{2}\cdot\frac{1}{4} + 4\cdot\frac{1}{2}\cdot\frac{3}{4} = \frac{28}{8} = \frac{7}{2} = 3{,}5.$$

Замечание. Те же самые значения, $x = 1/2$ и $v = 3{,}5$, можно было приближенно определить по рис. 4.6. Для точности, построения лучше проводить на листе в клеточку или миллиметровой бумаге.

Ответ: $\overline{p}^* = (1/2,\ 1/2)$, $\overline{q}^* = (0,\ 1/4,\ 0,\ 3/4)$, $v = 3{,}5$. □

Контрольные вопросы

1. Перечислите рассмотренные виды игр.
2. Дайте определение матричной игры.
3. Какие игры называются «играми с природой»?
4. Чем отличаются «игры с природой» от «игр с противником»?
5. Является ли матричная игра игрой с нулевой суммой? Как это отражено в платежной матрице?
6. Что такое платежная матрица?
7. Платежная матрица и матрица выигрышей, одно ли это и то же?
8. Что такое седловая точка игры?
9. Что такое цена игры и всегда ли она существует?
10. Какая стратегия называется максиминной?
11. В случае существования седловой точки игры, максиминная стратегия является оптимальной стратегий какого игрока: первого или второго?
12. Какая стратегия называется минимаксной?
13. В случае существования седловой точки игры, минимаксная стратегия является оптимальной стратегий какого игрока: первого или второго?
14. Что такое нижняя и верхняя цены игры?
15. Чему равна цена игры, если игра имеет седловую точку?
16. Как связана цена игры с верхней и нижней ценой игры?
17. Что такое *смешанная* стратегия?
18. Чем отличаются смешанные стратегии от чистых стратегий (ходов)?
19. Матричная игра является конечной или бесконечной игрой?
20. Смешанное расширение матричной игры является конечной или бесконечной игрой?
21. Как работает метод удаления доминируемых стратегий?
22. Какие матричные игры можно решать графическим способом?
23. Какие стратегии называются активными
24. Как определяются смешанные оптимальные стратегии?

ГЛАВА 5. ОСНОВЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Системы массового обслуживания

Основные понятия теории массового обслуживания

На практике, часто появляется необходимость моделировании работы систем или объектов многоразового использования. Возникающие при этом процессы получили название процессов обслуживания, а моделируемые системы – систем массового обслуживания (СМО). Помимо систем, включающих в себя большое количество однотипных элементов, вроде систем сигнализации, вычислительных кластеров или серверов, дата-центров; примерами СМО также являются логистические центры, системы связи, ремонтные службы, диспетчерские и колл-центры, склады и т.п.

Моделирование СМО предполагает, что система состоит из определенного числа обслуживающих единиц (элементов, приборов, устройств, пунктов, станций), которые называются ***каналами обслуживания***. Каналами могут быть линии связи, рабочие точки, вычислительные машины, диспетчеры и др. По числу каналов СМО подразделяют на многоканальные и, как частный случай, одноканальные.

Заявки поступают в СМО обычно не регулярно, а случайно, образуя так называемый ***случайный поток заявок*** (требований). Обслуживание каждой заявки, также требует какого-то промежутка времени, вообще говоря, случайной продолжительности. Случайный характер потока заявок и времени обслуживания приводит к неравномерной загрузке СМО: в какие-то периоды времени набирается слишком большое количество заявок (они либо становятся в очередь, либо покидают СМО необслуженными), в другие же периоды СМО работает недогруженная или простаивает.

Предметом теории массового обслуживания (ТМО) является построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, характер потока заявок и т.п.) с показателями эффективности, описывающими способность справляться с потоками заявок.

При формулировании задачи ТМО синтезируется система массового обслуживания: формируется ее структура и указываются вероятностные характеристики, управляющие ее поведением. Далее в задаче ТМО определяются показатели обслуживания, подлежащие определению, например, время ожидания, параметры очереди, интервалы занятости и другие параметры. Структура, вероятностные характеристики системы массового обслуживания и ее параметры имеют непосредственную связь с входными потоками событий.

Задача исследования СМО сводится к необходимости определения оптимального состава и параметров для обеспечения необходимого качества обслуживания.

СМО можно представить в виде следующей структуры (рис. 5.1):

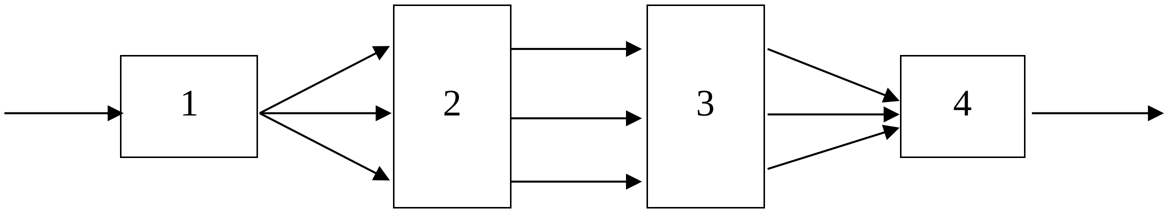


Рис. 5.1. Общая структура СМО.

где: 1) входной поток заявок (событий);

2) очередь (если есть);

3) каналы обслуживания, обрабатывающие поступившие заявки;

4) выходной поток, обработанных заявок, покидающих систему.

Более подробно, описывая составные части:

Входной поток – это совокупность заявок, нуждающихся в обслуживании (события). Примерами входных потоков являются поток запросов, поступающий на обработку на сервер, поток заявок в ремонтную службу, поток прибывающих транспортных средств, поток вызовов на диспетчерский пульт, сообщения о пожарах, отказы датчиков и т.д.

Очередь – это запоминающее устройство или элемент, фиксирующий время поступления и непосредственно сам факт заявки.

Каналы обслуживания – операторы, обслуживающие заявки.

Выходной поток – информация о об обработанных (или необработанных) заявках.

В ТМО расчеты часто производятся исходя из предположения, что входной поток является ***пуассоновским***, то есть задается ***формулой Пуассона***:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{5.1}$$

где

- λ – плотность потока заявок, то есть среднее число заявок, поступающих в систему за единицу времени (обычно берется значение, усредненное за какой-то достаточно большой промежуток времени);
- $P_k(t)$ – вероятность поступления ровно k требований за промежуток времени t.

Пуассоновский поток обладает следующими свойствам:

- ***стационарностью***, то есть обладает неизменными во времени вероятностными характеристикам;
- ***отсутствием последействия***, то есть вероятность поступление новой заявки не зависит от предыдущих поступлений;
- ***ординарностью***, то в каждый момент времени системы несколько заявок не могут поступить одновременно.

По сути дела, пуассоновский поток, обладает характеристиками потока абстрактной информации и является упрощающей моделью, удобной для проведения расчетов.

Классификация систем массового обслуживания

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 покидает сразу систему, а становится в очередь и ожидает обслуживания. Именно такие системы, называемые «СМО с очередью», всё чаще встречаются на практике и, в частности, теперь реализуются и в телефонной мобильной связи.

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 – «СМО с нетерпеливыми заявками». При анализе СМО учитывается также и «дисциплина обслуживания» – заявки могут обслуживаться либо в порядке поступления, либо с приоритетами или случайном порядке. Приоритет может быть, как ***абсолютным*** – когда заявка с более высоким приоритетом немедленно «вытесняет» обслуживаемую заявку с более низшим приоритетом, так и ***относительным*** – когда начатое обслуживание доводится до конца, а заявка с более высоким приоритетом имеет лишь право на следующее место в очереди.

Можно рассматривать также СМО с многофазовым обслуживанием, состоящими из нескольких последовательных этапов – ***фаз***.

Кроме предыдущих признаков, СМО делят также на следующие два класса: ***открытые*** и ***замкнутые*** **(*закрытые*)**. В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии находится сама СМО (сколько каналов занято, таким образом заявки как бы поступают из внешней среды, с неограниченным источником заявок). В замкнутой СМО поток заявок зависит от состояния СМО, например, при замене неисправных датчиков, чем больше их заменено или в работе, тем меньше поток заявок – заявки поступают из ограниченного (замкнутого) источника, являющегося частью системы.

Организация работы СМО может производиться под разные целевые критерии, например, с точки собственника СМО имеющиеся каналы должны быть максимально нагружены, даже в ущерб времени обслуживания. С точки зрения пользователя, наоборот, желательно времени обслуживания и уменьшение очереди. В любо случае, при организации работы СМО решается оптимизационная задача об наименьшем числе каналов с учетом заданных критериев эффективности обслуживания.

Простейшие системы массового обслуживания и их характеристики

Одноканальная система с отказами

Имеется один канал, на который поступает поток заявок с интенсивностью λ. Поток обслуживаний имеет интенсивность μ. Найти предельные вероятности состояний системы и показатели ее эффективности.

Система S (СМО) имеет два состояния: S_0 – канал свободен, S_1 – канал занят. Размеченный граф состояний представлен на рис. 5.2.

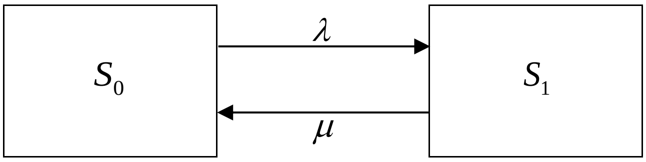


Рис. 5.2. Граф состояний одноканальной системы с отказами.

В предельном стационарном режиме система алгебраических уравнений для вероятностей состояний $p_0 = P(S_0)$ и $p_1 = P(S_1)$ имеет вид (потоки на входы в состояния равны потокам на выходах из них):

$$\begin{cases} \lambda p_0 = \mu p_1 \\ \mu p_1 = \lambda p_0 \end{cases}, \tag{5.2}$$

т.е. система вырождается в одно уравнение.

Учитывая условие $p_0 + p_1 = 1$, найдем предельные вероятности состояний

$$p_0 = \frac{\mu}{\lambda + \mu}, \qquad p_1 = \frac{\lambda}{\lambda + \mu}, \tag{5.3}$$

которые выражают среднее относительное время пребывания системы в состоянии S_0 (когда канал свободен) и S_1 (когда канал занят), т.е. определяют соответственно ***относительную пропускную способность*** Q системы и ***вероятность отказа*** $P_{отк}$:

$$Q = \frac{\mu}{\lambda + \mu}, \qquad P_{отк} = \frac{\lambda}{\lambda + \mu}. \tag{5.4}$$

Абсолютную пропускную способность найдем, умножив относительную пропускную способность Q на интенсивность потока заявок λ:

$$A = \lambda Q = \frac{\lambda\mu}{\lambda + \mu}. \tag{5.5}$$

Для удобства сведем рассмотренные характеристики одноканальной системы с отказами в одну таблицу (табл. 5.1).

Таблица 5.1. Сводная таблица характеристик одноканальной системы с отказами

Одноканальная система с отказами	
Характеристика системы	Формула
Вероятность обслуживания	$p_0 = \frac{\mu}{\lambda+\mu}$
Вероятность отказа в обслуживании	$p_1 = \frac{\lambda}{\lambda+\mu}$
Относительная пропускная способность	$p_0 = \frac{\mu}{\lambda+\mu}$
Абсолютная пропускная способность канала	$A = \lambda \cdot p_0 = \frac{\lambda \cdot \mu}{\lambda+\mu}$

Многоканальная система с отказами

Рассмотрим, так называемую, ***классическую задачу Эрланга***: имеется n каналов, на которые поступает поток заявок с интенсивностью λ. Поток обслуживаний каждого канала имеет интенсивность μ. Найти предельные вероятности состояний системы и показатели ее эффективности.

Система S (СМО) имеет следующие состояния (нумеруем их по числу заявок, находящихся в системе): $S_0, S_1, S_2, ..., S_k, \ldots, S_n$, где S_k – состояние системы, когда в ней находится k заявок, т.е. занято k каналов. Граф состояний СМО представлен на рис. 5.3.

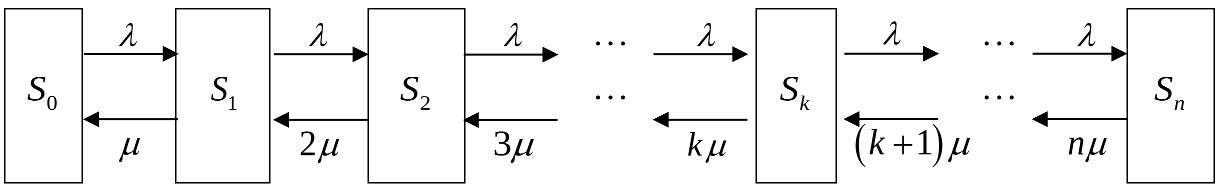


Рис. 5.3. Граф состояний многоканальной системы без очереди.

Поток заявок последовательно переводит систему из любого левого состояния в соседнее правое с одной и той же интенсивностью λ. Интенсивность же потока обслуживании, переводящих систему из любого правого состояния в соседнее левое, постоянно меняется в зависимости от состояния. Действительно, если СМО находится в состоянии S_2 (два канала заняты), то она может перейти в состояние S_1 (один канал занят), когда закончит обслуживание либо первый, либо второй канал, т.е. суммарная интенсивность их потоков обслуживаний будет 2μ. Аналогично суммарный поток обслуживании, переводящий СМО из состояния S_3 (три канала заняты) в S_2, будет иметь интенсивность 3μ, т.е. может освободиться любой из трех каналов, и т.д.

Запишем формулу для предельной вероятности состояния S_0:

$$p_0 = \left[1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \ldots + \frac{\lambda^k}{k!\mu^k} + \ldots + \frac{\lambda^n}{n!\mu^n}\right]^{-1}.$$

Величина $\rho = \frac{\lambda}{\mu}$ называется ***приведенной интенсивностью потока заявок*** или ***интенсивностью нагрузки канала***. Она выражает среднее число заявок, приходящих за среднее время обслуживания одной заявки. Теперь

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^k}{k!} + \ldots + \frac{\rho^n}{n!}\right]^{-1}, \tag{5.6}$$

$$p_1 = \frac{\rho}{1!} \cdot p_0,\ p_2 = \frac{\rho^2}{2!} \cdot p_0,\ \ldots,\ p_k = \frac{\rho^k}{k!} \cdot p_0,\ p_n = \frac{\rho^n}{n!} \cdot p_0. \tag{5.7}$$

Вероятность отказа этой СМО есть предельная вероятность того, что все n каналов системы будут заняты, т.е. $P_{отк} = \frac{\rho^n}{n!} \cdot p_0$.

Относительная пропускная способность – вероятность того, что заявка будет обслужена:

$$Q = 1 - P_{отк} = 1 - \frac{\rho^n}{n!} \cdot p_0. \tag{5.8}$$

Абсолютная пропускная способность:

$$A = \lambda Q = \lambda \cdot \left(1 - \frac{\rho^n}{n!} \cdot p_0\right). \tag{5.9}$$

Среднее число занятых каналов, т.е. математическое ожидание числа занятых каналов:

$$\bar{k} = \sum_{k=0}^{n} k p_k. \tag{5.10}$$

Однако среднее число занятых каналов можно найти проще, если учесть, что абсолютная пропускная способность системы A есть не что иное, как интенсивность потока обслуженных системой заявок (в единицу времени). Так как каждый занятый канал обслуживает в среднем μ заявок (в единицу времени), то среднее число занятых каналов $\bar{k} = \frac{A}{\mu}$ или

$$\bar{k} = \rho \cdot \left(1 - \frac{\rho^n}{n!} \cdot p_0\right). \tag{5.11}$$

Таблица 5.2. Сводная таблица характеристик многоканальной системы с отказами.

Многоканальная система с отказами	
Характеристика системы	Формула
Интенсивность потока заявок	$\rho = \frac{\lambda}{\mu}$
Вероятность обслуживания	$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^k}{k!} + \ldots + \frac{\rho^n}{n!}\right]^{-1}$
Вероятность отказа в обслуживании	$P_{отк} = \frac{\rho^n}{n!} \cdot p_0$
Относительная пропускная способность	$Q = 1 - P_{отк} = 1 - \frac{\rho^n}{n!} \cdot p_0$
Абсолютная пропускная способность канала	$A = \lambda Q = \lambda \cdot \left(1 - \frac{\rho^n}{n!} \cdot p_0\right)$
Среднее число занятых каналов	$\bar{k} = \rho \cdot \left(1 - \frac{\rho^n}{n!} \cdot p_0\right)$

Многоканальная система с очередью

Пусть имеется n каналов обслуживания, на которые поступает поток заявок с интенсивностью λ. Поток обслуживаний каждого канала имеет интенсивность μ. Помимо каналов обслуживания, пусть имеется m мест в очереди. Найти предельные вероятности состояний системы и показатели ее эффективности.

Система S (СМО) имеет следующие состояния (нумеруем их по числу заявок, находящихся в системе): $S_0, S_1, ..., S_n, S_{n+1}, \ldots, S_{n+m}$, где S_k – состояние системы, когда в ней находится k заявок, т.е. занято k каналов, если $k \leq n$, а если $k > n$, то занято еще и $k - n$ мест в очереди. Граф состояний СМО представлен на рис. 5.4.

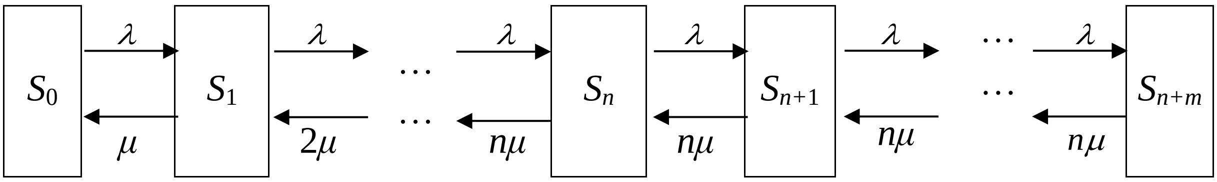


Рис. 5.4. Граф состояний многоканальной системы с очередью.

Поток заявок последовательно переводит систему из любого левого состояния в соседнее правое с одной и той же интенсивностью λ. Интенсивность же потока обслуживании, переводящих систему из любого правого состояния в соседнее левое, зависит от состояния:

Для $k \le n$, если СМО находится в состоянии S_k (k каналов заняты обслуживанием), то система может перейти в состояние $S_{k\text{-}1}$ (занято на один канал меньше), когда закончит обслуживание один из k занятых каналов, т.е. суммарная интенсивность их потоков обслуживаний будет $k\mu$.

Для $k > n$, обслуживанием заняты все n имеющихся каналов обслуживания, поэтому суммарный поток обслуживании, переводящий СМО из состояния S_k в $S_{k\text{-}1}$, будет иметь интенсивность $n\mu$, так как может освободиться любой из n имеющихся каналов обслуживания.

Пусть, как и ранее

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot M\left(t_{об}\right) \tag{5.12}$$

– среднее число поступающих заявок в течении среднего времени обслуживания одной заявки, т.е. приведенная плотность потока заявок, поступающих в систему.

Обозначим, также $w = \dfrac{\rho}{n}$.

Предельные вероятности состояний тогда равны:

$$p_1 = \frac{\rho}{1!} \cdot p_0,\ p_2 = \frac{\rho^2}{2!} \cdot p_0,\ \ldots,\ p_n = \frac{\rho^n}{n!} \cdot p_0, \tag{5.13}$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} \cdot p_0,\ \ldots,\ p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} \cdot p_0, \tag{5.14}$$

где

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n \cdot n!} + \ldots + \frac{\rho^{n+m}}{n^m \cdot n!}\right]^{-1}. \tag{5.15}$$

Тогда основные показатели эффективности рассматриваемой системы могут быть вычислены по следующим формулам (табл. 5.3):

Таблица 5.3. Сводная таблица характеристик многоканальной системы с очередью.

Многоканальная система с очередью	
Характеристика системы	Формула
Интенсивность потока заявок	$\rho = \frac{\lambda}{\mu}$
Интенсивность потока заявок на канал	$w = \frac{\rho}{n}$
Вероятность отказа	$P_{отк} = p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} \cdot p_0$
Относительная пропускная способность	$Q = 1 - P_{отк}$
Абсолютная пропускная способность	$A = \lambda Q$
Среднее число занятых каналов	$\bar{k} = \frac{A}{\mu}$
Среднее число заявок в очереди	$\bar{r} = \frac{\rho^{n+1} p_0}{n \cdot n!}\left[1 + 2w + 3w^2 + \ldots + mw^{m-1}\right]$
Среднее число заявок в системе	$\bar{z} = \bar{k} + \bar{r}$
Среднее время ожидания заявки в очереди	$t_{ож} = \frac{\bar{r}}{\lambda}$
Среднее время пребывания заявки в системе	$t_{сист} = t_{ож} + \frac{Q}{\mu}$

Вычисление показателей эффективности СМО

Напомним, что систему массового обслуживания (СМО) (например, диспетчерскую по обработке вызовов, информационную или телекоммуникационную системы, заявки на обслуживание или ремонт), математически представляют системой, находящейся в одном из нескольких взаимоисключающих состояний. Характеристики системы обслуживания таким образом зависят от того, какой процент времени система находится в каждом из состояний. Так, если система находится в состоянии «свободно», когда ни один из каналов обслуживания не занят, она может немедленно приступить к обработке вновь поступившей заявки, также и в состоянии, когда хотя бы один канал обслуживания свободен. Средний процент времени нахождения в каждом из состояний оценивается с помощью математического понятия ***вероятности***.

Для лучшего понимания расчета характеристик системы рассмотрим подробно следующий пример.

Пример 5.1 Имеется n-канальная СМО с m-ограниченной очередью, где

$$n = 2,\ m = 1,\ \lambda = 2,\ \mu = 3.$$

1. Построить размеченный граф состояний.
2. Составить и решить систему, состоящую из уравнений баланса для состояний.
3. Построить ***масштабированный*** граф состояний.
4. Вычислить показатели эффективности:

   $$L_{\text{сист}},\ L_{\text{очер}},\ L_{\text{обсл}},\ T_{\text{сист}},\ T_{\text{очер}},\ T_{\text{обсл}},$$

   используя матожидания.
5. Сравнить с показателями эффективности, вычисленными по формулам для соответствующей СМО.

Указание. Расчеты проводить с точностью до второго знака после запятой.
Решение.
1. В предположении пуассоновости потока событий, т.е.
   а) ординарности,
   б) стационарности,
   в) отсутствия последействий,

получаем диаграмму (рис. 5.5), известную также как ***модель гибели-размножения***:

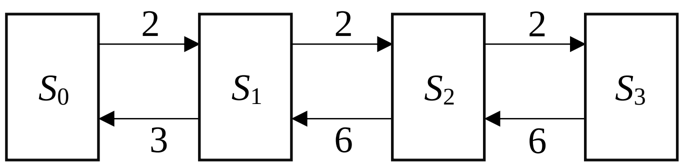


Рис. 5.5. Граф состояний 2-канальной системы с отказами и очередью длиной 1.

Здесь используются следующие обозначения:

S_0 – состояние системы, когда заявок на обслуживание в системе нет и все каналы обслуживания, и очередь свободны;
S_1 – состояние, когда занят один канал обслуживания;
S_2 – заняты оба канала обслуживания, очередь свободна;
S_3 – заняты оба канала обслуживания и занято единственное место в очереди.

По условию ординарности (предположение о том, что заявки могут поступать только по одной), находящаяся в состоянии S_i система может перейти только в соседние состояния S_{i+1} или S_{i-1}.

Из состояния S_0 – «заявок на обслуживании нет», система может перейти с вероятностной интенсивностью $\lambda = 2$ в состояние S_1 – «занят один канал обслуживания». И наоборот, из состояния S_1 система переходит с вероятностной интенсивностью μ = 3 в состояние S_0, так как с такой интенсивностью единственный занятый в этом состоянии канал обслуживания обрабатывает поступившую заявку. Из состояния S_2 система переходит в состояние S_1 уже с интенсивностью 2μ = 6, ввиду того, что над обслуживанием заявок работаю уже два канала обслуживания. Интенсивности остальных переходов обосновываются аналогично.

2. Составим решим систему уравнений баланса потока для каждого из состояний S_0, S_1, S_2, S_3 по схеме:

сумма интенсивностей входящих потоков событий равна сумме интенсивностей исходящих потоков событий

$$\begin{cases} 3p_1 = 2p_0 \\ 2p_0 + 6p_2 = 2p_1 + 3p_1 \\ 2p_1 + 6p_3 = 2p_2 + 6p_2 \\ 2p_2 = 6p_3 \end{cases} \quad \begin{vmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{vmatrix} \tag{5.16}$$

Здесь первое уравнение получено из расчета, что при устоявшемся режиме функционирования системы, интенсивность потока событий $2p_0$, переводящих состояние S_0 в S_1 и связанных с темпом поступлением новых заявок λ = 2, равна интенсивности потока событий $3p_1$, переводящих состояние S_0 в S_1 и связанных со средней скоростью μ = 3 обработки заявки одним каналом обслуживания. Остальные уравнения системы получаются аналогично.

Используя тот факт, что в каждом последующем уравнении, левая часть содержит правую часть, а правая – левую часть предыдущего уравнения, и их можно сократить, получим:

$$\begin{cases} 3p_1 = 2p_0 \\ 6p_2 = 2p_1 \\ 6p_3 = 2p_2 \\ 2p_2 = 6p_3 \end{cases}$$

Отсюда можно вывести следующие зависимости:

$$\begin{cases} p_1 = \dfrac{2}{3} p_0 \\ p_2 = \dfrac{1}{3} p_1 = \dfrac{1}{3}\dfrac{2}{3} p_0 = \dfrac{2}{9} p_0 \\ p_3 = \dfrac{1}{3} p_2 = \dfrac{1}{3}\dfrac{2}{9} p_0 = \dfrac{2}{27} p_0 \end{cases}$$

Используя, тот факт, что сумма вероятностей полной группы несовместных событий должна быть равна единице, получаем

$$\sum_{i=0}^{n+m} p_i = p_0 + p_1 + p_2 + p_3 = 1 \Rightarrow$$

$$\Rightarrow p_0 + \frac{2}{3}p_0 + \frac{2}{9}p_0 + \frac{2}{27}p_0 = p_0\left(1+\frac{2}{3}+\frac{2}{9}+\frac{2}{27}\right) = p_0\frac{53}{27} = 1.$$

Из последнего равенства получаем значения вероятностей нахождения в каждом из состояний системы:

$$p_0 = 1/\left(\frac{53}{27}\right) = \frac{27}{53} \approx 0,51;$$

$$p_1 = \frac{2}{3}p_0 = \frac{2}{3}\cdot\frac{27}{53} = \frac{18}{53} \approx 0,34;$$

$$p_2 = \frac{2}{9}p_0 = \frac{2}{9}\cdot\frac{27}{53} = \frac{6}{53} \approx 0,11;$$

$$p_3 = \frac{2}{27}p_0 = \frac{2}{27}\cdot\frac{27}{53} = \frac{2}{53} \approx 0,04.$$

Проверка: 0,51 + 0,34 + 0,11 + 0,04 = 1 – верно, ввиду того, что состояния системы несовместны и образуют полную группу событий.

3. Построим масштабированный граф состояний.

Для этого слева отложим вертикальный отрезок единичной длины (выбранная длина условна и может равняться любой величине, например, 2,5 см, выбор её обусловлен только удобством представления графа состояний).

Далее, на построенном отрезке единичной длины отложим отрезки с длинами, пропорциональные вероятностям p_0, p_1, p_2, p_3, то есть с длинами

0,51; 0,34; 0,11; 0,04.

Ориентируясь на полученные отрезки построим окружности также диаметрами

0,51; 0,34; 0,11; 0,04;

которые будут обозначать отдельные состояния системы, и соединим их стрелками-переходами.

Кружок отвечающий состоянию S_3 закрасим, чтобы выделить состояние, в котором происходит отказ обслуживания и визуально оценить вероятность этого, как правило, неблагоприятного события.

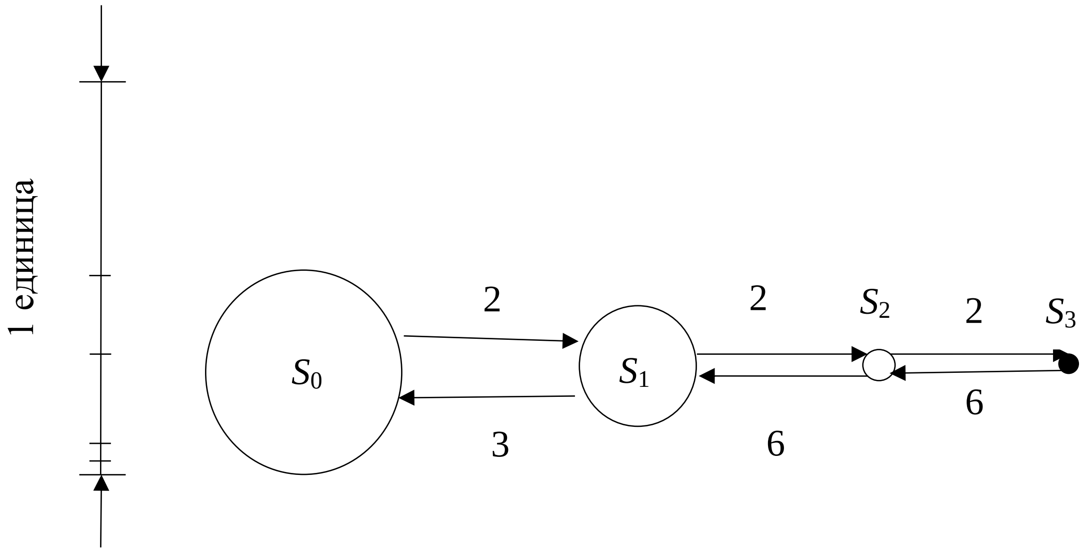


Рис. 5.6. Масштабированный граф состояний рассматриваемой СМО. Размеры (радиусы) кругов-состояний пропорциональны вероятностям, с которыми система находится в каждом из этих состояний.[6]

4. Найдем показатели эффективности системы.

Среднее число $L_{сист}$ заявок (длина последовательности заявок) в системе равна средней длине очереди заявок плюс среднее число заявок на обслуживании, и равна матожиданию случайной величины X – количества заявок в системе, имеющей следующее распределение

x_i	0	1	2	3
p_i	p_0	p_1	p_2	p_3

здесь x_i – возможные значения случайной величин, p_i –вероятность того, что соответствующее значение может случиться.

Само математическое ожидание (матожидание) $M[X]$ числовой дискретной случайной величины X рассчитывается по формуле:

$$M[X] = \sum_i x_i \cdot p_i \,. \tag{5.17}$$

- Если в системе находится 0 заявок, то это возможно только в состоянии S_0, вероятность нахождения в котором для системы равна p_0;
- в системе находится 1 заявка, когда она находится на обслуживании; это возможно, когда система находится в состоянии S_1, с вероятностью p_1;
- если в системе находится 2 заявки, то они обе находятся на обслуживании, что соответствует состоянию S_2;
- 3 заявки в системе означает, что две из них находятся на обслуживании, а одна в очереди – это соответствует состоянию S_3.

Тогда

[6] Иногда, вместо линейных размеров – радиусов, удобнее использовать площади кругов для выражения вероятностей.

$$L_{\text{сист}} = M[\text{количество заявок в системе}] =$$
$$= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 =$$
$$= 0 \cdot 0{,}51 + 1 \cdot 0{,}34 + 2 \cdot 0{,}11 + 3 \cdot 0{,}04 = 0{,}68$$

Аналогично, средняя длина очереди $L_{\text{очер}}$ (без учета заявок, находящихся на обслуживании) равна матожиданию случайной величины – количества заявок в очереди, имеющей следующее распределение:

0	1
$p_0 + p_1 + p_2$	p_3

Тогда можно вычислить среднюю длину очереди как матожидание случайной величины:

$$L_{\text{очер}} = M[\text{количество заявок в очереди}] =$$
$$= 0 \cdot (p_0 + p_1 + p_2) + 1 \cdot p_3 = 1 \cdot 0{,}04 = 0{,}04.$$

Аналогично, для среднего числа заявок, находящихся на обслуживании,

$$L_{\text{обсл}} = M[\text{количество заявок на обслуживании}] =$$
$$= 0 \cdot p_0 + 1 \cdot p_1 + 2(p_2 + p_3) =$$
$$= 0 \cdot 0{,}51 + 1 \cdot 0{,}34 + 2 \cdot (0{,}11 + 0{,}04) = 0{,}64.$$

Проверка: $L_{\text{сист}} = L_{\text{очер}} + L_{\text{обсл}}$, то есть должно быть 0,68 = 0,04 + 0,64 – верно.

Значение $\bar{k}$ – среднее количество занятых каналов обслуживания вычисляется как матожидание следующей случайной величины:

0	1	2
p_0	p_1	$p_2 + p_3$

Вероятность, что ни один канал обслуживания не занят, равна вероятности нахождения системы в состоянии S_0, то есть p_0; вероятность того, что занят один канал – вероятности нахождения в состоянии S_1; вероятность того, что заняты оба канала обслуживания равна вероятности нахождения либо в состоянии, когда оба канала заняты, а очередь свободна, либо оба заняты и единственное место в очереди занято, то есть

$$P(S_2 + S_3) = P(S_2) + P(S_3) = p_2 + p_3,$$

где первый переход делается по формуле сложения вероятностей для несовместных событий. Соответственно,

$$\bar{k} = M[\text{количество занятых каналов}]$$
$$= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot (p_2 + p_3) =$$
$$= 0 \cdot 0{,}51 + 1 \cdot 0{,}34 + 2 \cdot (0{,}11 + 0{,}04) =$$
$$= 0{,}34 + 0{,}30 = 0{,}64.$$

Таким образом, в среднем в каждый момент времени загружено менее одного (точнее, в среднем, 0,64) канала обслуживания, и система может показаться избыточной, ввиду значительной средней недогруженности, но тут надо смотреть и на другие критические параметры системы, например, на вероятность отказа в обслуживании (в нашем случае $p_{\text{отк}} = p_3 = 0{,}04$ – вероятность нахождения в состоянии, когда оба канала обслуживание и единственное место в очереди заняты). В случае, если отказ в обслуживании может привести к смерти людей (например, при вызове скорой или пожарных), даже такая величина может оказаться слишком большой и потребуется вводить новые каналы обслуживания (бригады или расчеты).

Используя, так называемую, ***теорему Литтла***, можно найти среднее время нахождения заявки в системе, в очереди и на обслуживании:

$$T_{\text{сист,очер,обсл}} = \frac{L_{\text{сист,очер,обсл}}}{\lambda} = 0{,}34;\ 0{,}02;\ 0{,}32 \text{ единиц времени,}$$

где единица времени, это продолжительность промежутка времени, выбранная за основу измерения времени в задаче, например, один час. Тогда значение $\lambda = 2$ интерпретируется, как поступление в среднем двух заявок в час; $\mu = 3$ – как обработка каналом обслуживания в среднем 3 заявок в час.

5. Сравним найденные нами значения с показателями эффективности вычисленными по формулам для СМО «**Многоканальная система с очередью»** (табл. 5.3), обращая внимание на различия в обозначениях. В прямоугольники будем заключать значения, совпавшие с предыдущими расчетами.

Имеем:

$$n = 2,\ m = 1,\ \lambda = 2,\ \mu = 3.$$

Тогда приведенная плотность потока заявок, поступающих в систему:

$$\rho = \frac{\lambda}{\mu} = \frac{2}{3}, \text{ соответственно, } w = \frac{\rho}{n} = \frac{2}{3} : 2 = \frac{1}{3};$$

предельные вероятности состояний:

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n \cdot n!} + \ldots + \frac{\rho^{n+m}}{n^m \cdot n!}\right]^{-1} =$$

$$= \left[1 + \frac{2}{3} \cdot \frac{1}{1!} + \left(\frac{2}{3}\right)^2 \cdot \frac{1}{2!} + \left(\frac{2}{3}\right)^3 \cdot \frac{1}{2 \cdot 2!}\right]^{-1} = \left[1 + \frac{2}{3} + \frac{2}{9} + \frac{2}{27}\right]^{-1} =$$

$$= \left[\frac{27}{27} + \frac{18}{27} + \frac{6}{27} + \frac{2}{27}\right]^{-1} = \left[\frac{53}{27}\right]^{-1} = \frac{27}{53} \approx \boxed{0{,}51};$$

$$p_1 = \frac{\rho}{1!} \cdot p_0 = \frac{2}{3} \cdot \frac{1}{1!} \cdot \frac{27}{53} = \frac{18}{53} \approx \boxed{0{,}34};$$

$$p_2 = \frac{\rho^2}{2!} \cdot p_0 = \left(\frac{2}{3}\right)^2 \cdot \frac{1}{2!} \cdot \frac{27}{53} = \frac{6}{53} \approx \boxed{0,11};$$

$$p_3 = p_{2+1} = \frac{\rho^3}{2 \cdot 2!} \cdot p_0 = \left(\frac{2}{3}\right)^3 \cdot \frac{1}{2 \cdot 2!} \cdot \frac{27}{53} = \frac{2}{53} \approx \boxed{0,04};$$

вероятность отказа:

$$P_{отк} = p_{n+m} = p_3 = \frac{2}{53} \approx 0,04;$$

относительная пропускная способность:

$$Q = 1 - P_{отк} = 1 - \frac{2}{53} = \frac{51}{53} \approx 0,96;$$

абсолютная пропускная способность:

$$A = \lambda Q = 2 \cdot \frac{51}{53} = \frac{102}{53} \approx 1,92;$$

среднее число занятых каналов (= $L_{обсл}$):

$$\bar{k} = \frac{A}{\mu} = \frac{102}{53} : 3 = \frac{34}{53} \approx \boxed{0,64};$$

среднее число заявок в очереди (= $L_{очер}$):

$$\bar{r} = \frac{\rho^{n+1} p_0}{n \cdot n!} \left[1 + 2w + 3w^2 + ... + mw^{m-1}\right] = \frac{2}{53} \cdot 1 = \frac{2}{53} \approx \boxed{0,04};$$

среднее число заявок в системе (= $L_{сист}$):

$$\bar{z} = \bar{k} + \bar{r} = \frac{34}{53} + \frac{32}{53} = \frac{36}{53} \approx 0,68;$$

среднее время ожидания заявки в очереди (= $T_{очер}$):

$$t_{ож} = \frac{\bar{r}}{\lambda} = \frac{2}{53} : 2 = \frac{1}{53} \approx \boxed{0,02};$$

среднее время пребывания заявки в системе (= $T_{сист}$):

$$t_{сист} = t_{ож} + \frac{Q}{\mu} = \frac{1}{53} + \frac{51}{53} : 3 = \frac{1}{53} + \frac{17}{53} = \frac{18}{53} \approx \boxed{0,34}.$$

Таким образом, результаты наших непосредственных вычислений совпали с результатами, полученными по теоретическим формулам. □

Контрольные вопросы

1. Что такое система массового обслуживания (СМО)?
2. Какие аспекты абстрактных моделей моделируют СМО?
3. Что такое пуассоновский поток событий?
4. Что такое *n*-канальная СМО?
5. Что такое *n*-канальная СМО с очередью?

6. Как выглядит граф состояний для n-канальной СМО с очередью длиной m?
7. Какие вероятностные характеристики СМО связаны с качеством обслуживания?
8. О чем говорят предельные вероятности состояний?

Заключение

В учебном пособии изложены необходимые теоретические сведения по дисциплине «Системный анализ и исследование операций». Материал пособия проиллюстрирован решениями примеров, связанными с задачами эффективного планирования и управления.

Системный анализ и исследование операций применяются на практике в целях эффективной организации работы, планирования проезда/доставки по наикратчайшему маршруту или с наименьшими затратами времени и топлива, также в задачах оптимального планирования производства и распределения ресурсов. Тем более актуальным становится применение этих методов в условиях недостатка ресурсов, когда имеется необходимость наиболее разумного и рационального распределения имеющиеся средства для поддержания каждодневной деятельности и развития.

Другой класс задач, к которым применимы методы исследования операций, связан с инженерными и техническими приложениями. К ним относятся задачи эффективного управления, регулирования и вывода на рабочие режимы трубопроводов, газопроводов, электросетей и систем теплоснабжения, особенно в условиях нарастающей цифровизации экономики и госуправления. Подходы системного анализа незаменимы при создании цифровых двойников инфраструктурных объектов.

Методы динамического программирования приложимы также к решению более узкого класса задач цифрового управления механотронными и робототехническими системами. В частности, наряду с родственными методами прикладной механики, соответствующие алгоритмы применяются для цифрового моделирования динамики механических систем.

Таким образом, понимание общих принципов системного анализа и исследования операций, знание существующих практических методов, предпосылок их применимости и имеющихся ограничений становится в современных условиях важной частью подготовки к решению задач управления и оптимального распределения ресурсов.

Литература

1. Алексеев В.М., Галеев Э.М., Тихомиров В.М. Сборник задач по оптимизации. Теория. Примеры. Задачи. Учебное пособие – 2-е изд., – М.: «Физматлит», 2005. – 256 с.
2. Ашманов С.А., Тимохов А.В. Теория оптимизации в задачах и упражнениях. Учеб. пособие. – 2-е изд., стер. – СПб.: «Лань», 2012. – 448 с.
3. Вдовин, В. М. Теория систем и системный анализ: учебник для бакалавров / В. М. Вдовин, Л. Е. Суркова, В. А. Валентинов. – 5-е изд., стер. – Москва: Издательско-торговая корпорация «Дашков и К°», 2020. – 642 с.
4. Гармаш, А. Н., Орлова, И. В. Математические методы в управлении: Учебное пособие. – Москва: Вузовский учебник: НИЦ Инфра-М, 2012. – 272 с.
5. Гасс С. Линейное программирование. – М.: «Физматгиз», 1961. – 304 с.
6. Гмурман В.Е. Теория вероятностей и математическая статистика: Учебное пособие для вузов / В.Е. Гмурман. 12-е изд., стер. – М.: Юрайт, 2014. – 479 с.
7. Каштанов, В. А., Зайцева, О. Б. Исследование операций (линейное программирование и стохастические модели): учебник / В.А. Каштанов, О.Б. Зайцева. – Москва: КУРС, 2017. – 256 с.
8. Колемаев, В. А. Математические методы и модели исследования операций: учебник для студентов вузов, обучающихся по специальности 080116 «Математические методы в экономике» и другим экономическим специальностям / В. А. Колемаев; под ред. В. А. Колемаева. – Москва: ЮНИТИ-ДАНА, 2012. – 592 с.
9. Лемешко, Б. Ю. Теория игр и исследование операций / Лемешко Б.Ю. – Новосибирск: НГТУ, 2013. – 167 с.
10. Соколов, Г. А. Основы теории массового обслуживания для экономистов: учебник / Г.А. Соколов. – М.: ИНФРА-М, 2019. – 128 с. – (Высшее образование: Бакалавриат).
11. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации. – 2-е изд., – М.: «Физматлит», 2005. – 544 с.
12. Тимченко, Т. Н. Системный анализ в управлении: учебное пособие / Т. Н. Тимченко. – Москва: РИОР, 2008. – 161 с.
13. Vanderbei R. Linear programming Foundations and Extensions. 2nd edition. 2001. – 466 pp.

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

учебное пособие

Авторы:

Матеров Евгений Николаевич – заведующий кафедрой физики, математики и информационных технологий Сибирской пожарно-спасательной академии ГПС МЧС России, кандидат физико-математических наук.

абёнышев Сергей Валерьевич – профессор кафедры физики, математики и информационных технологий Сибирской пожарно-спасательной академии ГПС МЧС России, кандидат физико-математических наук.