Содержание

РАЗДЕЛ 1. ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

ТЕМА 1. ОБЩИЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРОЦЕССОВ В МАШИНОСТРОЕНИИ

Лекция 1. Предмет, цель и задачи дисциплины. Понятие модели. Виды математических моделей. Классификация математических моделей. Требования, предъявляемые к математическим моделям. Этапы решения задачи методом математического моделирования. Погрешности результатов при математическом моделировании.

Предмет, цель и задачи дисциплины

Компьютеризация современного производства требует от будущего инженера хорошей математической подготовки, которую, в частности, должна обеспечить дисциплина «Математическое моделирование и алгоритмизация инженерных задач». Данная дисциплина является связующим звеном между общепрофессиональными и специальными дисциплинами.

Цель преподавания дисциплины «Математическое моделирование и алгоритмизация инженерных задач» – приобретение навыков разработки и использования математических моделей для описания, исследования и оптимизации процессов в машиностроении.

Основные задачи дисциплины «Математическое моделирование и алгоритмизация инженерных задач» состоят в изучении:

- общих понятий математического моделирования процессов в машиностроении (структуры, классификации и областей применения математических моделей, предъявляемых к ним требований);
- теоретических основ математического моделирования и оптимизации процессов в машиностроении;
- вопросов математического моделирования физических процессов в технологических системах;
- вопросов математического моделирования и оптимизации технологических станочных систем.
- приобретение практических навыков, необходимых для разработки алгоритмов решения проектных и научно-

исследовательских задач, связанных с математическим моделированием.

Понятие модели

Под объектами моделирования в машиностроительном производстве следует понимать:

1. Технологические системы (ТС) – участки из универсальных станков, автоматические линии, гибкие производственные системы (ГПС).
2. Технологические процессы (ТП).
3. Физические процессы (ФП) – процессы, протекающие при резании металлов, при функционировании технологического оборудования в упругой системе СПИД и т.д.

Моделирование представляет собой процесс замещения объекта исследования некоторой его моделью и проведение исследований на модели с целью получения необходимой информации об объекте.

Модель – это физический или абстрактный образ объекта, удобный для проведения исследований и позволяющий адекватно отображать интересующие исследователя физические свойства и характеристики объекта.

Различают предметное и абстрактное моделирование.

При *предметном моделировании* строят *физическую модель*, которая соответствующим образом отображает основные физические свойства и характеристики моделируемого объекта. При этом модель может иметь иную физическую природу по сравнению с объектом. Если физическая природа модели и объекта совпадают, то моделирование называются *физическими.*

При *абстрактном моделировании* строится абстрактная модель. Такая модель представляет собой математические соотношения, графы, схемы, диаграммы и т. п. Наиболее мощным средством построения абстрактных моделей является математическое моделирование.

Математическое моделирование 1)позволяет посредством известных математических функций, символов и зависимостей описать функционирование технического объекта во внешней среде; 2)определить выходные параметры и характеристики; 3)получить оценку показателей эффективности и качества; 4)осуществить поиск оптимальной структуры и параметров объектов.

Под ***математической моделью*** понимается совокупность математических объектов и отношений между ними, адекватно отображающих физические свойства объекта. ***Математическое моделирование*** – это процесс формирования математической модели и использования ее для анализа и синтеза.

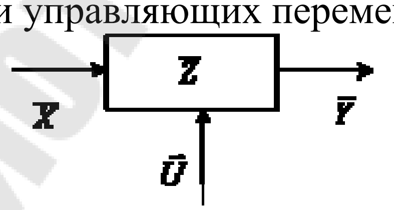
В конструкторской практике под математическим моделированием понимается процесс построения математической модели, а проведение исследований на модели в процессе проектирования называют *вычислительным экспериментом.*

Для осуществления вычислительного эксперимента на ЭВМ необходимо разработать алгоритм реализации математической модели.

Алгоритм – это предписание, определяющее последовательность выполнения операций вычислительного процесса. Для наглядности алгоритмы чаще всего представляют в виде схем и графов, иногда дают их вербальное (словесное) описание. Алгоритм, записанный в форме, воспринимаемой вычислительной машиной представляет собой *программную модель*. Процесс программирования называют *программным моделированием*.

Виды математических моделей

Математическую модель сложной системы (рис. 1.1) можно представить как функциональный оператор Ф, преобразующий пространство вектора входных $\vec{X}$, внутрисистемных параметров процесса $\vec{Z}$ и управляющих переменных $\vec{U}$ в выходные параметры $\vec{Y}$.



$\Phi(\vec{X},\vec{Z},\vec{U}) = \vec{Y}$

Рисунок 1.1 – Объект моделирования

Вид модели и степень ее детализации определяется не только свойствами моделируемого объекта, но и целью, с которой выполняется моделирование. Поэтому процесс разработки модели сложной системы состоит в последовательном анализе и моделировании отдельных ее подсистем с последующим установлением связей между этими подсистемами.

Процесс построения моделей можно представить следующим образом (рис.1.2).

Система-объект моделирования

Цель моделирования

Уровни детализации	Виды моделей или их реализации	Задачи моделирования
Содержательный	Концептуальная (содержательная модель)	Формулировка задачи моделирования
Структурный (топологический)	Формализованная модель (балансы и ограничения)	Математическая формулировка (постановка) задачи
Алгоритмический (функциональный)	Алгоритм решения формализованной модели	Разработка основных этапов решения математической модели
Параметрический	Программная реализация разработанного алгоритма	Идентификация параметров
Уровень адекватности модели	Критерии адекватности	Проверка адекватности математической модели

Рисунок 1.2 – Процесс создания математических моделей

На первом этапе создания модели выделяются признаки, характеризующие систему и системообразующие элементы, а также отношения, на которых реализуются эти признаки. Это позволяет определить исследуемый объект как систему.

На втором - определяется цель моделирования системы.

На третьем этапе на каждом уровне детализации разрабатываются математические модели и модели координаторов для взаимодействия между уровнями.

На первом уровне изучают интересующую систему (объект моделирования) и описывают ее содержательно. Такое описание называют концептуальной (содержательной) моделью, представляющей собой словесное описание математической формулировки задачи.

Затем формулируют концептуальную модель, для чего разрабатывают структуру модели. Это структурный или топологический уровень формирования модели, на котором модель записывается в виде балансовых соотношений и ограничений. Далее на алгоритмическом уровне разрабатывают алгоритм решения математической модели.

Программная реализация которого соответствует следующему уровню детализации – параметрическому, на котором определяются параметры модели.

И далее на последнем уровне проводится проверка адекватности модели моделируемому объекту.

Модель считается *адекватной*, если отражает исследуемые свойства с приемлемой точностью, которая оценивается степенью совпадения предсказанного в процессе эксперимента на модели значений выходных параметров с истинными значениями.

Классификация математических моделей

<u>По отношению ко времени</u> различают *статические и динамические модели*. Первые инвариантны ко времени, а вторые являются функцией времени.

<u>По характеру зависимости выходных параметров от входных</u> модели делятся *на детерминированные и стохастические*. Если существуют функциональные зависимости выходных параметров от входных, то модели являются детерминированными, если эти зависимости неизвестны, а известно лишь математическое описание выходов в виде функции входов, модели называются стохастическими.

<u>По характеру времени динамические модели</u> делятся *на непрерывные и дискретные.* Первые функционируют в непрерывном времени, а вторые - в дискретном. Примером непрерывных детерминированных моделей могут служить дифференциальные или интегро-дифференциальные уравнения; примером дискретных

детерминированных моделей – конечные автоматы, дискретных стохастических – вероятностные автоматы.

Основные виды моделей представленные на рисунке 1.3.

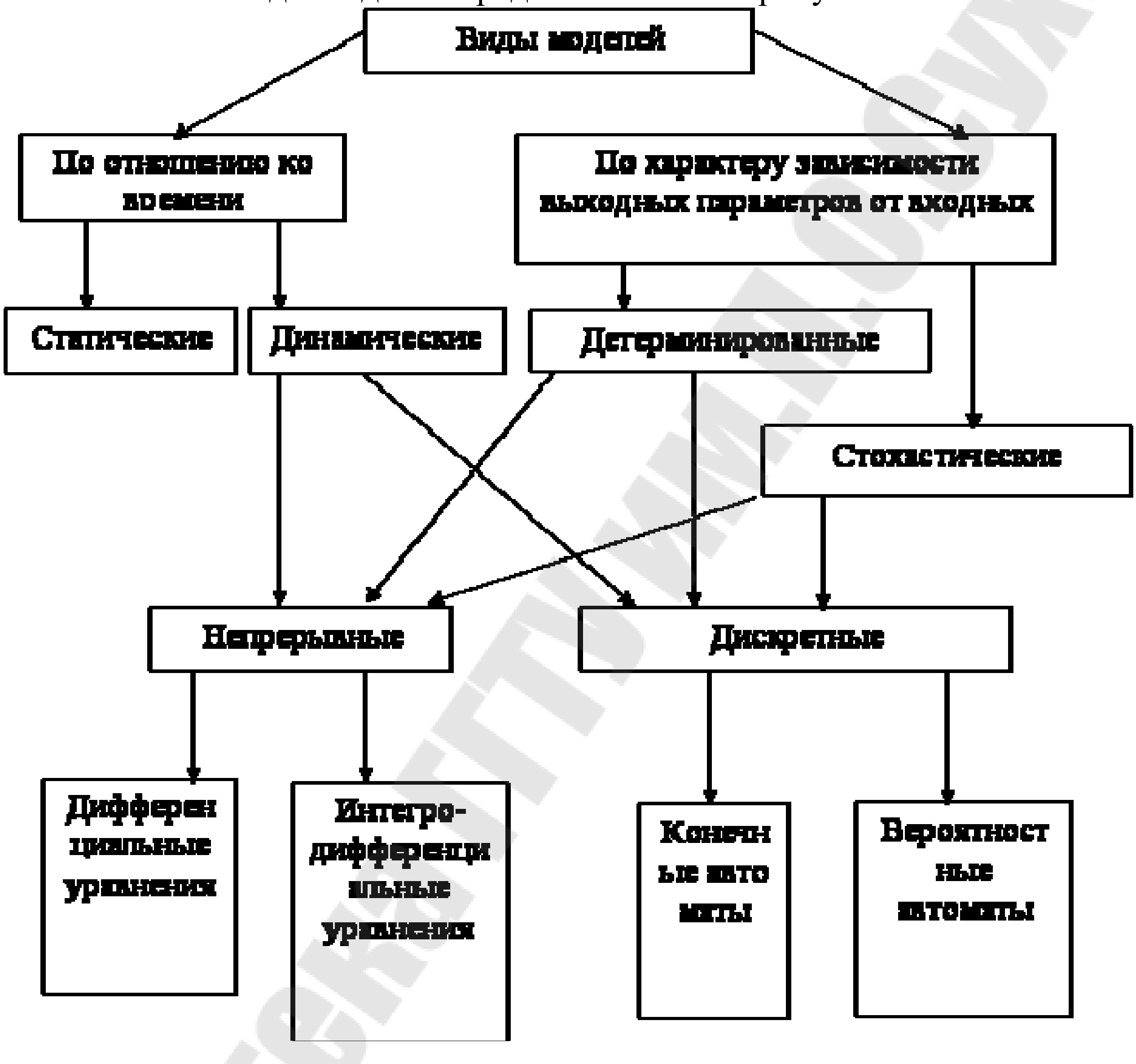


Рисунок 1.3 – Классификация математических моделей

Основные признаки классификации и типы ММ даны в таблице.

Таблица 1.1 – Классификация математических моделей

Признаки классификации	Виды математических моделей
1. Принадлежность к иерархическому уровню	Модели микроуровня Модели макроуровня

	Модели метауровня
2. Характер отображаемых свойств объекта	Структурные Функциональные
3. Способ представления свойств объекта	Аналитические Алгоритмические Имитационные
4. Способ получения модели	Теоретические Эмпирические
5. Особенности поведения объекта	Детерминированные Вероятностные

Приведенная классификация математических моделей может быть применена по отношению к любым объектам. Мы рассмотрим особенности различных видов моделей применительно к объектам (процессам) в машиностроении.

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

Математические модели на макроуровне производственного процесса описывают технологические процессы.

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

Структурные математические модели предназначены для отображения структурных свойств объектов. Например, в САПР ТП для представления структуры технологического процесса, расцеховки изделий используется структурно – логические модели.

Функциональные математические модели предназначены для отображения информационных, физических, временных процессов, протекающих в работающем оборудовании, в ходе выполнения технологических процессов и т.д.

Аналитические математические модели представляют собой явные математические выражения выходных параметров как функций от параметров входных и внутренних. Это, например, выражения для сил резания.

Пример 1.1. Пусть требуется описать и исследовать процесс резания (точения) с точки зрения действующих сил, которые представим известными в теории резания уравнениями:

$$P_x = C_{PX} t^{X_{px}} s^{Y_{px}} K_{ОБЩ}$$

$$P_Y = C_{PY} t^{X_{py}} s^{Y_{py}} K_{ОБЩ}$$
$$P_z = C_{PZ} t^{X_{pz}} s^{Y_{pz}} K_{ОБЩ}$$

где C_P– удельные силы резания;

t– глубина резания;

s– подача;

X, Y– показатели степени, выражающие влияние соответствующего параметра режима резания на силу резания (при точении $X_p \approx 1,\ Y_p \approx 0{,}75$);

$K_{ОБЩ} = K_M \cdot K_И \cdots$ коэффициент, зависящий от свойств обрабатываемого материала, инструментального материала и т.д.

Представленные выше три уравнения – это уже один из вариантов математической модели процесса точения с точки зрения действующих сил. Она, конечно же, простейшая и может служить для:

1. Описания процесса резания.
2. Исследования процесса резания.
3. Расчета сил резания.

Например, с помощью данной модели можно исследовать зависимость силы резания P_z от глубины резания t (рис. 1.4)

Рисунок 1.4 – Зависимость силы резания Pz от глубины резания t

Наиболее существенная характеристика аналитических моделей заключается в том, <u>что модель не является структурно подобной объекту моделирования.</u> Под структурным подобием здесь понимается однозначное соответствие элементов и связей модели элементам и связям моделируемого объекта.

Аналитическая модель может быть исследована следующим методами:

- *аналитическим,* когда стремятся получить в общем виде зависимости для искомых характеристик;
- *численными,* когда стремятся получить числовые результаты при конкретных начальных данных;
- *качественными*, когда имея решения в явном виде можно найти некоторые свойства решения (оценить устойчивость решения).

К аналитическим относятся модели, построенные на основе аппарата **математического программирования, корреляционного, регрессионного анализа**.

Аналитическая модель всегда представляет собой конструкцию, которую можно проанализировать и решить математическими средствами. Так, если используется аппарат математического программирования, то модель состоит в основе своей из целевой функции и системы ограничений на переменные. Целевая функция, как правило, выражает ту характеристику объекта (системы), которую требуется вычислить или оптимизировать. В частности, это может быть производительность технологической системы. Переменные выражают технические характеристики объекта (системы), в том числе варьируемые, ограничения – их допустимые предельные значения.

Аналитические модели являются эффективным инструментом для решения задач оптимизации процессов, протекающих в технологических системах, а также оптимизации и вычисления характеристик самих технологических систем.

Важным моментом является размерность конкретной аналитической модели. Часто для реальных технологических систем (автоматических линий, гибких производственных систем) размерность их аналитических моделей столь велика, что получение оптимального решения оказывается весьма сложным с вычислительной точки зрения. Для повышения вычислительной эффективности в этом случае используют различные приемы. Один из них связан с разбиением задачи большой размерности на подзадачи меньшей размерности так, чтобы автономные решения подзадач в определенной последовательности давали решение основной задачи. При этом возникают проблемы организации взаимодействия подзадач, которые не всегда оказываются простыми. Другой прием предполагает уменьшение точности вычислений, за счет чего удается сократить время решения задачи.

Алгоритмические математические модели выражают связи между выходными параметрами и параметрами входными и внутренними в виде алгоритма.

Имитационные математические модели – это алгоритмические модели, отражающие развитие процесса (поведение исследуемого объекта) во времени при задании внешних воздействий на процесс (объект). Например, это модели систем массового обслуживания, заданные в алгоритмической форме.

При *имитационном моделировании* процесс функционирования исследуемого объекта воспроизводится на ЭВМ в отсутствие аналитических зависимостей между входными, выходными параметрами и параметрами состояния системы. По результатам имитационного моделирования на ЭВМ можно прогнозировать поведение исследуемой системы.

Имитационные модели позволяют достаточно просто учитывать такие факторы, как наличие дискретных и непрерывных элементов, нелинейные характеристики системы, многочисленные случайные воздействия и другие.

Когда же результаты, полученные при воспроизведении на имитационной модели процесса функционирования системы, являются реализациями случайных величин и функций, то для нахождения характеристик процесса требуется его многократное воспроизведение с последующей статической обработкой информации. В этом случае в качестве метода машинной реализации имитационной модели следует использовать *метод статического моделирования (метод Монте – Карло).*

Работа с имитационной моделью заключается в *проведении имитационного эксперимента.* Процесс, протекающий в модели в ходе эксперимента, подобен процессу в реальном объекте. Поэтому исследование объекта на его имитационной модели сводится к изучению характеристик процесса, протекающего в ходе эксперимента.

Ценным качеством имитации является возможность управлять масштабом времени. Динамический процесс в имитационной модели протекает в так называемом системном времени. Системное время имитирует реальное время. При этом пересчет системного времени в модели можно выполнять двумя способами. *Первый способ* *заключается в «движении» по времени с некоторым постоянным шагом.* *Второй способ заключается в «движении» по времени от события к событию, при этом считается, что в промежутках времени между событиями в модели изменений не происходит*.

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

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

Детерминированные математические модели описывают поведение объекта с позиций полной определенности в настоящем и будущем. Примеры таких моделей: формулы физических законов, технологические процессы обработки деталей и т.д.

Вероятностные математические модели учитывают влияние случайных факторов на поведение объекта, т.е. оценивают его будущее с позиций вероятности тех или иных событий. *Примеры таких моделей:* описание ожидаемых длин очередей в системах массового обслуживания, ожидаемых объемов выпуска сверхплановой продукции производственным участком, точности размеров в партии деталей с учетом явления рассеяния и т.д.

Требования, предъявляемые к математическим моделям

К математическим моделям предъявляются следующие основные требования:

1. Универсальности.
2. Точности.
3. Адекватности.
4. Экономичности.

Универсальность математической модели характеризует полноту отражения в ней свойств реального объекта. Математическая модель отражают не все, а лишь некоторые свойства реального объекта. Например, формулы для сил резания, которые приведены в примере 1 , не учитывают температуру окружающего воздуха, влажность, экономические параметры и т.д.

Точность математической модели оценивается степенью совпадения значений выходных параметров реального объекта и значений тех же параметров, рассчитанных с помощью модели.

Пусть отражаемые в математической модели свойства объекта оцениваются вектором выходных параметров $Y = (y_1, y_2, ..., y_m)$, $y_{iмм}$ - i-ый параметр, рассчитанный с помощью модели, а y_{id} - истинное значение того же параметра. Тогда относительная погрешность математической модели по i – му параметру будет равна:

$$E_i = \frac{|y_{iмм} - y_{id}|}{y_{id}}$$

По этой формуле рассчитываются погрешности для каждого выходного параметра, в результате получается вектор погрешностей $E = (E_1, E_2, ..., E_m)$.

В целом для математической модели погрешность оценивается следующим образом:

$$E_{мм} = \max E_i, \quad i = \overline{1,m} .$$

Например, оценим погрешность рассмотренной в примере 1 математической модели. Вектор выходных параметров $Y = (P_X, P_y, P_z)$. Пусть $E_{PX} = E_1 = 0{,}05; E_{PY} = E_2 = 0{,}07; E_{PZ} = E_3 = 0{,}03;$;

тогда в целом погрешность математической модели $E_{мм} = \max E_i = 0{,}07$.

Адекватность математической модели – это ее способность отражать заданные свойства объекта с погрешностью, не выше заданной.

Так как выходные параметры модели являются функцией $Y = F(X,Q)$ от параметров внутренних и входных, то и точность модели зависит от их значений. Адекватность модели имеет место в ограниченной области изменения внутренних и входных параметров. Если обозначить область адекватности как OA, то

$$OA = \{X, Q \mid E_{мм} \le \delta\},$$

где δ - некоторое заданное число.

Экономичность математической модели характеризуется затратами вычислительных ресурсов на ее реализацию. Если работа с математической моделью осуществляется вручную, то ее экономичность определяется затратами личного времени проектировщика. Если модель используется при автоматизированном проектировании, то затратами машинного времени и памяти компьютера. Так как указанные величины определяются характеристиками конкретного компьютера, то использовать их для оценки экономичности математической модели не корректно. Поэтому, для оценки экономичности самой математической модели используют другие величины:

1. Среднее количество операций, выполняемых при одном обращении к математической модели.
2. Размерность системы уравнений в математической модели.
3. Количество используемых в модели внутренних параметров и т.д.

Требования высокой степени универсальности, точности, широкой области адекватности математической модели, с одной стороны, и высокой ее экономичности, с другой стороны, противоречивы. Поэтому компромиссные решения определяются решаемой задачей.

К математическим моделям предъявляется и целый ряд других требований, среди которых следует выделить следующие:

1. Вычислимость, т.е. возможность ручного или с помощью ЭВМ исследования качественных и количественных закономерностей функционирования объекта (системы).
2. Модульность, т.е. соответствие конструкций модели структурным составляющим объекта (системы).
3. Алгоритмизируемость, т.е. возможность разработки соответствующих алгоритма и программы, реализующей математическую модель на ЭВМ.
4. Наглядность, т.е. удобное визуальное восприятие модели.

Этапы решения задачи методом математического моделирования

Исследование реальных объектов методом математического моделирования в общем случае представляет собой последовательное выполнение следующих этапов:

1. *Формулировка цели исследования, которая должна быть достигнута при моделировании*. Целью исследования в значительной степени определяется сложность математической модели и сама возможность решения задачи.

2. *Анализ моделируемого объекта*. На этом этапе устанавливается, какие из признаков объекта оригинала являются существенными с точки зрения решаемой задачи, возможно, определяются исходные данные, характеризующие эти признаки, возможна разработка конструктивной или функциональной схемы объекта.
3. *Разработка физической модели и обоснование ее соответствия по существенным признакам моделируемому объекту.* На этом этапе формируются ограничения и предложения, которые упрощают описания объекта оригинала, не нанося при этом существенного ущерба, достоверности получаемого результата исследования. Этот этап важен, поскольку реальные объекты в своем большинстве крайне сложны, и без подобного упрощения их математическое моделирование зачастую право невозможно.
4. *Разработка математической модели.*
5. *Разработка алгоритма или алгоритмов компьютерной модели*. Здесь следует отметить, что на основании одной и той же модели могут быть построены различные алгоритмы. Это зависит от того, какие из параметров в рамках принимаемого исследования будут известными или неизвестными искомыми. Например, параметры, которые считаются известными при решении задач анализа, будут неизвестными при решении задач синтеза и наоборот. Изучение процесса проектирования на различных уровнях показывает, что ему характерен определенный состав и последовательность решаемых задач, основным из которых является задачи анализа из синтеза.

   Анализ технических объектов предлагает изучение их свойств. При анализе не создаются новые объекты, а исследуются заданные. Синтез технических объектов направлен на создание новых вариантов. Поскольку анализ направлен для оценки этих вариантов, то синтез этих вариантов и анализ выступают в процессе проектирования в диалектическом единстве.
6. *Разработка компьютерной программы или пакета программ*. Обычно для этих целей используют языки высокого уровня, такие как Паскаль, СИ, Бейсик, и др.
7. *Отладка и тестирование программ*. Этот этап часто является самым трудоемким. Здесь выявляются ошибки не только в программе, но и в алгоритме, и в самой математической

модели. Не редко на этапе отладки программ является некорректность или недопустимость допущений и ограничений, принятых на этапе разработки физической модели. Это в свою очередь может привести к необходимости отказа от разработанной математической модели и разработки новой.

8. *Численные исследования, анализ и интерпретация результатов численных исследований в терминологии объекта оригинала.*

Приведенная схема решения задач методов математического моделирования является условной в определенной степени. На рис. 1.5 приведена общая схема компьютерного математического моделирования.

Погрешности результатов при математическом моделировании

Необходимо понимать, что при решении определенного класса задач методом математического моделирования неизбежны погрешности определяемых результатов, которые по источнику их происхождения можно разбить на четыре группы.

Погрешности формализации. Возникают при переходе от объектов оригинала к его физической модели и обусловлены приближенным его соответствия этой модели. Количественную оценку этих погрешностей в общем случае указать нельзя. Обычно такую оценку получают при экстремальных исследованиях на самих объектах оригинала или их местах.

Погрешности исходных данных связаны с физическими применениями объекта оригинала. При измерениях в технических системах погрешность измерения технических условий обычно 1-5 %, погрешность измерения динамических величин 5-20 %. Это приводит к погрешностям получаемых результатов.

Погрешности вычислительного алгоритма связаны с приближенным решением математических соотношений численными методами. Погрешность вычислительного алгоритма должна быть в 2-5 раз меньше погрешности исходных данных. Следует иметь в виду, что неудачно разработанный алгоритм может привести к тому, что результаты численного решения не будут соответствовать исследуемой математической модели.

<u>*Погрешности машинного округления*</u> связаны с тем, что в ЭВМ все вычисление выполняются с определенным числом значащих цифр. Эта погрешность зависит от типа и программы. Погрешность машинного вычисления должна быть в 5-10 раз меньше погрешности вычислительного алгоритма.

Таким образом, для получения достоверных результатов при математическом моделировании необходимо глубокое понимание всех существенных особенностей исследуемого объекта, его модели, вычислительных алгоритмов, программы, возможностей вычислительной техники.

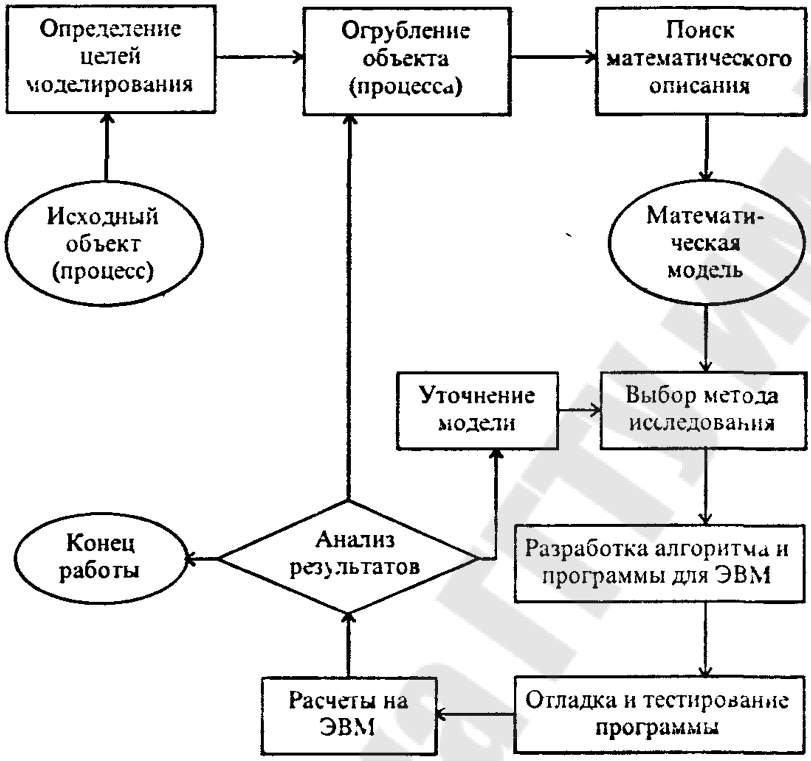


Рисунок 1.5 – Общая схема процесса компьютерного математического моделирования

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Что такое математическая модель?
2. Чем отличаются предметное и абстрактное моделирование?
3. Что позволяет описать, определить, осуществить, получить математическое моделирование?
4. Какие уровни детализации необходимы для создания математической модели?
5. Признаки классификации математических моделей.
6. Перечислите требования к математическим моделям.
7. Назовите погрешности результатов моделирования.

РАЗДЕЛ 2. РЕШЕНИЕ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ С ПОМОЩЬЮ ТЕОРИИ ГРАФОВ

ТЕМА 2. ОСНОВЫ ТЕОРИИ ГРАФОВ

Лекция 2. Теоретико-множественное определение графов. Матричные способы задания графов. Упорядочение элементов орграфа. Алгоритм Фалкерсона. Матричный способ упорядочивания вершин на примере орграфа.

Теоретико-множественное определение графов

Теория графов обладает широким разнообразием возможностей её применения в различных областях знаний.

В частности она является эффективным аппаратом формализации задач экономической и планово-производственной практики. В последнее время теория графов нашла важные применения в автоматизации управления производством, в календарном и сетевом планировании, при оптимизации размещения производства, в теории массового обслуживания, при рационализации схем перевозок продукции, при наиболее компактной записи и обработке экономической информации и др. Основным объектом этой теории является *граф*.

Формально граф G определяется заданием двух множеств X и U и обозначается $G=\{X,U\}$. Элементы множества X называют *вершинами.* Будем их обозначать буквами $x_1, x_2, \ldots, x_n$. Вершины изображают точками плоскости или пространства. Элементами множества U являются пары связанных между собой элементов множества X. Их изображают отрезками кривых или прямых линий. Обозначать их будем буквами $u_1, u_2, \ldots u_m$. Взаимное расположение, формы и длины упомянутых отрезков значения не имеют. Важно лишь, что они соединяют две данные вершины множества X. Если в паре вершин x_i и x_j указано направление связи, т.е. какая из них является первой, то соединяющий их отрезок u_k называется ***дугой*,** если же ориентация не указана, **-** ***ребром***; а вершины, определяющие дугу (ребро) u_k, называют ***концевыми вершинами*** дуги (ребра) u_k. Если концевые вершины совпадают, то дугу (ребро) называют ***петлёй***. На графе могут существовать дуги (ребра) с одинаковыми концевыми вершинами. Будем называть такие дуги (ребра) ***параллельными***.

Говорят, что дуга (ребро) инцидентна своим концевым вершинам, и обратно.

Если в графе *G* все элементы множества *U* изображаются дугами, то граф называется ***ориентированным (орграфом)***, если ребрами, - ***неориентированным***. Иногда рассматривают ***смешанные*** графы, состоящие из дуг и ребер. Геометрическое изображение графа часто бывает весьма наглядным. Этим и пользуются на практике (см рисунок 2.1). Вершины X={a, b, c, d, e, g, h}, дуги – U= {(a, a), (c, b), (c, d), (c, e), (d, c), (d, d), (g, h)}

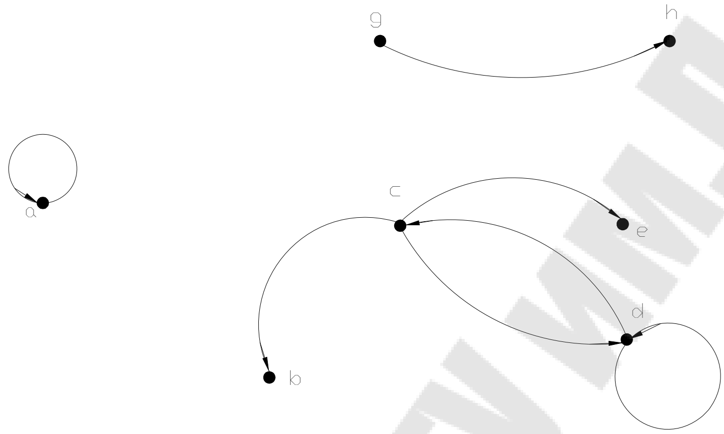


Рисунок 2.1 –Общий вид графа

Примерами графов могут служить схемы железных или шоссейных дорог, схемы связи поставщиков и потребителей, структурные формулы молекул и т. д. В приложениях обычно используются графы, в которых множества *X* и *U* состоят из конечного числа элементов. Такие графы называют ***конечными***.

Иногда бывает удобно дать графу другое определение.

Можно считать, что множество направленных дуг *U*, соединяющих элементы множества X, отображает это множество само в себя.

Таким образом, ***граф G есть пара (X, Г), состоящая из множества X и отображения Г, заданного на этом множестве G={X, Г}.***

Так для графа на рисунке 2.1 отображение *Г* задано следующим образом.

$\Gamma_a = \{a\}$; $\Gamma_{\mathrm{b}} = \varnothing$;

$\Gamma_{\mathrm{c}} = \{b, d, e\}$; $\Gamma_{\mathrm{d}} = \{c, d\}$; $\Gamma_{\mathrm{e}} = \{d\}$; $\Gamma_{\mathrm{g}} = \{h\}$; $\Gamma_{\mathrm{h}} = \varnothing$.

Дуга, соединяющая вершины *a* и *b* и направленная от *a* к *b*, обозначается *u*=(a, b).

Путем в орграфе G называется последовательность дуг $\mu = (u_1, \cdots, u_k)$, в которой конец каждой предыдущей дуги совпадает с началом последующей.

Путь, проходящей через все вершины и притом только по одному разу, называется ***гамильтоновым***.

На рис.2.2 представлен гамильтонов путь $x_1x_2x_3x_4x_5$.

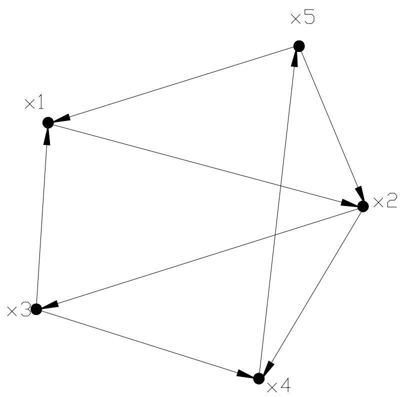


Рисунок 2.2 – Орграф, имеющий гамильтонов путь

Путь, содержащий все дуги графа и при том только по одному разу, называется ***эйлеровым.***

На рис. 2.3 представлен орграф с эйлеровым путем $x_1x_2x_3x_1x_4x_3$.

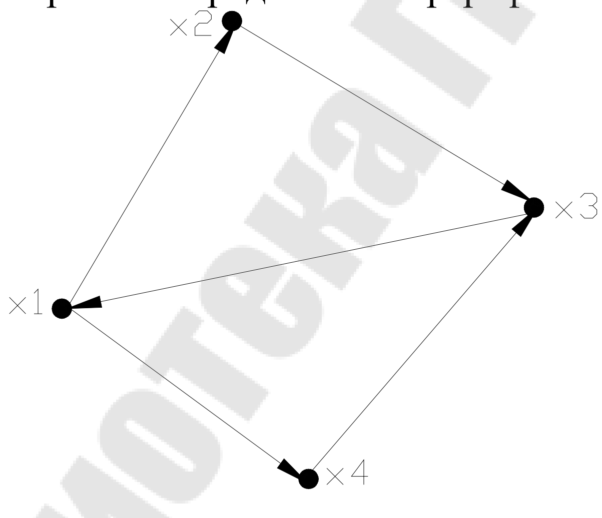


Рисунок 2.3 – Орграф, имеющий эйлеровый путь

Длиной пути $\mu = (u_1, \cdots, u_k)$ называется число $l(\mu)=k$, равное числу дуг, составляющих путь μ. Путь может быть *бесконечным* или *конечным*. В случае бесконечного пути полагаем $l(\mu) = \infty$.

Путь, в котором никакая дуга не встречается дважды, называется ***простым***.

Контур – это конечный путь $\mu = (x_1, \cdots, x_k)$, у которого начальная вершина x_1 совпадает с конечной x_k. При этом контур называется элементарным (гамильтоновым), если все его вершины различны за исключением начальной и конечной, которые совпадают.

На рисунке 2.2 можно выделить, например, контур $x_1x_2x_3x_1$, а также гамильтонов контур $x_1x_2x_3x_4x_5x_1$.

Иногда граф рассматривают без учета ориентации его дуг. В этом случае его называют неориентированным графом. Для неориентированного графа понятия дуга, путь, контур заменяются понятием ребро, цепь, цикл.

Ребро, - это отрезок, соединяющий две вершины.

Цепью называется последовательность ребер.

Циклом называется конечная цепь, у которой начальная и конечная вершины совпадают.

Говорят, что ***граф связан***, если любые две его вершины можно соединить цепью.

Граф сильно связан, если для любых двух вершин x и y $(x \neq y)$ существует путь, идущий из x в y.

Деревом называется конечный связанный неориентированный граф, не имеющий циклов. Начальная вершина называется ***корнем*** дерева (см рис.2.4).

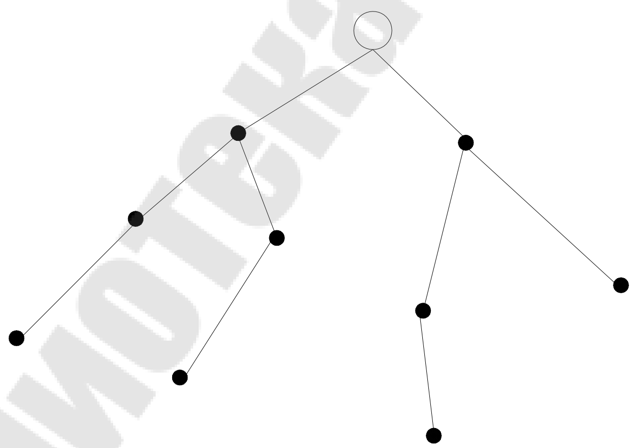


Рисунок 2.4 – Неориентированный граф - дерево

Дерево можно построить, последовательно добавляя ребра в его вершинах. Простейшее дерево состоит из двух вершин, соединенных ребром. Каждый раз, когда мы добавляем еще одно ребро в конце его прибавляется также и вершина. Следовательно, дерево с n вершинами имеет n-1 ребро.

Несвязанный граф, представляющий объединение деревьев, называется ***лесом***.

Матричные способы задания графов

До сих пор мы изображали графы рисунками. При большом числе вершин и дуг (ребер) рисунок графа теряет наглядность. В таких случаях для задания графов и работы с ними используются матрицы. Особенно удобно использовать матричный эквивалент графа при его исследовании на ЭВМ.

Дадим два определения.

Две вершины x и y **являются смежными**, если они различны и если существует дуга, идущая из x в y.

Дуга u_i ***называется инцидентной*** вершине x, если она заходит в эту вершину или исходит из нее.

Обозначим через $x_1, x_2, \cdots, x_n$ вершины графа. Введем далее числа

$$s_{i,j} = \begin{cases} -1, \text{ если } u_j \text{ исходит из } x_i \\ 1, \text{ если } u_j \text{ заходит в } x_i \\ 0, \text{ если } u_j \text{ не инцидентна } x_i \end{cases}$$

Матрица $S = [s_{i,j}]$ порядка nxm называется ***матрицей инциденций для дуг графа.***

В случае неориентированного графа элементами матрицы будут 1 и 0.

Рассмотрим пример 2.1**.** На рисунке 2.5 представлен орграф, а в таблице 2.1 матрица инцинденций для дуг данного орграфа.

Таблица 2.1– Матрица инцинденций для дуг орграфа

	u1	u2	u3	u4	u5	u6	u7
x1	-1	0	0	0	-1	-1	0
x2	1	-1	0	-1	0	0	0
x3	0	0	0	1	1	0	-1
x4	0	1	1	0	0	0	0
x5	0	0	-1	0	0	1	1

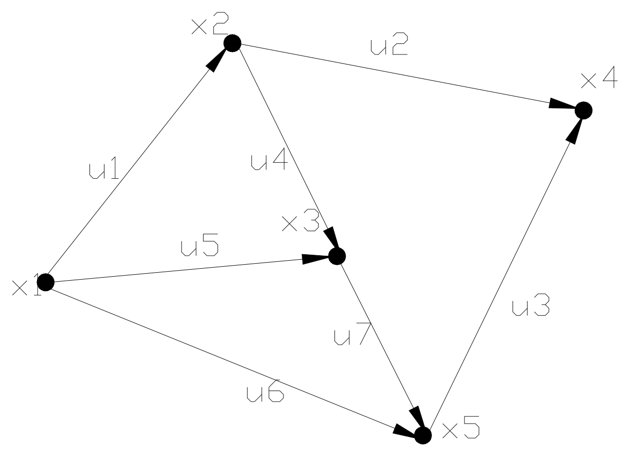


Рисунок 2.5 – Пример 2.1 орграфа

Пример для неориентированного графа тот же, только на месте -1 будут находится 1.

Строки матрицы инцинденций называют *векторами инцинденций*.

Как для орграфов, так и неориентированных графов можно определит матрицу смежности вершин.

Матрица смежности вершин орграфа *G*, содержащего *n* вершин, это квадратная матрица $P[p_{i,j}]$ *n*-го порядка, строки и столбцы которой соответствуют вершинам графа *G*. Элементы $p_{i,j}$ матрицы равны числу дуг, идущих из *i*-ой вершины в *j*-ую. Если орграф не содержит параллельных дуг, то матрица состоит из 1 и 0.

В случае неориентированного графа ему вместе с ребром (x_i, x_j) принадлежит и ребро (x_j, x_i), поэтому матрица будет симметричной. Справедливо и обратное утверждение: любой симметричной матрице с целыми неотрицательными элементами можно поставить в соответствие граф.

Пример2.2 Построение матрицы смежности вершин для орграфа (см рис. 2.6 и табл. 2.2) и неориентированного графа (см табл. 2.3).

Таблица 2.2 – Матрица смежности вершин орграфа

	x1	x2	x3	x4
x1	1	0	1	0
x2	1	0	0	0
x3	0	1	0	1
x4	0	1	0	1

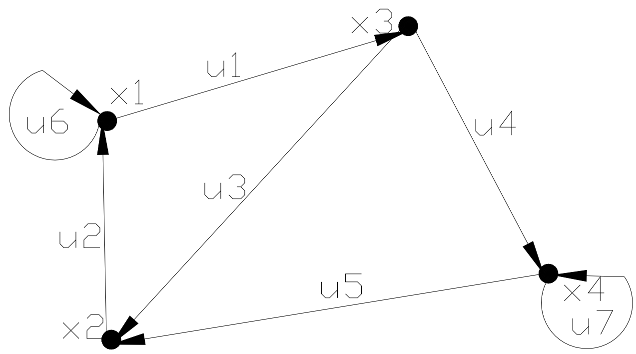


Рисунок 2.6 – Пример 2.2 орграфа

Таблица 2.3– Матрица смежности вершин неориентированного графа

	x1	x2	x3	x4
x1	1	1	1	0
x2	1	0	1	1
x3	1	1	0	1
x4	0	1	1	1

По матрице смежности вершин легко определить полустепени захода и исхода вершин, а значит, и степень вершины. ***Полустепень захода*** $P^{+}(x_i)$ вершины x_i равна сумме элементов i-ого столбца; ***полустепень исхода*** $P^{-}(x_i)$ - сумме элементов i-й строки.

Вычислим полустепени захода и исхода для каждой вершины орграфа (см рис. 2.6), матрица смежности которого представлена в таблице 2.2, получим таблицу 2.4.

Таблица 2.4 – Матрица смежности вершин орграфа и полустепени заходов и исходов

	x1	x2	x3	x4	$P^{-}(x_i)$
x1	1	0	1	0	2
x2	1	0	0	0	1
x3	0	1	0	1	2
x4	0	1	0	1	2
$P^{+}(x_i)$	2	2	1	2	

Например, для вершины x_3: $P^{+}(x_3) = 1, P^{-}(x_3) = 2$, так что $P\ \ (x_3) = 3$.

В дальнейшем используется понятие непосредственного предшествования дуг. Будем говорить, что *дуга* u_i *непосредственно предшествует дуге* u_j если конец дуги u_i является началом дуги u_j

Граф G можно задать *матрицей смежности дуг (ребер)*. Матрица смежности дуг орграфа - это квадратная матрица $[q_{i,j}]$ m-го порядка (m-число дуг). Строки и столбцы матрицы соответствуют дугам (ребрам) графа. Элементы $q_{i,j}$ равны 1, если дуга u_i непосредственно предшествует дуге u_j, и 0 в остальных случаях.

Матрицей смежности ребер неориентированного графа является матрица m-го порядка (m-число ребер) с элементами $q_{i,j}$, равными 1, если ребра u_i и u_j смежны, и 0 в остальных случаях.

Пример 2.3 Построение матрицы смежности дуг для орграфа (см рис. 2.7 и табл. 2.5) и ребер для неориентированного графа (см рис. 2.8 и табл. 2.6).

Таблица 2.5 – Матрица смежности дуг орграфа

	u1	u2	u3	u4	u5
u1	0	1	0	0	0
u2	0	0	0	1	0
u3	0	0	0	1	0
u4	0	0	0	0	0
u5	0	0	0	0	0

Таблица 2.6– Матрица смежности ребер орграфа

	u1	u2	u3	u4	u5
u1	0	1	1	0	1
u2	1	0	1	1	0
u3	1	1	0	1	1
u4	0	1	1	0	1
u5	1	0	1	1	0

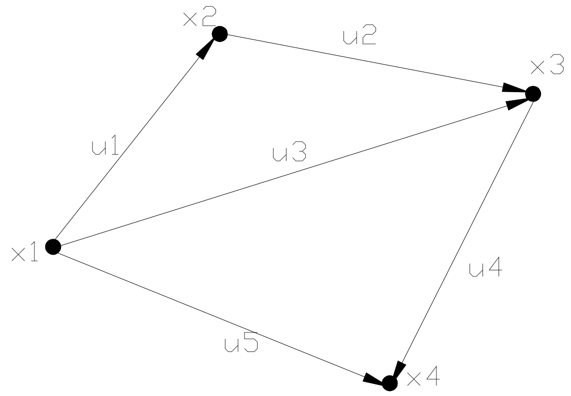


Рисунок 2.7– Пример орграфа 2.3

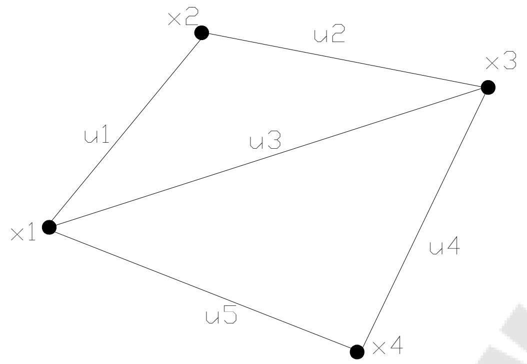


Рисунок 2.8 –Пример неориентированного графа

Упорядочение элементов орграфа. Алгоритм Фалкерсона

Расчеты в задачах, связанных с графами упрощаются, если их элементы упорядочены. Здесь потребуется понятие предшествования вершин.

Говорят, что вершина x_i предшествует вершине x_j, если существует путь из x_i в x_j, тогда x_i называют предшествующей вершине x_j, а x_j – последующей за x_i.

Под упорядочением вершин связного орграфа без контуров понимают такое разбиение его вершин на группы, при котором:

1. Вершины первой группы не имеют предшествующих, а вершины последней – последующих.
2. Вершины любой другой группы не имеют предшествующих в следующей группе.
3. Вершины одной и той же группы дугами не соединяются.

Можно показать, что описанное разбиение всегда возможно.

Аналогичным образом вводится понятие *упорядочения дуг.*

В результате упорядочения элементов получают граф, изоморфный данному. **Изоморфны** - содержится одна и та же информация. На рисунке 2.9 а) и б) изображен по существу один и тот же граф.

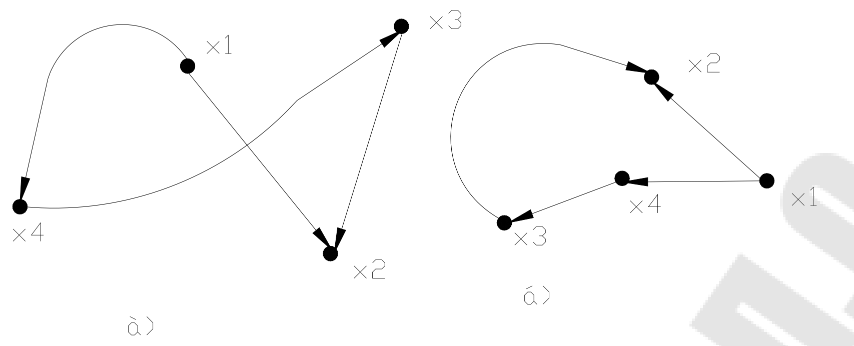


Рисунок 2.9 – Варианты изображения одного и того же орграфа

Упорядочение элементов выполняют графическим или матричным способом. Графический способ упорядочения вершин (*алгоритм Фалкерсона*) состоит из следующих шагов.

1. Находят вершины графа, в которые не входит ни одна дуга. Они образуют первую группу. Нумеруют вершины группы в натуральном порядке 1, 2,… При этом присвоение номеров вершинам внутри группы может быть сделано не единственным образом, что не имеет значения.

2. Мысленно вычеркивают все пронумерованные вершины и дуги, из них выходящие. В получившемся графе найдется по крайней мере одна вершина, в которую не входит ни одна дуга. Этой вершине, входящей во вторую группу, присваивают очередной номер и т. д. Этот шаг повторяют до тех пор, пока все вершины не будут упорядочены (пронумерованы).

Аналогичным образом упорядочивают дуги орграфа. Сначала находят дуги, не имеющие непосредственно предшествующих (они образуют I группу). После вычеркивания дуг I группы в оставшемся графе вновь выделяют дуги, не имеющие непосредственно предшествующих (они образуют II группу). И так до тех пор, пока все дуги не будут разбиты на группы. В заключение упорядоченным дугам присваивают новые обозначения с индексами 1,2,…

<u>Пример 2.4</u> Упорядочить вершины данного графа и построить изоморфный граф см рис 2.10-2.11).

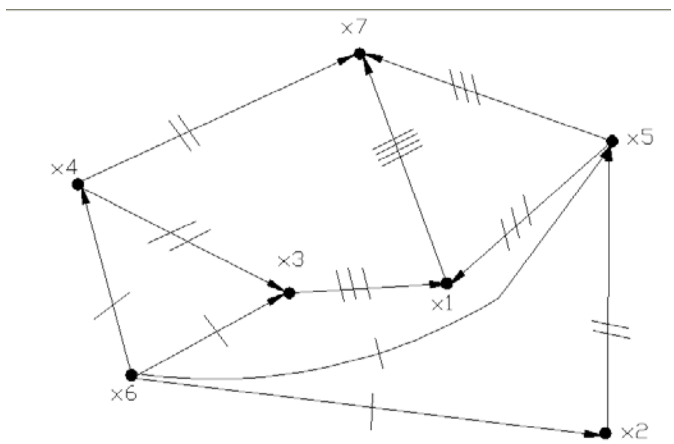


Рисунок 2.10 –Связанный орграф без контуров

<u>*Решение.*</u> Анализируя данный граф (рис. 2.10), замечаем, что в вершину x_6 не входит ни одна дуга. Следовательно, x_6 не имеет предшествующих, а потому относим ее к I группе. Больше подобных вершин на графе нет. Исключаем из рассмотрения вершину x_6 и дуги, и дуги из нее исходящие (на рис. 2.10, эти дуги отмечены одной черточкой - первое вычеркивание). В оставшемся графе опять находим вершины, в которые не заходит ни одна дуга. Таковыми будут x_2 и x_4 . Они образуют II группу. Выполняем второе вычеркивание и т. д. (см. рис. 2.10).

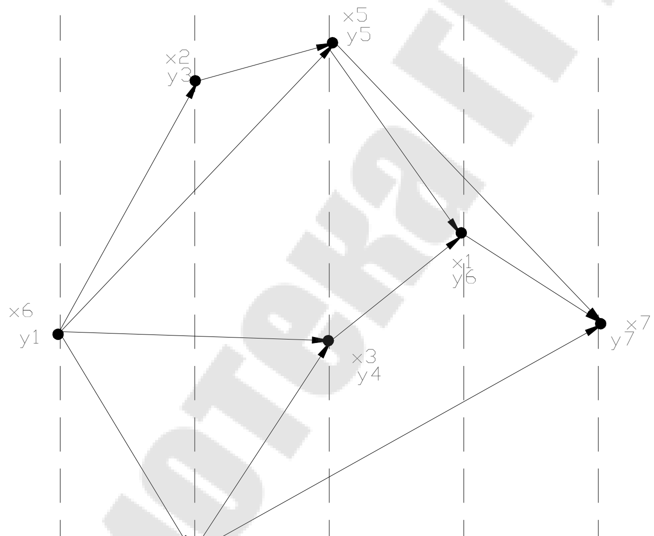


Рисунок 2. 11 – Построение изоморфного графа

На рис. 2.11 для наглядности проведены вертикали, соответствующие группам разбиения, на которых последовательно отмечались точки: сначала x_6, затем x_2 и x_4 и т. д. В заключение эти точки соединяли дугами так, как на данном графе, и получили изоморфный граф с упорядоченными вершинами. Остаётся перенумеровать его вершины в натуральном порядке.

Матричный способ упорядочивания вершин на примере графа

Рассмотрим матричный способ упорядочения вершин для рисунка 2.10. Сначала построим матрицу смежности вершин для орграфа (см рис.2.10) и вычислим полустепени захода вершин, получим таблицу 2.7.

Таблица 2.7 – Матрица смежности вершин для рис. 2.10

	x1	x2	x3	x4	x5	x6	x7
x1	0	0	0	0	0	0	1
x2	0	0	0	0	1	0	0
x3	1	0	0	0	0	0	0
x4	0	0	1	0	0	0	1
x5	1	0	0	0	0	0	1
x6	0	1	1	1	1	0	0
x7	0	0	0	0	0	0	0
$P^{+}(x_i)$	2	1	2	1	2	0	3

Обозначим через Vx1, Vx2, …, Vx6 векторы, являющиеся строками матрицы смежности. Вычислим компоненты вектора V1=Vx1+Vx2+…+Vx6 и припишем их внизу к матрице смежности. Обратим внимание на то, что V1 – это не что иное как полустепени захода вер $P^{+}(x_i)$.

Таблица 2.8

	x1	x2	x3	x4	x5	x6	x7	Группы
x1	0	0	0	0	0	0	1	
x2	0	0	0	0	1	0	0	
x3	1	0	0	0	0	0	0	
x4	0	0	1	0	0	0	1	
x5	0	0	0	0	0	0	0	
x6	0	1	1	1	1	0	0	I
x7	0	0	0	0	0	0	0	
V1	2	1	2	1	2	0	3	
V2	2	0	1	0	1	-	3	

В вершину х6 не заходит ни одна дуга, она образует первую группу. Исключаем в дальнейшем из рассмотрения вершину *x6* и дуги (*x6, x4*), (*x6, x3*), (*x6, x2*), (*x6, x5*). Вычислим компоненты вектора V2=V1-Vx6 и запишем во второй дополнительной строке (см табл. 2.8).

В этой таблице в строке V2 появилось два нуля, соответствующие столбцам *x2* и *x4*. Эти нули свидетельствуют о том, что в вершины *x2* и *x4* не заходит ни одна дуга, т.е. вершины *x2* и *x4* не имеют предшествующих в графе без вершины *x6* и дуг (*x6, x4*), (*x6, x3*), (*x6, x2*), (*x6, x5*). Следовательно, вершины *x2* и *x4* образуют вторую группу.

Затем находим вектор V3=V2-(Vx2+Vx4) с двумя нулевыми компонентами – *x3* и *x5* (см табл. 2.9). Вершины *x3* и *x5* образуют третью группу.

Таблица 2.9

	x1	x2	x3	x4	x5	x6	x7	Группы
x1	0	0	0	0	0	0	1	
x2	0	0	0	0	1	0	0	II
x3	1	0	0	0	0	0	0	
x4	0	0	1	0	0	0	1	II
x5	0	0	0	0	0	0	0	
x6	0	1	1	1	1	0	0	I
x7	0	0	0	0	0	0	0	
V1	2	1	2	1	2	0	3	
V2	2	0	1	0	1	-	3	
V3	2	-	0	-	0	-	2	

Таблица 2.10

	x1	x2	x3	x4	x5	x6	x7	Группы
x1	0	0	0	0	0	0	1	
x2	0	0	0	0	1	0	0	II
x3	1	0	0	0	0	0	0	III
x4	0	0	1	0	0	0	1	II
x5	0	0	0	0	0	0	0	III
x6	0	1	1	1	1	0	0	I
x7	0	0	0	0	0	0	0	
V1	2	1	2	1	2	0	3	
V2	2	0	1	0	1	-	3	
V3	2	-	0	-	0	-	2	
V4	0	-	-	-	-	-	1	

Затем находим вектор V4=V3-(Vx3+Vx5) с одной нулевой компонентой– *x1* (см табл. 2.10). Вершина *x1* образует четвертую группу.

Затем находим вектор V5=V4 - Vx1 с одной нулевой компонентой– x7 (см табл. 2.11). Вершина *x7* образует последнюю пятую группу.

Таблица 2.11

	x1	x2	x3	x4	x5	x6	x7	Группы
x1	0	0	0	0	0	0	1	IV
x2	0	0	0	0	1	0	0	II
x3	1	0	0	0	0	0	0	III
x4	0	0	1	0	0	0	1	II
x5	0	0	0	0	0	0	0	III
x6	0	1	1	1	1	0	0	I
x7	0	0	0	0	0	0	0	V
V1	2	1	2	1	2	0	3	
V2	2	0	1	0	1	-	3	
V3	2	-	0	-	0	-	2	
V4	0	-	-	-	-	-	1	
V5	-	-	-	-	-	-	0	

В результате упорядочения вершин получаем орграф, изоморфный данному (см рис.2.11).

При необходимости надо перенумеровать вершины, начав с вершины первой группы. Внутри группы вершины нумеруются в произвольном порядке.

Если орграф упорядочен правильно и вершины перенумерованы в натуральном порядке, то стрелки всех дуг направлены вправо, а номера начал дуг, меньше номеров их концов.

ТЕМА 3. ПРАКТИЧЕСКИЕ ПРИЛОЖЕНИЯ ТЕОРИИ ГРАФОВ В МАШИНОСТРОЕНИИ

Лекция 3. Задача о кратчайшем пути. Нахождение кратчайшего пути в графе с ребрами единичной длины. Нахождение кратчайшего пути в графе с ребрами произвольной длины. Нахождение длиннейшего пути в графе с ребрами произвольной длины. Формирование технологических операций. Балансировка технологического маршрута. Оснащение обрабатывающего центра.

Задача о кратчайшем пути

В практических приложениях имеет большое значение задача о нахождении кратчайшего пути между двумя вершинами связанного графа. К такой задаче сводятся многие задачи выбора наиболее экономического (с точки зрения расстояния или времени, или стоимости) маршрута на имеющейся карте дорог, многие задачи выбора наиболее экономичного способа перевода динамической системы из одного состояния в другое и т.д.

В математике разработан ряд методов решения подобных задач.

Задача о кратчайшем пути на графе в общем виде может быть сформулирована следующем образом.

Дан граф $G=\{X, U\}$. Каждому ребру этого графа приписано некоторое число $l(u) \geq 0$, называемое длиной ребра. В частных случаях $l(u)$ может быть расстоянием между вершинами, соединяемыми ребром u, временем или стоимостью проезда по этому ребру и т.п. При этом любая цепь μ будет характеризоваться длиной

$$l(\mu) = \sum_{u \in \mu} l(u) \qquad (3.1)$$

Требуется для двух произвольных вершин a и b графа G найти путь μ_{ab}, причем такой, чтобы его полная длина была наименьшей.

Сначала рассмотрим правило для решения задачи частного вида, когда длина каждого ребра равна 1.

Нахождение кратчайшего пути в графе с ребрами единичной длины

Иногда приходится иметь дело с графами, ребра которых имеют одинаковую длину, принимаемую за единицу. Вершины такого графа обычно представляют собой состояние некоторой системы, в которой с точки зрения все переходы, делаемые за один шаг, эквивалентны.

Перейдем к задаче нахождения в графе кратчайшего пути, соединяющего начальную вершину с конечной.

Общее правило для нахождения кратчайшего пути в графе состоит в том, чтобы каждой вершине x_i приписать индекс λ_i, равный длине кратчайшего пути из данной вершины в конечную. Приписывание индексов в случае графа с ребрами единичной длины производится в следующем порядке.

1. Конечной вершине x_0 приписывается индекс 0.
2. Всем вершинам, из которых идет ребро в конечную вершину, приписывается индекс 1.
3. Всем вершинам, еще не имеющим индексов, из которых идет ребро в вершину с индексом λ_i, приписывается индекс $\lambda_i + 1$.

Этот процесс продолжается до тех пор, пока не будет помечена начальная вершина. По окончании разметки индекс у начальной вершины будет равен длине кратчайшего пути. Сам кратчайший путь найдем, если будем двигаться из начальной вершины в направлении убывания индексов.

Описанный способ является частным случаем нахождения оптимального решения по методу динамического программирования.

Нахождение кратчайшего пути в графе с ребрами произвольной длины

Задача приписывания вершинам графа числовых индексов усложняется, если ребра графа имеют произвольную длину. Усложнение вызвано тем, что в сложном графе путь, проходящий через наименьшее число вершин, зачастую имеет большую длину, чем некоторые обходные пути.

Процесс приписывания индексов для такого графа заключается в следующем.

1. Каждая вершина x_i помечается индексом λ_i. Первоначально конечной вершине x_0 приписывается индекс $\lambda_0 = 0$. Для остальных вершин предварительно полагаем $\lambda_i = \infty$.
2. Ищем такую дугу (x_i, x_j), для которой $\lambda_j - \lambda_i > l(x_i, x_j)$, и заменяем индекс λ_j индексом $\lambda_j' = \lambda_i + l(x_i, x_j) < \lambda_j$.

Продолжаем этот процесс замены до тех пор, пока остается хотя бы одна дуга, для которой можно уменьшить λ_j.

Отметим два свойства, которыми будут обладать приписанные вершинам индексы.

1. Пусть (x_k, x_s)- произвольное ребро. Для него обязательно выполняется условие
   $\lambda_k - \lambda_s \le l(x_k, x_s)$, так как если бы оно не выполнялось, индекс λ_s нужно было бы уменьшать.
2. Пусть x_p- произвольная вершина. При рассмотренном процессе приписывания индексов индекс λ_p монотонно уменьшается. Пусть x_q- последняя вершина, послужившая для его уменьшения. Тогда $\lambda_p = \lambda_q + l(x_p, x_q)$. Следовательно, для произвольной вершины x_p с индексом λ_p найдется вершина x_q, соединенная ребром с x_p, такая что $\lambda_p - \lambda_q = l(x_q, x_p)$.

Эти свойства позволяют сформулировать следующее правило для нахождения кратчайшего пути.

Пусть $x_n = a$ - начальная вершина с индексом λ_n. Ищем вершину x_{p_1}, такую, что $\lambda_{p_n} - \lambda_{p_1} = l(x_{p_n}, x_{p_1})$, и т.д. до тех пор, пока не дойдем до конечной вершины $x_{p_{k+1}} = x_0 = b$. Путь $\mu(x_n, x_{k_1}, x_{k_s}, x_0)$, длина которого равна λ_n, является кратчайшим.

На рис. 3.1 представлен пример нахождения кратчайшего пути из вершины a в вершину b.

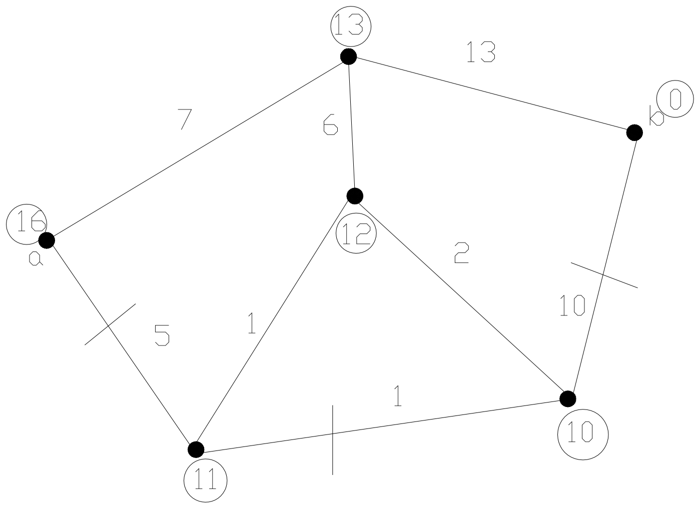


Рисунок 3.1 – Нахождение кратчайшего пути

Нахождение длиннейшего пути в графе с ребрами произвольной длины

Алгоритм нахождения длиннейшего пути представляет собой процесс приписывания индексов для вершин графа и заключается в следующем.

1 Каждая вершина x_i помечается индексом λ_i. Первоначально начальной вершине x_0 приписывается индекс $\lambda_0 = 0$. Для остальных вершин предварительно полагаем $\lambda_i = -\infty$.

2 Ищем такую дугу (x_i, x_j), для которой $\lambda_j - \lambda_i < l(x_i, x_j)$, и заменяем индекс λ_j индексом $\lambda'_j = \lambda_i + l(x_i, x_j) > \lambda_j$.

Продолжаем этот процесс замены до тех пор, пока остается хотя бы одна дуга, для которой можно увеличить λ_j.

Пусть $x_n = b$ - конечная вершина с индексом λ_n. Ищем вершину x_{p_1}, такую, что $\lambda_{p_1} - \lambda_{p_n} = l(x_{p_n}, x_{p_1})$, и т.д. до тех пор, пока не дойдем до начальной вершины $x_{p_{k+1}} = x_0 = a$. Путь $\mu(x_0, ..., x_{k_s}, ..., x_{k_1}, x_n)$, длина которого равна λ_n, является длиннейшим.

Формирование технологических операций

Технологической операцией будем называть последовательность *переходов,* выполняемых одним станком. Например, последовательность переходов $П_1$, $П_2$, $П_3$, $П_4$, $П_5$, выполняемых соответственно станками C_1, C_1, C_2, C_2, C_1 разбивается на три технологические операции: O_1 — переходы $П_1$, $П_2$ на станке C_1 , O_2 — переходы $П_3$, $П_4$ на станке C_2, O_3 — переход $П_5$ на станке C_1. Станок C_1 производит две операции, станок C_2 — одну. При формировании технологических операций такого типа возникает следующая задача.

Дана последовательность переходов $П_1,\dots, П_n$ в обработке некоторой детали. Для каждого перехода заданы варианты выбора станков $C_1, \dots, C_m$, которые могут его выполнить. Требуется так выбрать станки для выполнения переходов, чтобы число получаемых при этом операций было минимально. На рис. 3.2 приведены пример этой задачи и одно из его решений. Покажем сводимость к задаче КРАТЧАЙШИЙ ПУТЬ.

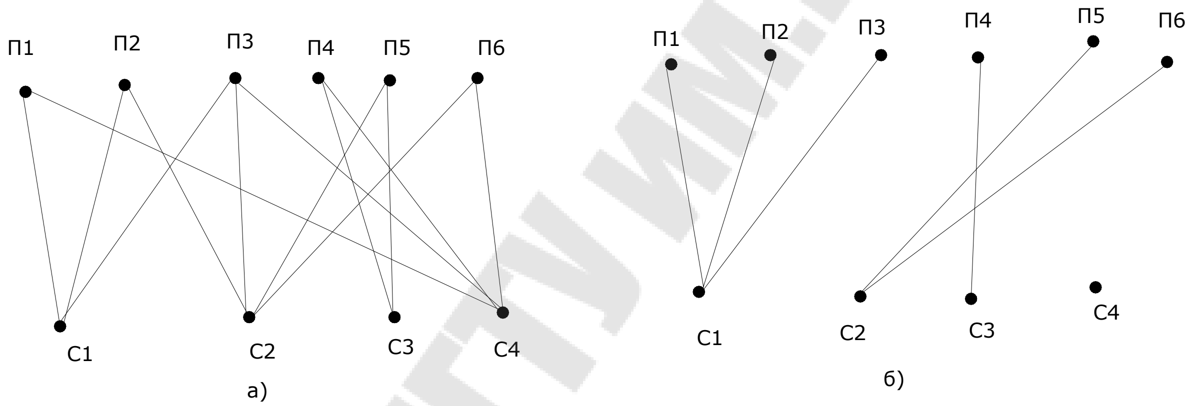


Рисунок 3.2 – Пример задачи формирования технологических операций а) и одного его решения б)

Построим орграф с вершинами s, t и вершинами $v_{i,j}$ для каждого перехода $П_j$. и каждого станка C_i , который может его выполнить. Множество дуг будут составлять дуги $v_{i,j} \to v_{k,j+1}, s \to v_{i,1}, v_{i,n} \to t$, которые следует провести для всех допустимых наборов значений индексов i, j, k. Все дуги вида $v_{i,j} \to v_{i,j+1}$ нагрузим нулем, остальные — единицей. На рис.3.3 приведен пример построения орграфа для примера задачи (см рис. 3.2).

Произвольный путь из истока s в сток t на построенном взвешенном орграфе обладает следующими свойствами:

а) промежуточные вершины пути $v_{i_1,1}, v_{i_2,2}, \dots, v_{i_n,n}$ определяют разбиение последовательности переходов на операции, поскольку, $i_1, i_2, \dots, i_n$ это номера станков, которые их выполняют.

Причем нет никакого варианта разбиения, которому не соответствовал бы некоторый путь из *s* в *t*;

б) каждый набор промежуточных вершин, лежащих вдоль неудлиняемых участков пути с нулевой длиной, определяет одну операцию. Такими наборами в приведенном примере будут $\{v_{1,1}, v_{1,2}, v_{1,3}\}, \{v_{4,4}\}, \{v_{2,5}, v_{2,6}\}$;

в) каждая промежуточная дуга с единичной длиной определяет смену станка. Следовательно, <u>длина пути без единицы равна числу операций.</u>

Из этих свойств вытекает сводимость, поскольку по кратчайшему пути из истока в сток на построенном взвешенном орграфе однозначно находится разбиение последовательности переходов на минимальное число операций. Для приведенного примера получаются три операции: O_1 — переходы $П_1, П_2, П_3$ на станке C_1; O_2 - переход $П_4$ на станке C_4; O_3 — переходы $П_5, П_6$ на станке C_2.

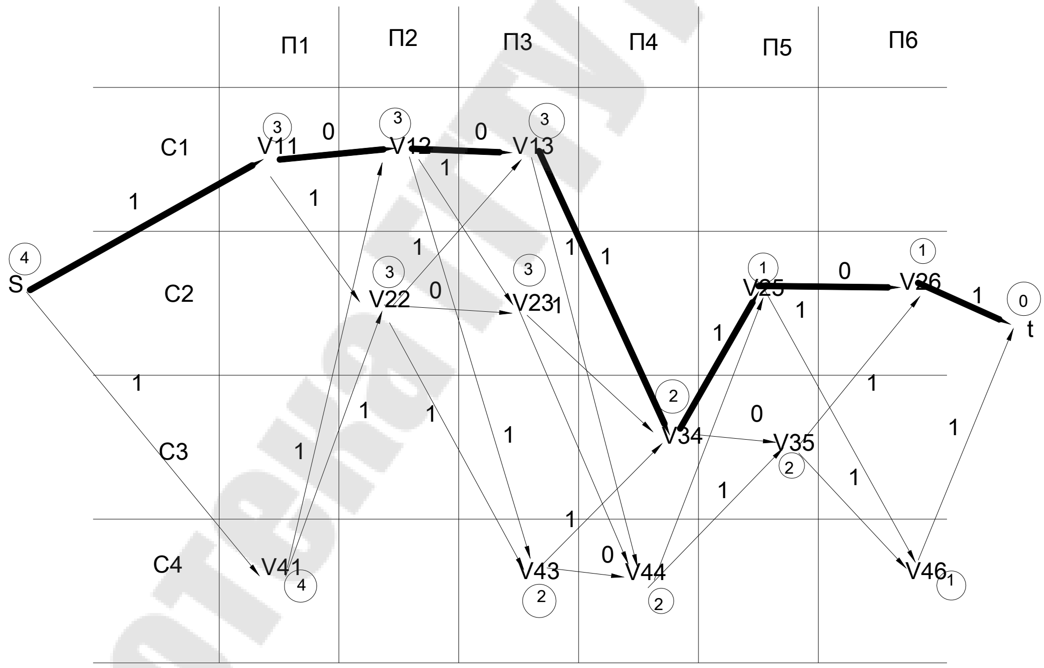


Рисунок 3.3 – Пример формирования технологических операций как поиска Кратчайшего Пути на орграфе

На рис. 3.3 горизонтальные дуги нагружены нулем, остальные — единицей; жирные дуги обозначают кратчайший путь, соответствующий решению *(б)* на рис. 3.2; длина пути без единицы равна числу операций, т е. трем операциям:

- операция O_1 включает переходы $П_1$, $П_2$, $П_3$ и выполняется на станке C_1;
- операция O_2 включает переход $П_4$ и выполняется на станке C_3;
- операция O_3 включает переходы $П_5$, $П_6$ и выполняется на станке C_2.

Балансировка технологического маршрута

Следующая производственная задача, сводимая к задаче поиска пути на графе, возникает при выборе технологических маршрутов обработки деталей в гибкой производственной системе (ГПС).

Пусть на вход некоторой ГПС поступила партия одинаковых деталей для изготовления изделий одной номенклатуры. Задана установленная для этой номенклатуры последовательность технологических операций. Для каждой операции определены: допустимые назначения на станки ГПС и время ее выполнения каждым подходящим станком. Время выполнения операции может зависеть от станка, который ее выполняет. Известно время транспортировки детали от одного станка к другому. Требуется так назначить операции на станки, чтобы получаемый при этом технологический маршрут прохождения станков был *сбалансирован* (т. е. длительности обработки и транспортировки детали на всех участках маршрута, по возможности, выравнены). Балансировка технологического маршрута приводит к наиболее равномерной загрузке оборудования ГПС.

Пусть $O_1, O_2, ..., O_n$— последовательность операций, а $C_1, C_2, ..., C_m$— станки ГПС. Построим орграф с вершинами *s, t* и вершинами $u_{i,j}, v_{i,j}$ для каждой операции O_j и каждого станка C_i, который может ее выполнить. Вершины соединим дугами $s \to u_{i,1}, v_{i,n} \to t, u_{i,j} \to v_{i,j}, v_{i,j} \to u_{k,j+1}$ для всех допустимых наборов значений *i, j, k.*

Все дуги вида $u_{i,j} \to v_{i,j}$ нагрузим временем выполнения операции O_j станком C_i, все дуги вида $v_{i,j} \to u_{k,j+1}, i \neq k$,—

временем транспортировки от станка C_i; к станку C_k, а все дуги вида $s \to u_{i,1}, v_{i,n} \to t, v_{i,j} \to u_{i,j+1}$ — нулем.

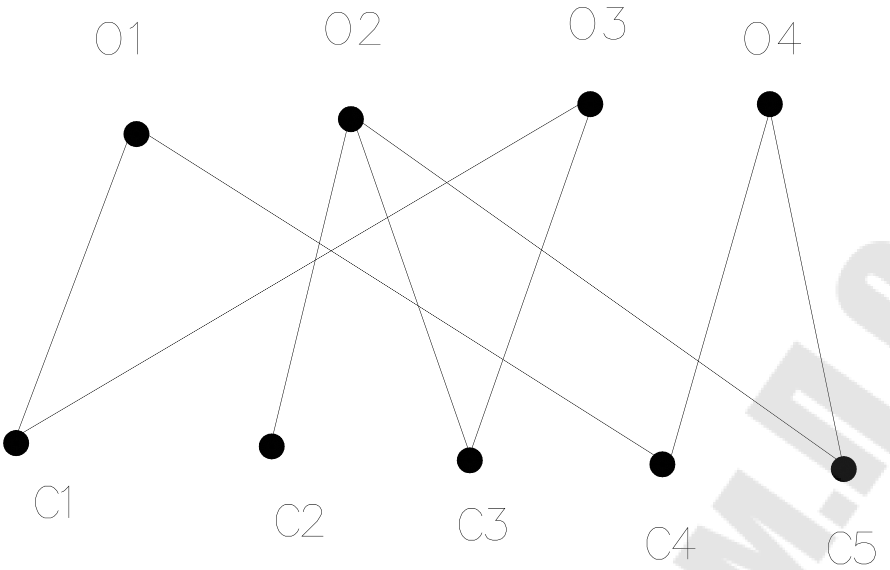


Рисунок 3.4 – Пример допустимых назначений операций на станки

Рассмотрим пример построения орграфа при *n=4,* *m*=5. Допустимые назначения операций на станки определены ребрами двудольного графа на рис. 3.4, а время выполнения операций подходящими станками и время транспортировки представлены в таблицах 3.1 и 3.2. Клетки, соответствующие недопустимым назначениям операции на станки, не заполнены. На рис. 3.5 изображен соответствующий орграф.

Произвольный путь на построенном орграфе из s в *t* проходит через промежуточные вершины $u_{i_1,1}, v_{i_1,1}, u_{i_2,2}, v_{i_2,2}, ..., u_{i_n,n}, v_{i_n,n}$ и обладает следующими свойствами:

а) он определяет вариант допустимого назначения операций на станки, поскольку $i_1, i_2, ..., i_n$- это номера станков, которые должны выполнять операции $O_1, O_2, ..., O_n$. Причем нет ни одного варианта допустимого назначения операций, которому не соответствовал бы некоторый путь из *s* в *t.*

б) соответствующий путь маршрут проходит станки в последовательности $C_{i_1}, C_{i_2}, ..., C_{i_n}$. Самая длинная дуга пути определяет самый длительный по времени процесс на маршруте. Если такой дугой окажется $u_{i_k,k} \to v_{i_k,k}$, то самым длительным процессом будет операция O_k, если $v_{i_k,k} \to u_{i_{k+1},k+1}$ то — транспортировка от станка C_{i_k} к станку $C_{i_{k+1}}$

Сбалансированному маршруту поставим в соответствие путь из *s* в *t* с наиболее короткой длиннейшей дугой. Тем самым время наиболее длительного процесса на маршруте мы сделаем по возможности меньше и приблизим его к времени остальных. Такая интерпретация критерия балансировки окончательно сводит исходную задачу к задаче ТОНЧАЙШИЙ ПУТЬ.

Существуют следующие две модификации задачи балансировки. В первой требуется найти самый производительный, а не сбалансированный маршрут. Самым производительным является маршрут, который деталь проходит за минимальное время. Такая задача, очевидно, сводится к задаче КРАТЧАЙШИЙ ПУТЬ. Во второй модификации задано не время выполнения операций и транспортировки, а их скорость. Такая задача сводится к исходной заменой всех скоростных параметров *P* на временные *1/P*. Однако можно поступить иначе. Сбалансированным можно считать маршрут, вдоль которого деталь проходит с наибольшей скоростью. Эта скорость будет определяться самым непроизводительным станком или самым низкоскоростным транспортным средством на маршруте. Им будет соответствовать самая короткая дуга на широчайшем пути из *s* в *t,* если все нулевые длины дуг в указанном взвешенном орграфе заменить на очень большие (превышающие практически возможные). Отсюда вытекает сводимость второй модификации к задаче ШИРОЧАЙШИЙ ПУТЬ.

Таблица 3.1

	C1	C2	C3	C4	C5
O1	3			4	
O2		2	5		6
O3	1		4		
O4				5	3

Таблица 3.2

	C1	C2	C3	C4	C5
C1	0	7	9	5	3
C2	7	0	8	4	4
C3	9	8	0	3	6
C4	3	4	3	0	5
C5	3	4	6	5	0

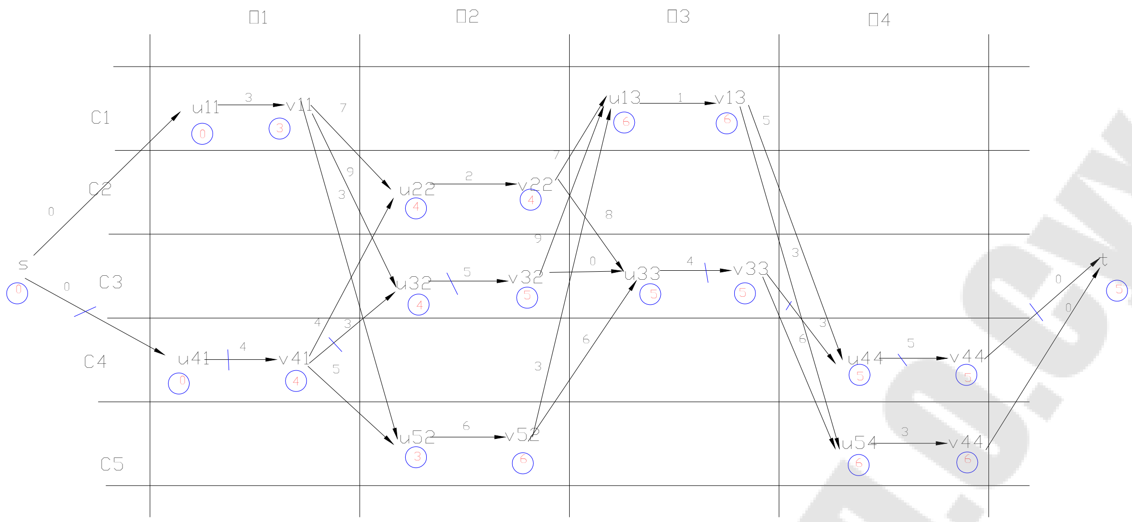


Рисунок 3.5– Орграф для задачи балансировки с исходными данными на рис.3.4 и табл. 3.1, табл. 3.2.

Под вершинами указана тонкость тончайшего пути, ведущего к ним из *s*; штрих - линией отмечен тончайший путь из *s* в *t* с тонкостью 5.

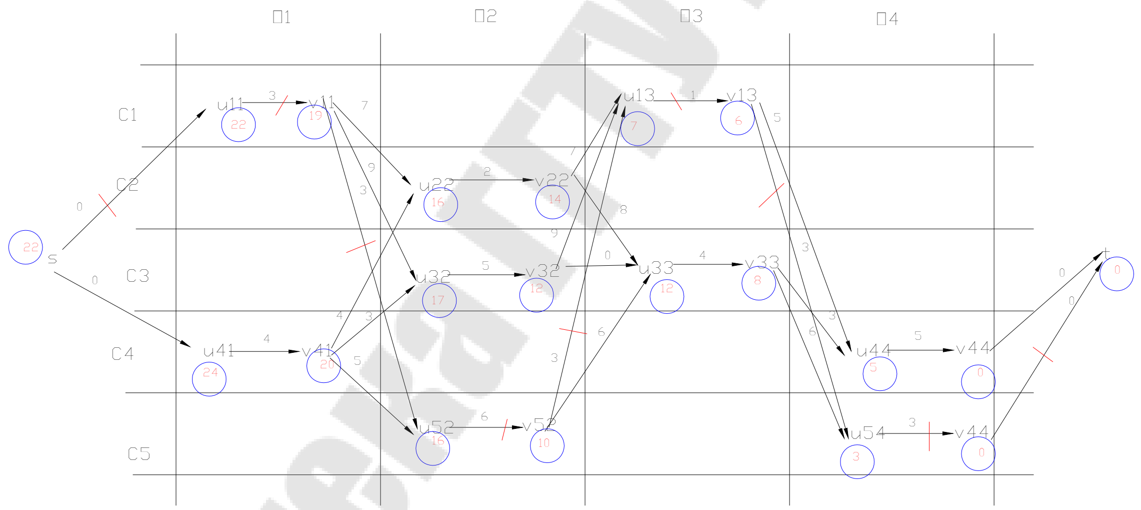


Рисунок 3.6- Орграф для задачи самый Производительный маршрут с исходными данными на рис.3.4 и табл. 3.1, табл. 3.2.

Минимальное время, за которое деталь проходит деталь равно 22 –{О1 – С1; О2 – С5; О3 – С1; О4 – С5}.

Оснащение обрабатывающего центра

К задачам поиска путей на графах сводятся многие производственные задачи. Типичными примерами здесь могут служить задачи выбора маршрутов транспортных роботов и маршрутов прохождения деталей по станкам, поиска критических путей в сетевых графиках проектов и вообще всякого рода «маршрутные» задачи. Сводимость в подобных случаях удается в силу явного структурного сходства задач. В этом разделе представлен пример менее очевидной сводимости.

Пусть имеется станок типа «обрабатывающий центр», магазин которого может оснащаться инструментами различных типов. Для инструмента каждого типа известно, сколько гнезд он занимает в магазине и каково время его работы. В магазине станка можно размещать несколько экземпляров инструмента одного типа. Требуется так оснастить станок инструментами, чтобы общее время их последовательной работы было максимальным, а емкость магазина была использована полностью.

Математическим описанием этой задачи является следующая задача рюкзачного типа.

ЦЕЛОЧИСЛЕННЫЙ РЮКЗАК С ОГРАНИЧЕНИЕМ - РАВЕНСТВОМ. Для положительных целых чисел $a_1, a_2, ..., a_n, b$ и неотрицательных вещественных чисел $c_1, c_2, ..., c_n$ найти неотрицательные целые значения переменных $x_1, x_2, ..., x_n$, максимизирующие целевую функцию

$$z = c_1 x_1 + ... + c_n x_n$$

при ограничении

$$a_1 x_1 + ... + a_n x_n = b \qquad (3.2)$$

Параметр b можно рассматривать как объем рюкзака, в который можно укладывать предметы n типов, причем — по нескольку экземпляров каждого типа, а параметры a_j, c_j — как объем и стоимость одного предмета j-го типа. Равенство $x_j = k$ означает, что в рюкзак укладывается k экземпляров предмета j-го типа. Рюкзак нужно упаковать предметами так, чтобы его объем был использован полностью, а стоимость — максимальной. Объем рюкзака можно интерпретировать как объем магазина обрабатывающего центра, предметы — как инструменты с временем работы $c_1,...,c_n$ и числами занимаемых гнезд $a_1,...,a_n$ соответственно. Покажем сводимость сформулированной рюкзачной задачи к задаче ДЛИННЕЙШИЙ ПУТЬ.

Без ограничения общности можно считать, что коэффициенты уравнения (3.2) различны. Если же найдутся такие *u* и *v,* что $a_u = a_v, c_u \le c_v$, то переменную x_u можно приравнять нулю, поскольку предметы типа *v* имеют тот же объем и не меньшую стоимость, чем предметы типа *u.* Находя все пары *u* и *v* с указанным свойством и устраняя переменные x_u, можно добиться того, что коэффициенты при оставшихся переменных будут различны. Полученную таким образом задачу будем называть *приведенной*.

Ясно, что решение исходной задачи получается из решения приведенной приравниванием нулю устраненных переменных. Для сводимости необходима именно приведенная задача.

Построим орграф с вершинами 0, 1, 2, …, *b.* Дугу $i \to k$ проведем всякий раз, когда среди индексов коэффициентов 1, 2, ..., *n,* найдется такой индекс *j*, что $a_j = k - i$. Поскольку коэффициенты $a_1,...,a_n$ различны, для каждой дуги существует единственный такой индекс. Это позволяет нагрузить дуги $i \to k$ числами c_j.

Рассмотрим пример построения орграфа для задачи:

$$z = 5x_1 + 2x_2 + 4x_3 + 3x_4,$$

$$3x_1 + x_2 + 2x_3 + 3x_4 = 6.$$

В ограничении коэффициенты при x_1, x_4 одинаковы. Устраняя x_4, получаем приведенную задачу:

$$\begin{aligned} &z = 5x_1 + 2x_2 + 4x_3, \\ &3x_1 + x_2 + 2x_3 = 6. \end{aligned} \qquad (3.3)$$

Орграф, соответствующий этой задаче, изображен на рис. 3.7. Взаимосвязь решений приведенной задачи и путей из вершины 0 в вершину *b* на построенном орграфе устанавливают следующие легко проверяемые утверждения:

а) если $x_1^*,...,x_n^*$ — это произвольное неотрицательное целочисленное решение уравнения (3.2) со значением целевой функции z^*, то в орграфе есть путь, имеющий длину z^* и проходящий через вершины

$$0,\quad a_1 1,\quad a_1 2,\ \ldots,\quad a_1 x_1^*,$$

$$a_1 x_1^* + a_2 1,\quad a_1 x_1^* + a_2 2,\ \ldots,\ a_1 x_1^* + a_2 x_2^*,$$

.....

$$\sum_{j=1}^{n-1} a_j x_j^* + a_n 1,\qquad \sum_{j=1}^{n-1} a_j x_j^* + a_n 2,\ \ldots,\ \sum_{j=1}^{n-1} a_j x_j^* + a_n x_n^* = b$$

б) если $\pi: 0 = v_1 \to v_2,\ \ v_2 \to v_3, v_{k-1} \to v_k = b,$ — это произвольный путь с длиной λ из вершины 0 в вершину b, то ему соответствует решение $x_1^\pi, ..., x_n^\pi$ уравнения (1) со значением целевой функции $z^\pi = \lambda$. Это решение находится следующим образом. Для каждого j = 1, 2, ,..., n нужно просмотреть данный путь, найти все дуги $v_i \to v_{i+1}$ со свойством $a_j = v_{i+1} - v_i$ и посчитать их число. Оно –то и равно x_j^π.

Утверждения «а» и «б» доказывают сводимость рюкзачной задачи к задаче ДЛИННЕЙШИЙ ПУТЬ.

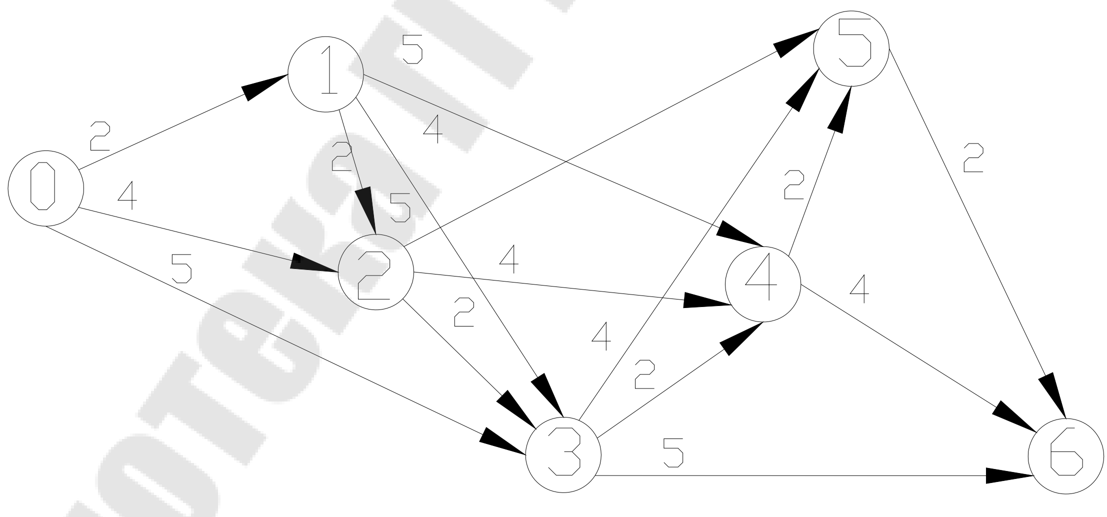


Рисунок 3.7 – Орграф для примера (3.3)

На рис. 3.8 длиннейший путь выделен штрих - линией. Ему соответствует решение задачи (2): $x_1^* = 0, x_2^* = 4, x_3^* = 1,\quad z^* = 12$.

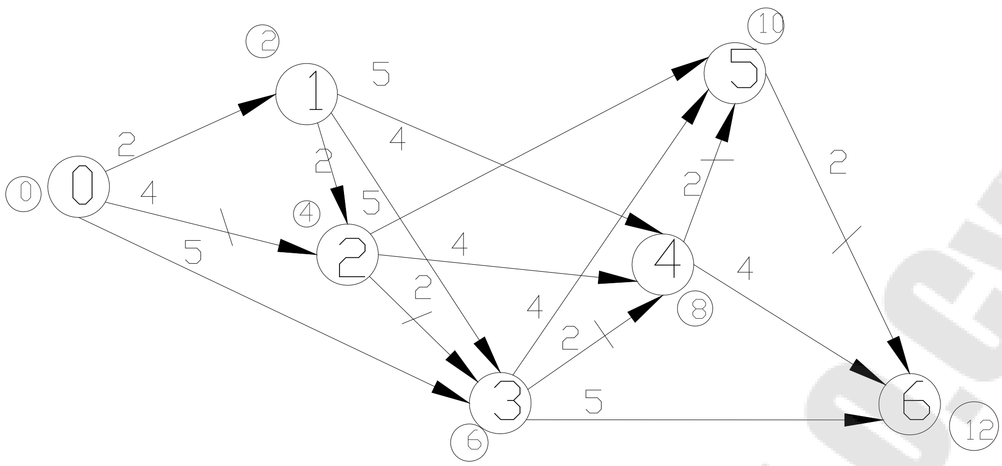


Рисунок 3.8 – Вариант решения 1 для примера (33)

На рис. 3.9 длиннейший путь выделен штрих - линией. Ему соответствует решение задачи (2): $x_1^* = 0, x_2^* = 2, x_3^* = 2, \quad z^* = 12$.

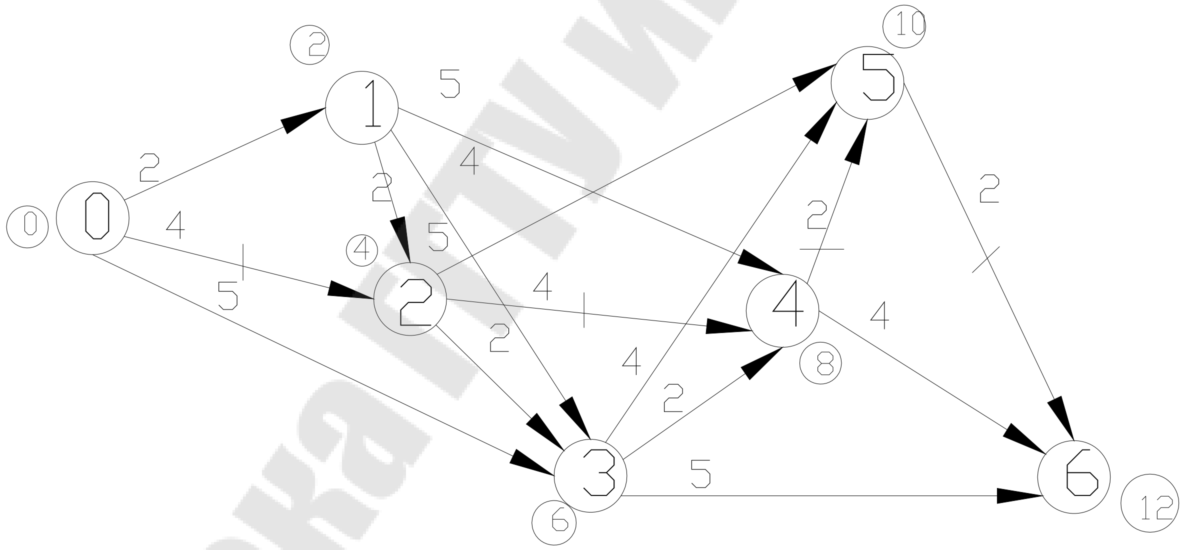


Рисунок 3.9– Вариант решения 2для примера (3.3)

На рис. 3.10 длиннейший путь выделен штрих - линией. Ему соответствует решение задачи (2): $x_1^* = 0, x_2^* = 0, x_3^* = 3, \quad z^* = 12$.

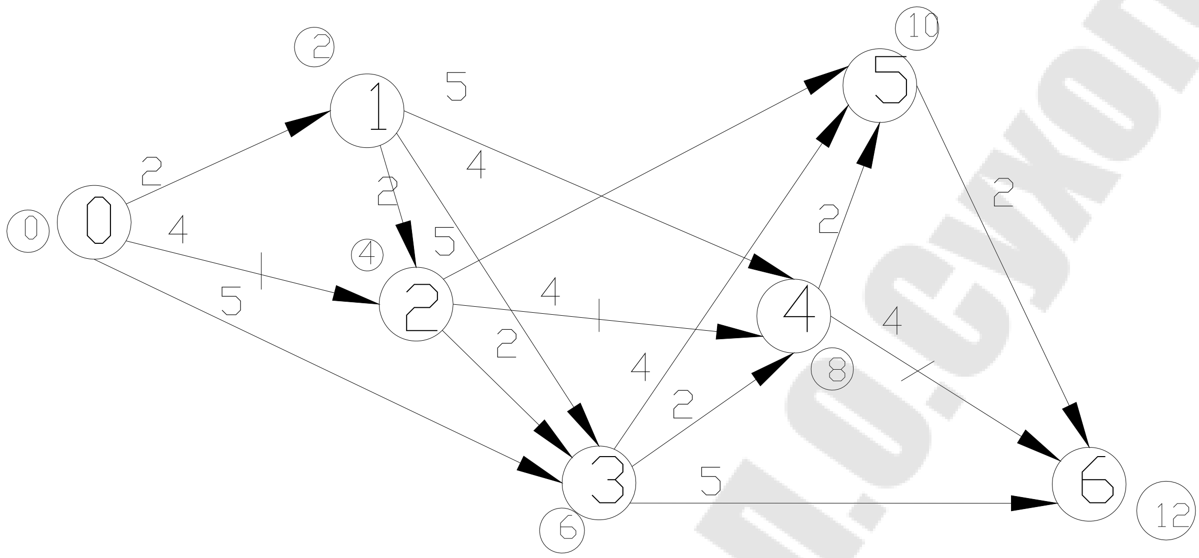


Рисунок 3.10– Вариант решения 3для примера (3.3)

На рис.11 длиннейший путь выделен штрих - линией. Ему соответствует решение задачи (2): $x_1^* = 0, x_2^* = 6, x_3^* = 0, \quad z^* = 12$.

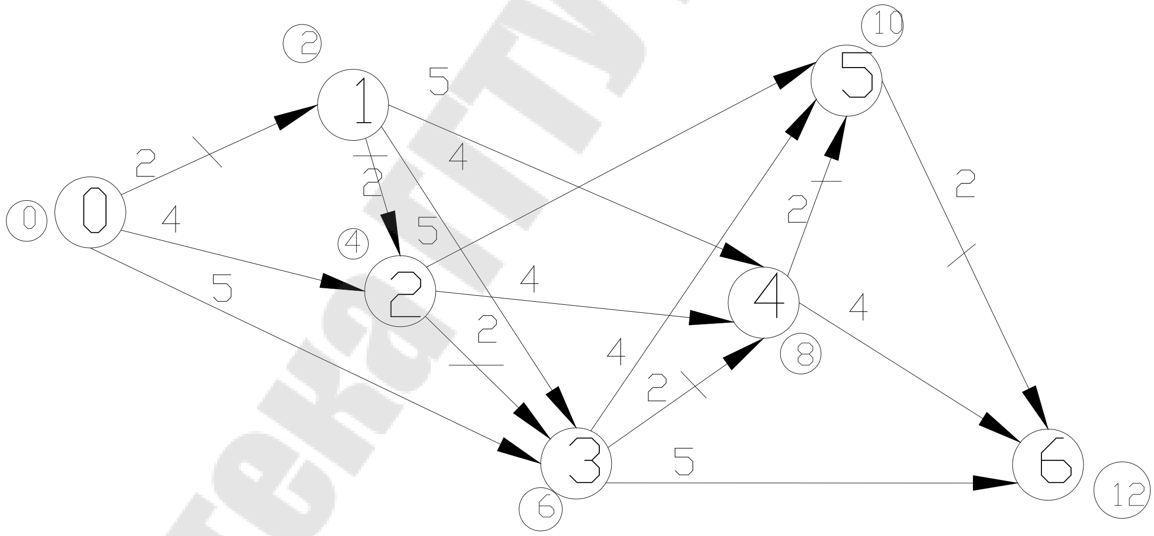


Рисунок 3.11– Вариант решения 4 для примера (3.3)

Аналогичная рюкзачная задача, в которой нужно минимизировать *z,* сводится к задаче КРАТЧАЙШИЙ ПУТЬ. При сводимости нужно лишь изменить правило построения приведенной задачи, устраняя вместо x_u переменную x_v.

В ограничении (3.2) коэффициенты при x_1, x_4 одинаковы. Устраняя x_1, получаем приведенную задачу:

$$z = 2x_2 + 4x_3 + 3x_4,$$
$$x_2 + 2x_3 + 3x_4 = 6. \qquad (3.4)$$

На рис. 3.12 представлен орграф для задачи (3.4).

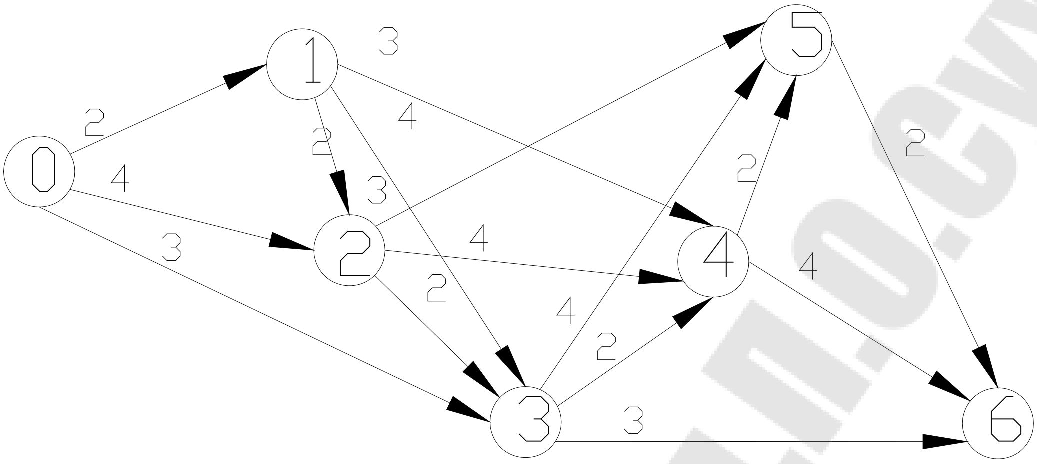


Рисунок 12 – Орграф для примера (3.4)

Для задачи (3.4) есть только один кратчайший путь, проходящий через вершины 0, 3, 6. Ему соответствует решение задачи (3): $x_4^* = 2, x_2^* = 0, x_3^* = 0, \quad z^* = 6$ (см рис.3.13).

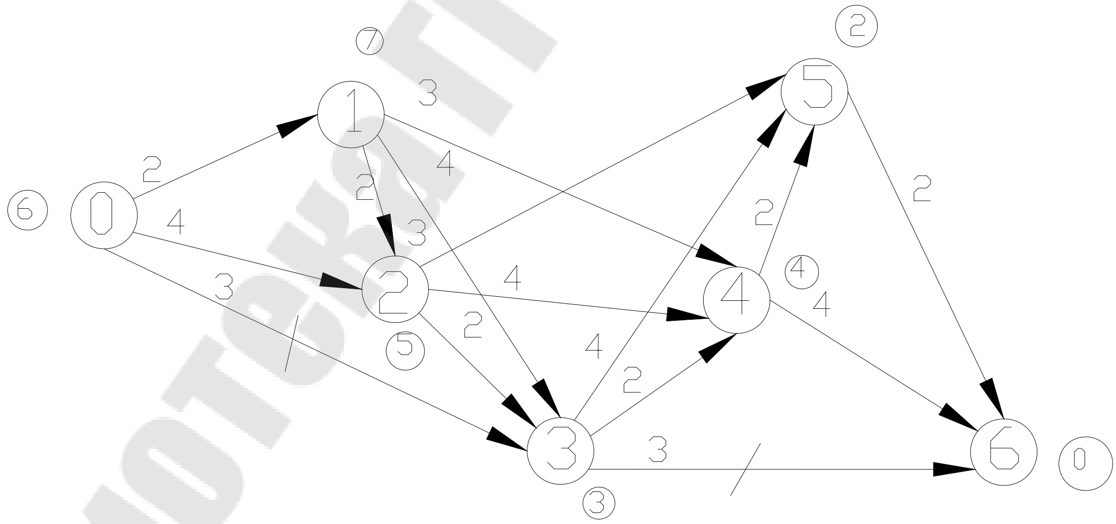


Рисунок 3.13– Вариант решения 1 для примера (3.4)

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Дайте два теоретико-множественных определений графа.
2. Разница между путем, гамильтоновым и эйлеровым путем.
3. Чем отличается контур от гамильтонового контура
4. Дайте определения: ребро, цепь, цикл.
5. Что такое дерево?
6. Как построить матрицу инцинденций для дуг орграфа и ребер неориентированного графа?
7. Как построить матрицу смежности для вершин орграфа и неориентированного графа?
8. Что понимают под упорядочением вершин связного орграфа без контуров?
9. Что такое изоморфный граф?
10. Алгоритм Фалкерсона.
11. Упорядочивание вершин матричным способом.
12. Сформулируйте задачу о кратчайшем пути.
13. Алгоритм поиска КРАТЧАЙШИЙ ПУТЬ
14. Алгоритм поиска ДЛИННЕЙШИЙ ПУТЬ
15. Сформулируйте задачу «Формирование технологических операций».
16. Правило построения орграфа для задачи «Формирование технологических операций».
17. Как определить количество операций с помощью орграфа в задаче «Формирование технологических операций»?
18. Сформулируйте задачу «Балансировка технологического маршрута».
19. Правило построения орграфа для задачи «Балансировка технологического маршрута».
20. Как определить самый сбалансированный маршрут в задаче «Балансировка технологического маршрута»?
21. Как определить самый производительный маршрут в задаче «Балансировка технологического маршрута»?
22. Сформулируйте задачу «Оснащение обрабатывающего центра»
23. Что такое приведенная задача?
24. Правило построения орграфа на max для задачи «Оснащение обрабатывающего центра».
25. Правило построения орграфа на min для задачи «Оснащение обрабатывающего центра».

РАЗДЕЛ 3. ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ПРИ РЕШЕНИИ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

ТЕМА 4. МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Лекция 4. Основные понятия. Формы записи задачи линейного программирования. Некоторые модели задач линейного программирования: задача о выборе оптимальных технологий, задача оптимального использования ресурсов, задача о распределении производственной программы (о размещении заказов или загрузке взаимозаменяемых групп оборудования), задача загрузки невзаимозаменяемых групп оборудования, задачи распределения производственной программы по календарным периодам, задача производственного планирования, задача о смесях, задача о раскрое материалов.

Основные понятия

Общая задача математического программирования формулируется следующим образом. Требуется найти значения n переменных $x_1, x_2, \cdots, x_n$, которые удовлетворяют m требованиям и неравенствам:

$$\varphi_i(x_1, x_2, x_n) \begin{Bmatrix} \le \\ = \\ \ge \end{Bmatrix} b_i \quad (i = 1, 2, m), \tag{4.1}$$

и максимизировать (минимизировать) функцию

$$z = f(x_1, x_2, \cdots, x_n). \tag{4.2}$$

Условия (4.1) называются *ограничениями*, а функция (4.2) – целевой функцией. Предполагается, что функции φ_i и f известны, b_i заданные постоянные. Величины m и n между собой не связаны. На переменные могут налагаться условия неотрицательности, целочисленности.

Набор чисел $x = (x_1, x_2, \cdots, x_n)$, удовлетворяющий (4.1) называется *планом* задачи математического программирования. План X задачи (4.1) и (4.2) с неотрицательными компонентами называется *допустимым.* План $X^* = (x_1^*, x_2^*, \cdots, x_n^*)$, доставляющий экстремум функции (4.2) называется *оптимальным.*

Если все функции в (4.1) и (4.2) линейны, то задача называется *линейной,* а раздел математического программирования будет называться *линейным программированием.* Его методы широко применяются на промышленных предприятиях:

- при оптимизации производственной программы; распределении её по цехам; по участкам; по временным интервалам;
- при ассортиментной загрузке оборудования;
- в задачах текущего и перспективного технологического планирования;
- при планировании грузопотоков;
- при определении плана товарооборота и его распределении;
- при составлении оптимальных смесей;
- раскройно-транспортных и производственно-транспортных задач.

Если хотя бы одна из функций φ_i или $z(x)$ нелинейны – это раздел *нелинейного программирования.* В промышленности оно используется:

- при расчете оптимальной партии выпуска деталей;
- в управлении комплектными поставками и запасами;
- в распределении ограниченных ресурсов;
- в оптимизации некоторых показателей производственно-экономической деятельности.

Если хотя бы одна из функций φ_i или система ограничений изменяется во времени или целевая функция *z(x)* имеет оптимальную структуру:

$$z(x) = \sum_j z_j(x_j), \textit{или} \quad z(x) = \prod_j z_j(x_j),$$

или сам процесс принятия решений имеет многошаговый характер, то такие задачи решаются методом *динамического программирования*.

В управлении промышленным предприятием таким методом решаются:

- задачи текущего и перспективного планирования;
- управление производством;
- распределением ограниченных ресурсов;
- оптимального размещения;
- замены оборудования.

Если на все или некоторые переменные x_j наложено условие дискретности, например, целочисленности, то такой раздел программирования считается *дискретным (целочисленным*). К такому разделу относятся задачи:

- выпуска делимой продукции;
- задачи маршрутизации;
- задачи управления поставками при заданных транзитных нормах отпуска;
- размещение производственно-складской структуры.

Формы записи задачи линейного программирования

Общей задачей линейного программирования (ОЗЛП), называется задача, в которой требуется максимизировать (минимизировать) функцию:

$$f = \sum_{j=1}^{n} c_j x_j \qquad (4.3)$$

при условиях

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \left(i = 1, ..., m_1\right), \qquad (4.4)$$

$$\sum_{j=1}^{n} a_{ij} x_j \ge b_i, \left(i = m_1 + 1, ..., m_2\right), \qquad (4.5)$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \left(i = m_2 + 1, ..., m\right), \qquad (4.6)$$

$$x_j \ge 0, \quad \left(j = 1, ..., n_1\right), \qquad (4.7)$$

$$x_j - \textit{произвольное}, \quad \left(j = n_1 + 1, ..., n\right) \qquad (4.8)$$

Функция f называется *целевой* функцией. Условия (4.4)-(4.8) называются *ограничениями*. Величины m и n между собой несвязаны. На переменные могут накладываться условия неотрицательности, целочисленности.

Набор чисел $x = (x_1, x_2, ..., x_n)$, удовлетворяющий ограничениям (4.4)-(4.8), называется *планом* ОЗЛП. План **X** задачи (4.3)-(4.8) с неотрицательными компонентами называется *допустимым*. План $x^* = (x_1^*, x_2^*, ..., x_n^*)$, доставляющий экстремум функции (4.3), называется *оптимальным*.

Если в задаче линейного программирования требуется максимизировать целевую функцию (4.3) при ограничениях (4.4), (4.7) и $m_1=m$, $n_1=n$, то задача задана в *симметричной* форме записи.

Задача линейного программирования, в которой требуется найти максимум функции (4.3) при ограничениях (4.6), когда $m_2=0$, и (4.6), задана в *канонической форме* записи.

При использовании тех или иных методов решения задач линейного программирования приходится осуществлять переходы от одной формы записи задачи к другой. Делается это путем следующих математических преобразований:

- неравенства (4.4) путем умножения левых и правых частей на (-1) превращаются в неравенства (4.4) и наоборот;
- если какая-то переменная не подчинена условию неотрицательности (4.7), то ее заменяют разностью двух других переменных. Например, произвольную переменную x_k заменяют на $(x_k' - x_k'')$, где $x_k' \ge 0$ и $x_k'' \ge 0$;
- неравенства (4.4) превращаются в равенства (4.6) путем добавления *дополнительной или балансовой* переменной x_{n+1};
- неравенства (4.5) превращаются в равенства (4.6) путем вычитания *дополнительной или балансовой* переменной x_{n+1};
- задачу минимизации можно формально заменить задачей максимизации. Минимальное значение функции f равно максимальному значению функции ($-f$), взятому с противоположным знаком, т.е. $\min f = -\max(-f)$.

Некоторые модели задач линейного программирования

Задача о выборе оптимальных технологий

Пусть для выпуска некоторой однородной продукции можно использовать n, $j=\overline{1,n}$ технологий. Для этого требуется:

m видов исходных ресурсов, заданных $b_1, b_2, \cdots, b_m$;

c_j – стоимость конечной продукции (в рублях), производимой в единицу времени по j-му технологическому способу;

$a_{i,j}$ – расход i-го исходного ресурса в единицу времени по j-му технологическому способу.

В качестве неизвестной величины x_j примем время, в течение которого предприятие вырабатывает продукцию по *j*-му технологическому способу.

Пренебрегая временем переналадок, необходимых для перехода от одного технологического способа к другому, получим следующую математическую модель задачи.

Максимизировать объём выпуска продукции

$$\max : z(x) = \sum_{j=1}^{n} c_j x_j, \tag{4.9}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \left(i = 1, \ldots, m\right), \tag{4.10}$$

$$x_j \geq 0,\ j = \overline{1, n}\ . \tag{4.11}$$

Задача оптимального использования ресурсов

Имеется *m* видов ресурсов, заданных вектором $A_0 = (b_1, b2, \cdots, b_m)$. Задана матрица $A = \left\| a_{i,j} \right\|$, где $a_{i.j}$ – норма расхода *i*-го ресурса на единицу *j*-й продукции $(j = \overline{1,n})$. Эффективность выпуска единицы *j*-й продукции характеризуются, например, прибылью p_j. Требуется определить план выпуска продукции $X = (x_1, x_2, ..., x_n)$, максимизирующий прибыль предприятия при заданных ресурсах, т.е.

$$\max : z = \sum_{j=1}^{n} p_j x_j \tag{4.12}$$

при ограничениях

$$\sum_{j=1}^{n} a_{i,j} x_j \leq b_i, \quad i = \overline{1, m} \tag{4.13}$$

$$x_j \geq 0, \quad j = \overline{1,n} \tag{4.14}$$

Часто ассортимент продукции устанавливается вышестоящей организацией, т.е. его объемы должны быть заключены в некоторых границах d_j и D_j:

$$d_j \leq x_j \leq D_j, \quad j = \overline{1,n} \tag{4.15}$$

и (или) удовлетворять условию комплектности для сборки $x_1 : x_2 : ... : x_n = l_1 : l_2 : ... : l_n$.

Эти условия можно записать так:

$$l_{j+1}x_j - l_j x_{j+1} = 0, \quad j = \overline{1,n} \qquad (4.16)$$

Условию комплектности могут удовлетворять только некоторые из изготовляемых продуктов.

Модель задачи оптимального использования ресурсов лежит в основе моделей оптимизации годовой производственной программы предприятия. В простейшем случае задача решается на уровне предприятия или цеха. В модель включаются только ограничения по фонду времени работы оборудования. Сохраняя прежние обозначения, запишем через α_j и c_j соответственно отпускную цену и затраты на единицу *j*-й продукции. В качестве критерия оптимальности могут быть приняты:

- максимум прибыли

$$\max: z1 = \sum_{j=1}^{n}(\alpha_j - c_j)x_j \qquad (4.17)$$

- минимум затрат на производство

$$\min: z2 = \sum_{j=1}^{n} c_j x_j \qquad (4.18)$$

- максимум выпускав стоимостном выражении:

$$\max: z3 = \sum_{j=1}^{n} \alpha_j x_j \qquad (4.19)$$

- максимум использования оборудования

$$\max: z4 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} x_j \qquad (4.20)$$

и другие при ограничениях на фонд времени оборудования (4.13), ассортимент выпуска продукции (4.14) и при условии неотрицательности (4.15).

В более развернутом виде наряду с системой ограничений по фонду времени работы оборудования учитываются ограничения:

- по лимитирующим материалам

$$\sum_{j=1}^{n} r_{l_j} x_j \le R_l, \quad l = \overline{1,L} \qquad (4.21)$$

где r_{l_j} – -норма расхода *l*-го вида материала на выпуск единицы *j*- ой продукции; R_l – фонд *l*-го вида материала;

- по затратам труда

$$\sum_{j=1}^{n} t_{k_j} x_j \le T_k, \quad k = \overline{1,K} \qquad (4.22)$$

где t_{k_j} – норма времени, затрачиваемая рабочим k-ой специальности на выпуск единицы j-ой продукции; T_k – фонд времени рабочего k-ой специальности;

- по фонду заработной платы

$$\sum_{k=1}^{K}\sum_{j=1}^{n} f_k t_{k_j} x_j \le F \qquad (4.23)$$

где f_k – стоимость часа труда рабочего k-ой специальности; F – фонд заработной платы рабочих;

- по складским площадям

$$\sum_{j=1}^{n} s_j x_j \le S \qquad (4.24)$$

где s_j – норма расхода складской площади за единицу j-ой продукции; S – общий фонд таких площадей;

- по выпуску товарной продукции

$$\sum_{j=1}^{n} w_j x_j \le W \qquad (4.25)$$

где W – годовое плановое задание по производству товарной продукции в стоимостном выражении.

Дальнейшее усложнение задач оптимизации производственной программы (4.17-4.25) включают требования расшивки узких мест производства в рамках фонда капитальных вложений, предназначенных на пополнение оборудования. Еще более сложный тип моделей учитывает структуру продукции, технологические способы и т.д.

Задача о распределении производственной программы (о размещении заказов или загрузке взаимозаменяемых групп оборудования)

Речь идет о задаче распределении заказов между $m\,(i=\overline{1,m})$ предприятиями (цехами, станками, исполнителями) с различными производственными и технологическими характеристиками по взаимозаменяемым в смысле выполнения заказов. Требуется составить такой план размещения заказов (загрузки оборудования), при котором задание было бы выполнено, а показатель эффективности достигал экстремального значения.

Сформулируем задачу математически. Пусть на $m\,(i=\overline{1,m})$ однородных группах оборудования можно изготовить $n\,(i=\overline{1,n})$ видов продукции. План выпуска каждого вида на определенный период задан набором $x_j, j=\overline{1,n}$. Мощность каждого вида оборудования ограничена и равна b_i. Известна технологическая матрица $A=\left\|a_{i,j}\right\|$, где $a_{i,j}$ – число единиц продукции *j*, выпускаемой в единицу времени на *i*-м оборудовании.

Пусть $c_{i,j}$ – затраты, связанные с выпуском *j*-ой продукции на *i*-м оборудовании; $x_{i.j}$ – неизвестная величина: объем выпуска *j*-й продукции на *i*-м оборудовании.

Модель задачи примет вид:
целевая функция – минимизация расходов на реализацию всех заказов

$$\min: z=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{i,j}x_{i,j} \qquad (4.26)$$

при ограничениях

- на ресурсы оборудования

$$\sum_{j=1}^{n}a_{i,j}x_{i,j}\le b_i \quad i=\overline{1,m} \qquad (4.27)$$

- на выпуск продукции

$$\sum_{i=1}^{m}x_{i,j}=x_j \quad j=\overline{1,n} \qquad (4.28)$$

- условие неотрицательности

$$x_{i,j}\ge 0 \qquad (4.29)$$

Задачу (4.26)-(4.29) иногда называют *распределительной*.

Если по некоторым видам продукции допускается превышение плана, то ограничение (4.28) примет вид

$$\sum_{i=1}^{m}x_{i,j}\ge x_j \quad j=\overline{1,n_1}.$$

По продукции, заказы на которую принимаются по кооперации, для полной загрузки оборудования, ограничения можно записать так:

$$\sum_{i=1}^{m}x_{i,j}\le x_j \quad j=\overline{n_1+1,n_2}.$$

Для продукции, выпуск которой должен соответствовать плану, ограничения таковы:

$$\sum_{i=1}^{m}x_{i,j}=x_j \quad j=\overline{n_2+1,n}.$$

В качестве целевой функции можно принять:

- максимум прибыли

$$\max: z = \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha_j - c_{i,j})x_{i,j},$$

где α_j – отпускная цена j-ой продукции;

- минимум затрат станочного времени

$\min: z = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}x_{i,j}$, и другие.

Математическая модель задачи о распределении производственной программы в случае выпуска комплектной продукции имеет вид (x – число комплектов, α_j – число единиц j-ой продукции, входящей в комплект)

$$\max: z = x \tag{4.30}$$

при ограничениях:

- на ресурсы оборудования

$$\sum_{j=1}^{n} a_{i,j}x_{i,j} \le b_i \quad i = \overline{1,m} \tag{4.31}$$

- на комплектность продукции

$$\sum_{i=1}^{m} x_{i,j} \ge \alpha_j x \qquad j = \overline{1,n} \tag{4.32}$$

- условие неотрицательности

$$x_{i,j} \ge 0 \tag{4.33}$$

Задача загрузки невзаимозаменяемых групп оборудования

Пусть имеется $m\ (i = \overline{1,m})$ групп оборудования, на котором нужно выпускать $n\ (j = \overline{1,n})$ видов продукции. Предполагается, что группы оборудования различны. В пределах данной технологии одна и та же продукция на другом оборудовании не производится, т.е. оборудование взаимнозаменяется лишь при замене технологического способа производства.

Пусть $k\ (k = \overline{1,K})$ – номер технологического способа. Обозначим через $a_{i,j}^{k}$ норму затрат времени на обработку единицы j-ой продукции на i-ом оборудовании по k-му технологическому способу. Пусть b_i – общий полезный фонд времени работы i -ой группы оборудования.
Неизвестная величина задачи $x_{j,k}$ – объём продукции j-го вида, вырабатываемой по k-ой технологии; x_j – заданный сверху план j- ой

продукции; $c_{j,k}$ – затраты связанные с изготовлением единицы j-ой продукции по k-ой технологии, которые следует минимизировать.

Математическая модель задачи:

$$\min : z = \sum_{j=1}^{n} \sum_{k=1}^{K} c_{j,k} x_{j,k} \qquad (4.34)$$

при ограничениях :

- на фонд времени работы каждой группы оборудования

$$\sum_{j=1}^{n} \sum_{k=1}^{K} a_{i,j}^{k} x_{j,k} \le b_i, \quad i = \overline{1,m} \qquad (4.35)$$

- \- на выпуск продукции

$$\sum_{k=1}^{k} x_{jk} \ge x_j, \; j = \overline{1,n} \qquad (4.36)$$

- \- условие неотрицательности

$$x_{j,k} \ge 0 \qquad (4.37)$$

Задачи распределения производственной программы по календарным периодам

Эти задачи чаще всего целочисленны с булевыми переменными. Рассмотрим простейшую постановку такой задачи.

Известна годовая *производственная* программа $v_{j,t}, \quad j = \overline{1,n}, t = \overline{1,T}$, т.е. потребность в t-м интервале на j-ю продукцию. Пусть $b_{i,t}$ – фонд i-ых ресурсов в t-м интервале; $a_{i,j}$ – норма расхода i-ого ресурса на выпуск j-ой продукции. Неизвестные величины: $x_{j,t}$ –объемы выпуска j-ой продукции в t-м интервале.

Для математической формулировки задачи, например, экспертным путем, определяют приоритеты каждого вида продукции p_j. Наиболее важным видам приписывается приоритет 1 и т.д.

Целевая функция задачи – максимум выпуска с учетом приоритетности продукции

$$\max : z = \sum_{j=1}^{n} \sum_{t=1}^{T} \frac{1}{p_j} x_{j,t} \qquad (4.38)$$

при ограничениях:

- на ресурсы в каждом t-м интервале

$$\sum_{j=1}^{n} a_{i,j} x_{j,t} \le b_{i,t}, \quad i = \overline{1,m}, t = \overline{1,T} \qquad (4.39)$$

- на объем выпуска

$$\sum_{\tau=1}^{t} x_{j,\tau} \ge \sum_{\tau=1}^{t} v_{j,\tau},\ j=\overline{1,n},\ t=\overline{1,T} \qquad (4.40)$$

- на условие неотрицательности

$$x_{j,t} \ge 0,\ j=\overline{1,n},\ t=\overline{1,T} \qquad (4.41)$$

Задача производственного планирования

Рассматривается производство, в котором участвует m ингредиентов (сырье, промежуточные и конечные продукты). Имеется n технологических способов организации производства. Обозначим через неизвестную величину x_j – интенсивность использования технологического способа; $a_{i,j}$ – затраты (если $a_{i,j} \ge 0$) или выпуск (если $a_{i,j} < 0$) *i*-го ингредиента по *j*-му технологическому способу с единичной интенсивностью; b_i – величина *i*-го ингредиента. Если $b_i < 0$, то *i*-ый ингредиент выпускается, если $b_i \ge 0$, то *i*-ый ингредиент потребляется.

Пусть c_j — прибыль (если $c_j < 0$) или затраты (если $c_j \ge 0$) от использования *j*-ой технологии с единичной интенсивностью.

Общая модель производственного планирования примет вид:

- целевая функция – максимизировать прибыль от использования всех способов производства

$$\max : z = \sum_{j=1}^{n} c_j x_j \qquad (4.42)$$

- при ограничениях, означающих, что затраты *i*-го ингредиента не должны превышать b_i, если $b_i \ge 0$ или *i*-ый ингредиент должен быть выпущен в количестве не меньше $|b_i|$, если $b_i < 0$, т.е.

$$\sum_{j=1}^{n} a_{i,j} x_j \le b_i \quad i=\overline{1,m} \qquad (4.43)$$

- и условия неотрицательности

$$x_j \ge 0 \qquad (4.44)$$

Задача о смесях

Критерием оптимальности можно принять минимум растрат на исходные материалы.

Построим математическую модель задачи.

Пусть $a_{i,j}$ – доля i-го ($i=\overline{1,m}$) химического элемента в исходном шихтовом материале j-го ($j=\overline{1,n}$) вида. Обозначим через x_j долю j-го материала, включаемого в состав шихты. Тогда содержание i-го элемента в шихте выбранного состава выразится суммой

$$\sum_{j=1}^{n} a_{i,j} x_j, \quad i=\overline{1,m}$$

Однако в результате окислительного процесса содержание некоторых химических элементов в расплаве существенно уменьшается (происходит их угар). В связи с этим в состав расплава следует вводить поправочный коэффициент. Если обозначить величину угара через u_i, то поправочный коэффициент будет равен $(1-u_i)$. Вот почему фактическое содержание химических элементов в расплаве следует находить по формуле

$$\sum_{j=1}^{n}(1-u_i)a_{i,j}x_j,\ i=\overline{1,m}$$

Реальная величина i-го химического элемента в расплаве должна колебаться между нижними b_i и верхними B_i пределами. Обозначим через c_j стоимость единицы j-го исходного материала.

Модель задачи об оптимальном составе шихты принимает вид:

$$\min: z=\sum_{j=1}^{n} c_j x_j \tag{4.45}$$

при ограничениях:

- на нижний и верхний уровни содержания в расплаве химического элемента

$$b_i \le \sum_{j=1}^{n}(1-u_i)a_{i,j}x_j \le B_i, \quad i=\overline{1,m} \tag{4.46}$$

- на комплектность долевого участия исходных материалов

$$\sum_{j=1}^{n} x_j = 1 \tag{4.47}$$

- граничные условия, определяемые технологией выплавки и ограниченностью отдельных исходных материалов, таковы

$$0 \le x_j \le D_j,\ j=\overline{1,n} \tag{4.48}$$

Задача о раскрое материалов

Рациональный раскрой промышленных материалов - важный источник экономии ресурсов. Он повышает коэффициент использования материалов. Централизация раскроя на снабженческосбытовых базах позволяет сократить ассортимент материалов, что ведет к укреплению заказов. Сущность задачи об оптимальном раскрое состоит в разработке таких технологически допустимых планов раскроя, при котором получается необходимый комплект заготовок, а отходы по площади, весу или стоимости к минимуму.

В настоящее время наиболее изучены вопросы раскроя длинномерных и листовых материалов.

Задача комплектности раскроя формулируется так: на раскрой поступает $m, (i = \overline{1,m})$ материалов в объеме b_i каждого. Требуется изготовить K различных комплектующих изделий в количестве пропорциональных: $l_1 : l_2 : ... : l_K$. Каждая единица i-го материала может быть раскроена n $(j = \overline{1,n})$ различными способами. При раскрое единицы i-го материала j-м способом получается $a_{i,j}^k$ единиц k-го изделия. Обозначим через $x_{i,j}$ количество единиц i-го материала, раскраиваемого j-м способом, а количество изготовляемых комплектов – через x.

Математическая модель задачи имеет вид:

$$\max : z = x \tag{4.49}$$

при ограничениях

- на объем каждого вида материалов

$$\sum_{j=1}^{n} x_{i,j} \le b_i, \quad i = \overline{1,m} \tag{4.50}$$

- условия комплектности

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^k x_{i,j} = l_k x, \quad k = \overline{1,K} \tag{4.51}$$

- условия неотрицательности

$$x_{i,j} \ge 0 \tag{4.52}$$

В частности, если на раскрой поступает один вид материала в объеме b, то модель примет вид:

$$\max : z = x$$

$$\sum_{j=1}^{n} x_j = b$$

$$\sum_{j=1}^{n} a_{j,k} x_j = l_k x, \quad k = \overline{1,K}$$

$$x_j \ge 0, \quad j = \overline{1,n}$$

Приведем простейшую модель раскройной задачи в случае критерия – минимум отходов при раскрое. Рассмотрим задачу оптимального раскроя по одному измерению длинномерных материалов (прутков, труб, профильного проката и др.).

Пусть

- L – длина исходного материала;
- Δ_i – длина заготовки *i- го* вида;
- b_i – нужное количество заготовок *i- го* вида;
- $a_{i,j}$ – количество *i- го* вида, полученного при раскрое единицы исходного материала j-му варианту;
- c_j – отход при раскрое единицы исходного материала j-му варианту.

Неизвестная величина x_j – интенсивность использования j-го варианта, т.е. сколько единиц исходного материала будет раскроена по j-му варианту.

Математическая модель

$$\min: z = \sum_{j=1}^{n} c_j x_j$$

при условии получения заготовок в соответствии с ассортиментной программой

$$\sum_{j=1}^{n} a_{i,j} x_j = b_i, \quad i = \overline{1,m},$$

при условии неотрицательности

$$x_j \geq 0, \ j = \overline{1,n}$$

Рассмотрим критерий минимум числа прутков для удовлетворения потребности. Обозначим число прутков через x. Модель примет вид:

- целевая функция - минимум числа единиц раскраиваемого материала

$$\min: z = x$$

при ограничениях:

- заявки потребителей удовлетворяются полностью

$$\sum_{j=1}^{n} a_{i,j} x_j = b_i$$

- на объем исходного материала, поступившего на раскрой

$$\sum_{i=1}^{m} \Delta_i \sum_{j=1}^{n} a_{i,j} x_j + m \sum_{i=1}^{n} c_j x_j = xL$$

- условие неотрицательности $x_j \geq 0, \quad j = \overline{1,n}$

ТЕМА 5. МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ (ЗЛП)

Лекция 5. Графический метод решения ЗЛП. Решение задач линейного программирования симплексным методом. Симплексное преобразование. Указания к нахождению начального опорного плана. Нахождение оптимального плана.

Графический метод решения ЗЛП

Графический метод используется для решения задач линейного программирования с двумя переменными, заданными в симметричной форме, и многими переменными, заданными в канонической форме (при условии, что они содержат не более двух свободных переменных).

Задачу линейного программирования с двумя переменными можно записать так:

$$\max: f = c_1x_1 + c_2x_2 \tag{5.1}$$

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 \le b_1 \\ &\dots\dots\dots\dots\dots\dots \\ &a_{m1}x_1 + a_{m2}x_2 \le b_m \end{aligned} \tag{5.2}$$

$$x_1 \ge 0, \quad x_2 \ge 0 \tag{5.3}$$

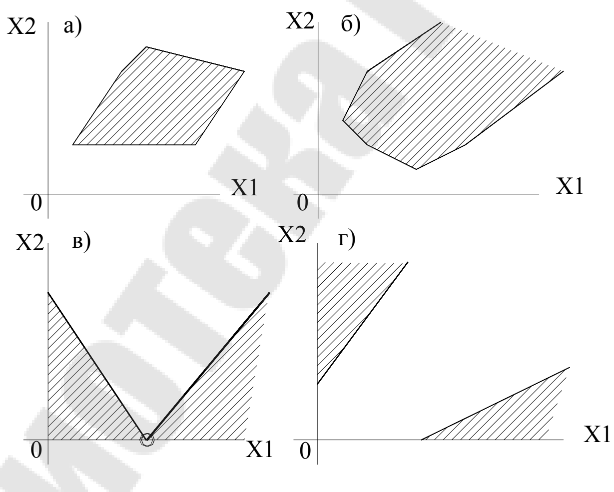


Рисунок 5.1 – Область допустимых решений

Область допустимых решений задачи (5.1)-(5.3) представляет собой либо выпуклый многоугольник (рис. 5.1, а)), либо выпуклую многоугольную область (рис. 5.1, б)), либо единственную точку (рис. 5.1, в)), либо пустое множество (рис. 5.1, г)).

Функция (5.1) представляет собой на плоскости X_1OX_2 семейство параллельных прямых, каждой из которых отвечает определенное значение f. Перпендикулярный к этим прямым вектор $\boldsymbol{p}=(c_1, c_2)$ указывает направление наискорейшего возрастания функции f (5.1), и задача (5.1)-(5.3) заключается в следующем: *необходимо найти точку допустимой области, через которую проходит прямая семейства f, отвечающая наибольшему значению функции (5.1).*

При решении задачи линейного программирования графическим способом могут встретиться следующие случаи:

- задача имеет единственное решение (рис. 5.2, а));
- задача имеет бесконечное множество решений (рис. 5.2, б));
- функция f не имеет экстремального значения (рис. 5.2, в)).

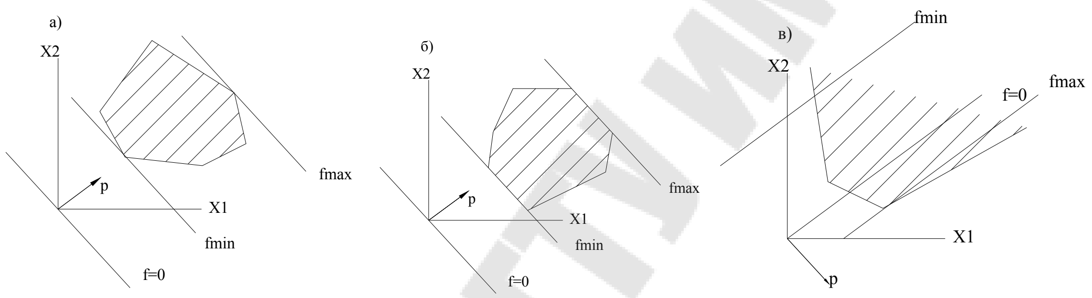


Рисунок 5.2 – Графическое решение ОЗЛП

Решение задачи графическим способом проводиться в такой последовательности:

- записывают уравнения граничных прямых $a_{k1}x_1 + a_{k2}x_2 = b_k$, $k=1, \ldots, m$, и строят их на плоскости X_1OX_2;
- определяют полуплоскости, соответствующие исходным ограничениям-неравенствам (5.2). Для этого берут произвольную точку, лежащую по ту или иную сторону от граничной прямой, и ее координаты подставляют в левую часть ограничения-неравенства. Если оно удовлетворяется, то искомой будет полуплоскость, которая содержит выбранную точку; если оно не удовлетворяется, то искомой будет полуплоскость, которой данная точка не принадлежит;
- выделяют область допустимых решений как общую часть $m+2$ полуплоскостей, где m полуплоскостей соответствуют исходным неравенствам (5.2), а 2 полуплоскости – условию

неотрицательности переменных ($x_1 \geq 0$ – правая координатная полуплоскость; $x_2 \geq 0$ – верхняя координатная полуплоскость);

- строят вектор ***p***= (c_1; c_2) и перпендикулярно к нему одну из прямых семейства *f,* например, *f=0*;
- определяют экстремальную точку многоугольника решений путем параллельного перемещения вспомогательной прямой *f=0* в направлении вектора ***p***. Это будет наиболее удаленная крайняя точка, в которой прямая *f* встречается с областью допустимых решений. Если необходимо найти точку, которой соответствует минимальное значение функции *f,* то вспомогательную прямую перемещают в направлении вектора ***p*** до пересечения с первой точкой допустимой области (либо прямую *f=0* перемещают в направлении вектора (-***p***));
- вычисляют координаты оптимальной точки и значения функции *f*.

  Если в канонической записи задачи со многими переменными разность между числом переменных *n* и числом линейно независимых уравнений *r* равна двум, т.е. *n-r=2*, то задачу можно решать графическим методом. Модель этой задачи преобразуют к симметричной форме. Полученную задачу с двумя переменными решают графически. Найдя координаты оптимума, подставляют их в ограничительные уравнения канонической формы. Решая полученную систему уравнений, определяют и остальные координаты оптимума.

Решение задач линейного программирования симплексным методом

Симплексный метод (симплекс-метод) является одним из универсальных методов решения ЗЛП, называемый также методом последовательного улучшения плана. Симплекс- метод позволяет вести расчеты вручную и на ЭВМ.

Дана задача линейного программирования в канонической форме:

максимизировать $f = \sum_{j=1}^{n} c_j x_j$ (5.4)

при условии

$$\sum_{j=1}^{n} a_{ij} x_j = a_{i0} \quad (i = \overline{1,m}); \qquad (5.5)$$

$x_j \geq 0 \qquad (j = \overline{1, n})$; (5.6)

Предположим, что в системе (5.5) $m<n$ и все m уравнений линейно независимы (ранг системы $r=m$). В этом случае система имеет бесчисленное множество решений. Ее можно разрешить относительно m переменных x_1, ..., x_m, если векторы-коэффициенты $\mathbf{a_1}$, …, $\mathbf{a_m}$ при этих переменных линейно независимы:

$$x_i = b_{i0} - \sum_{j=1}^{n-m} b_{ij} x_{m+j} \quad (i = \overline{1, m}). \tag{5.7}$$

В этом случае x_1, ..., x_m - *базисные переменные*, а x_{m+1}, ..., x_n – *свободные переменные*. Тогда и целевую функцию можно выразить через свободные переменные x_{m+j}:

$$f = b_{00} - \sum_{j=1}^{n-m} b_{0j} x_{m+j} . \tag{5.8}$$

Если все $b_{i0}>0$, то план $\mathbf{x}_0=(b_{10};\ b_{20};\ldots;b_{m0};\ 0;\ldots0)$ будет являться опорным и $f(\mathbf{x}_0)=b_{00}$. Если при этом опорном плане значение целевой функции максимально, то опорный план является *оптимальным.*

Решение задачи (5.4)-(5.6) симплексным методом складывается из двух этапов:

1) нахождение начального опорного плана;
2) определение среди опорных планов задачи оптимального.

Решение задачи производится с использованием *симплексных таблиц*. Симплексная таблица, содержащая ограничения (5.7) и целевую функцию (5.8) имеет вид табл. 5.1.

Таблица 5.1 – Вид симплекной таблицы

БП	СЧ	Переменные (БП, СП)					
		$-x_1$	…	$-x_m$	$-x_{m+1}$	...	$-x_n$
$x_1=$	b_{10}	1	0	0	b_{11}	…	$b_{1,n-m}$
…	…	0	1	0	….	...	….
$x_m=$	b_{m0}	0	0	1	b_{m1}	…	$b_{m,n-m}$
f=		0	0	0	b_{01}	…	$b_{0,n-m}$

Обозначения в таблице:

- БП – базисные переменные;
- СП – свободные переменные;
- СЧ – свободные члены;
- последняя строка называется f – строкой или строкой целевой функции.

Признаком оптимальности опорного плана является неотрицательность *f*-строки.

Симплексное преобразование

Преобразование системы (5.5) к новому базису называется *симплексным преобразованием*.

Правило выбора переменных при направленном преобразовании (при переходе от одного опорного плана к другому, более близкому к оптимальному) одного базиса в другой: *в базис вводят переменную* x_{m+j}, *соответствующую отрицательному элементу f – строки с наибольшей абсолютной величиной.*

Столбец коэффициентов при переменной, включенной в базис, называется *разрешающим.*

Для определения переменной, подлежащей исключению из базиса, составляют отношения свободных членов к положительным элементам разрешающего столбца (такие отношения называются *симплексными*) и находят среди них наименьшее, которое и определяет переменную, исключаемую из базиса. Строка, которой соответствует переменная, исключаемая из базиса, называется *разрешающей*.

Элемент симплекс-таблицы, стоящий на пересечении разрешающей строки и разрешающего столбца, называется *разрешающим.*

Предполагая, что разрешающий элемент выбран, можно сформулировать следующее правило перерасчета элементов симплекс-таблицы при переходе к новому опорному плану:

- разрешающий элемент заменяется обратной величиной;
- остальные элементы разрешающей строки делятся на разрешающий элемент;
- остальные элементы разрешающего столбца делятся на разрешающий элемент и меняют свои знаки;
- прочие элементы вычисляют по формуле :

$$b_{ij}^{'} = (b_{ij}b_{ks} - b_{is}b_{kj}) / b_{ks} \quad (i = 0, ..., m, i \neq k, \ j = 0, ..., n - m, \ j \neq s) \quad (5.9)$$

где b_{ks} –разрешающий элемент.

При вычислении элементов по формуле (5.9) удобно пользоваться правилом прямоугольника. Элементы, входящие в эту формулу, расположены в вершинах воображаемого «прямоугольника».

$$
\begin{array}{cccccccccccccc}
. & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & b_{ij} & . & . & . & . & b_{is} & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & b_{kj} & . & . & . & . & \boxed{b_{ks}} & . & . & . & . & .
\end{array}
$$

Диагональ этого прямоугольника, на которой расположены разрешающий b_{ks} и преобразуемый b_{ij} элементы, называется главной, а другая - побочной. Преобразованный элемент b_{ij}' равен разности произведений элементов главной и побочной диагоналей, деленной на разрешающий элемент.

Сформированного правила придерживаются независимо от того, в какой вершине прямоугольника расположен разрешающий элемент.

Указания к нахождению начального опорного плана

Симплексное преобразование можно применить для нахождения опорного плана, т.е. для преобразования системы (5.5) к виду (5.7).

Система (5.5) записывается в виде:

$$
\begin{aligned}
&0 = a_{10} + a_{11} \cdot (-x_1) + ... + a_{1n} \cdot (-x_n), \\
&.., \\
&0 = a_{m0} + a_{m1} \cdot (-x_1) + ... + a_{mn} \cdot (-x_n)
\end{aligned}
\tag{5.10}
$$

а целевую функцию f –в виде:

$$
f = 0 + (-c_1) \cdot (-x_1) ... + (-c_n) \cdot (-x_n) \tag{5.11}
$$

Равенства (5.10) и (5.11) вносят в симплекс – таблицу, некоторые уравнения системы (5.10) умножив на –1, чтобы элементы столбца свободных членов были положительными. В результате получается таблица 5.2.

Таблица 5.2

	СЧ	$-x_1$	…	$-x_n$
0=	a_{10}	a_{11}	…	a_{1n}
…	…	...	…	…
0=	a_{m0}	a_{m1}	…	a_{mn}
f=	0	$-c_1$	…	$-c_n$

Таблица 5.2 подвергается симплексным преобразованиям. На каждом шаге один из x и 0 меняются местами. Через r шагов (r- ранг матрицы коэффициентов системы (5.10)) r переменных x поменяются с 0 первого столбца. Пусть это будут переменные x_1, x_2, ..., x_r. Получится таблица вида таб. 5.3.

Таблица 5.3

	СЧ	0	…	0	$-x_{r+1}$	…	$-x_n$
x_1	b_{10}	b_{11}	…	b_{1r}	$b_{1,r+1}$	…	b_{1n}
…	…	…	…	…	…	…	…
x_r	b_{r0}	b_{r1}	…	b_{rr}	$b_{r,r+1}$	…	b_{rn}
0=	$b_{r+1,0}$	$b_{r+1,1}$	…	$b_{r+1,r}$	0	…	0
…	…	…	…	…	…	…	…
0=	b_{m0}	b_{m1}	…	b_{mr}	0	…	0
f=	b_{00}	b_{01}	…	b_{0r}	$b_{0,r+1}$	…	b_{0n}

Система (5.5) будет совместной, если в таблице 5.3 свободные члены $b_{r+1,0},\dots,b_{m0}$ равны нулю. Если хотя бы один из них отличен от нуля, то система будет несовместна.

При расчетах опускают разрешающие столбцы и строки, целиком состоящие из нулей. Если в ходе преобразований встретится 0-строка, в которой все элементы неположительны (свободный член положителен), то соответствующая система не имеет опорных решений. Если система (5.5) совместна в области неотрицательных решений, то через r шагов перейдем к таблице аналогичной табл. 5.1 и содержащей начальный опорный план $x^0=(b_{10},\ ,b_{r0},0,\dots,0)$, а также $f(x^0)=b_{00}$. Найденный опорный план соответствует базису $Б_0=\{x_1,\dots,x_r\}$, относительно которого оказалась разрешима система.

<u>Пример 5.1</u> Найти начальный опорный план задачи:

$$f = 2x_1 + 3x_2 - x_3 + 4x_4 \quad (\max)$$

$$\left.\begin{aligned} 3x_1 - 3x_2 - x_3 + x_4 &= 8 \\ x_1 - x_2 + 2x_3 + x_4 &= 4 \end{aligned}\right\}$$

$$x_j \ge 0, \quad j = \overline{1,4}$$

<u>Решение</u>. Систему уравнений запишем в виде:

$$0=8-3x_1+3x_2+x_3-x_4$$

$$0=4-x_1+x_2-2x_3-x_4$$

и занесем в симплексную таблицу (табл. 5.4)

Таблица 5.4

	СЧ	Переменные			
		$-x_1$	$-x_2$	$-x_3$	$-x_4$
0=	8	3	-3	-1	1
0=	4	1	-1	2	$\boxed{1}$ ⇒
f=	0	-2	-3	1	-4

⇑

В табл. 5.4 за разрешающий столбец можно взять любой столбец за исключением второго, так как он не содержит ни одного положительного элемента. Возьмем, например, четвертый столбец и вычислим симплексные отношения: min (8/1; 4/1)=4 (вторая строка разрешающая). Выполним симплексное преобразование и получим табл. 5.5. Разрешающим столбцом в табл. 5.5 можно взять только первый, и разрешающей строкой будет только первая, т.к. min (4/2;4/1)=2. После симплексного преобразования с разрешающим элементом 2 получим табл. 5.6 с начальным опорным планом x^o=(2; 0; 0; 2), при котором $f(x^o)$=12.

Таблица 5.5

	СЧ	СП		
		$-x_1$	$-x_2$	$-x_3$
0=	4	2	-2	-3
x_4=	4	1	-1	2
f=	16	2	-7	9

Таблица 5.6

БП	СЧ	СП	
		$-x_2$	$-x_3$
x_1=	2	-1	-3/2
x_4=	2	0	7
f=	12	-5	12

Нахождение оптимального плана

Если в симплексной таблице, содержащей опорный план, все элементы f-строки (не считая свободного члена) неотрицательны, то данный опорный план является оптимальным. Полученный оптимальный план будет единственным, если все элементы *f*-строки положительны. Если среди неотрицательных элементов встречается хотя бы один нулевой, то задача имеет бесконечное множество оптимальных планов.

Если хотя один элемент *f* –строки отрицательный, то оптимальный опорный план находят по алгоритму:

- выбирают разрешающий столбец по отрицательному элементу *f*-строки (если в *f*-строке отрицательных элементов несколько, то наибольший по абсолютной величине отрицательный элемент укажет на разрешающий элемент);
- разрешающая строка находится по минимальному симплексному отношению;
- делают симплексное преобразование с выбранным разрешающим элементом и получают новый опорный план, который опять проверяют на оптимальность. Решение проводится до тех пор, пока

не будет получен оптимальный план, либо установлена неразрешимость задачи.

Если в f-строке симплексной таблицы, содержащей опорный план, есть хотя бы один отрицательный элемент, а в соответствующем этому элементу столбце нет ни одного положительного, то целевая функция не ограничена в области допустимых решений, т.е. $f \to \infty$.

Пример 5.2

Механический завод при изготовлении двух типов деталей использует токарное, фрезерное и сварочное оборудование. При этом обработку каждой детали можно вести двумя различными технологическими способами. Полезный фонд времени работы каждой группы оборудования (в станко-ч), нормы расхода времени при обработке детали на соответствующем оборудовании по данному технологическому способу и прибыль от выпуска единицы детали каждого вида приведены в таблице 5.7. Составить оптимальный план «загрузки оборудования», обеспечивающий заводу максимальную прибыль.

Таблица 5.7 – Исходные данные

Оборудование	Детали				Полезный фонд времени, станко-ч
	1		2		
	Технологические способы				
	1	2	3	4	
Фрезерное	2	2	3	0	20
Токарное	3	1	1	2	37
Сварочное	0	1	1	4	30
Прибыль,руб.	11	6	9	6	

Решение. Обозначим через:

- x_1 – количество деталей первого типа, обработанных по 1-му технологическому способу;
- x_2 – количество деталей первого типа, обработанных по 2-му технологическому способу;
- x_3 – количество деталей второго типа, обработанных по 3-му технологическому способу;
- x_4 – количество деталей второго типа, обработанных по 4-му технологическому способу.

Составим математическую модель задачи.

$$\max: f(x) = 11x_1 + 6x_2 + 9x_3 + 6x_4$$

$$\begin{cases} 2x_1 + 2x_2 + 3x_3 \le 20 \\ 3x_1 + x_2 + x_3 + 2x_4 \le 37 \\ x_2 + x_3 + 4x_4 \le 30 \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0.$$

Приведем ее к каноническому виду, т. е. ограничения неравенства заменяем на ограничения равенства:

$$\max: f(x) = 11x_1 + 6x_2 + 9x_3 + 6x_4$$

$$\begin{cases} 2x_1 + 2x_2 + 3x_3 + x_5 = 20 \\ 3x_1 + x_2 + x_3 + 2x_4 + x_6 = 37 \\ x_2 + x_3 + 4x_4 + x_7 = 30 \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0, x_5 \ge 0, x_6 \ge 0, x_7 \ge 0$$

Решение ЗЛП разбивается на два этапа:

- на первом этапе находится начальный опорный план или убеждаются в том, что такого плана не существует;
- на втором этапе производиться последовательное улучшение плана (базиса).

В нашей системе уравнений (*m*=3) базисные переменные легко определяются. Это – x_5, x_6, x_7, которые выражаются через x_1, x_2, x_3, x_4.

Для перехода ко 2-му этапу решения задачи, необходимо систему уравнений записать в таком виде:

$$\begin{cases} x_5 = 20 - (2x_1 + 2x_2 + 3x_3) \\ x_6 = 37 - (3x_1 + x_2 + x_3 + 2x_4) \\ x_7 = 30 - (x_2 + x_3 + 4x_4) \end{cases}$$

Для определения значений базисных переменных x_5, x_6, x_7 необходимо приравнять к нулю свободные переменные x_1, x_2, x_3, x_4. Полученное таким образом решение называется базисным - (0, 0,0, 0, 20, 37, 30).

После этого переходим ко второму этапу решения задачи. Рассмотрим симплекс - метод.

Таблица 5.8

БП	СЧ	БП, СП						
		$-x_1$	$-x_2$	$-x_3$	$-x_4$	$-x_5$	$-x_6$	$-x_7$
x_5=	20	$\boxed{2}$	2	3	0	1	0	0 $\Rightarrow$
x_6=	37	3	1	1	2	0	1	0
x_7=	30	0	1	1	4	0	0	1
F=	0	-11	-6	9	-6	0	0	0
		$\Uparrow$						

После заполнения исходной симплекс-таблицы (табл. 5.8.) начинается подготовка к заполнению второй таблицы. Для этого используется алгоритм симплекс-метода.

1) Проверка базисного решения на оптимальность. Просматриваем знаки коэффициентов функции f (последняя строка таблицы, кроме коэффициентов при свободном члене и базисных переменных). Наличие отрицательных коэффициентов, говорит о том, что исходное решение еще не оптимально.
2) Проверяем задачу на наличие решения. Так как под всеми отрицательными коэффициентами целевой функции нет ни одного столбца с неположительными числами, это значит, что задача имеет решение.
3) Выбираем из небазисных переменных ту, которая способна при введении ее в базис увеличить значения целевой функции: $\max\limits_{j \in n} \{\|c_j\|\}, c_j < 0$. Отмечаем этот столбец $\Uparrow$, $\boxed{x_1}$
4) Определяем, какая из базисных переменных должна будет выведена из базиса. Для этого определяем минимальное частное от деления соответствующих свободных членов и положительных коэффициентов

Таблица 5.9

БП	СЧ	БП, СП						
		$-x_1$	$-x_2$	$-x_3$	$-x_4$	$-x_5$	$-x_6$	$-x_7$
x_1=	10	1	1	3/2	0	½	0	0
x_6=	7	0	-2	-7/2	$\boxed{2}$	-3/2	1	0 $\Rightarrow$
x_7=	30	0	1	1	4	0	0	1
f=	110	0	5	15/2	-6	11/2	0	0
					$\Uparrow$			

— это переменная x_5. Эту строку отметим $\Rightarrow$.

5) Вводимую в базис переменную x_1 выражаем через переменную x_5, выводимую из базиса и небазисные переменные x_2, x_3, x_4. Для этого составляем новую симплекс- таблицу (табл. 5.9.).

В ней базис выражается переменными x_1, x_6, x_7. Делим строку табл. 5.8., отмеченную $\Rightarrow$, на разрешающий элемент.

6) Все остальные базисные переменные x_6, x_7 и целевую функцию выражаем через новые небазисные переменные x_2, x_3, x_4, x_5, используя правило прямоугольника. После заполнения табл. 5.9. расчет повторяется с пункта 1).

Таблица 5.10

БП	СЧ	БП, СП						
		$-x_1$	$-x_2$	$-x_3$	$-x_4$	$-x_5$	$-x_6$	$-x_7$
x_1=	10	1	1	3/2		1/2	0	0
x_4=	7/2	0	-1	-7/4	1	-3/4	1/2	0
x_7=	16	0	5	$\boxed{8}$	0	3	-2	0$\Rightarrow$
f=	131	0	-1	-3	0	1	3	1

$\Uparrow$

Таблица 5.11

БП	СЧ	БП, СП						
		$-x_1$	$-x_2$	$-x_3$	$-x_4$	$-x_5$	$-x_6$	$-x_7$
x_1=	7	1	1/16	0	0	-1/16	3/8	-3/16
x_4=	7	0	3/32	0	1	-3/32	1/2	7/32
x_3=	2	0	5/8	1	0	3/8	-1/4	1/8
	137	0	7/8	0	0	17/8	9/4	3/8

В таблице 5.11. последняя строка не содержит отрицательных коэффициентов, следовательно, находим оптимальное решение, максимизирующее критерий оптимальности. Решение задачи следующее:

$X^{опт}$=(7, 0,2,7), f_{max}= 11*7+0*6+9*2+6*7=137

Лекция 6. Двойственность задач линейного программирования. Влияние изменения параметров исходной задачи на значение целевой функции. Совместное решение двойственных задач.

Двойственность задач линейного программирования

Каждой задаче линейного программирования соответствует двойственная задача. Двойственная задача по отношению к исходной задаче строится по следующим правилам:

1. Если исходная задача ставится на максимум, то двойственная ставится на минимум и наоборот.
2. Коэффициенты целевой функции исходной задачи становятся правыми частями ограничений двойственной задачи. Правые части ограничений исходной задачи становятся коэффициентами целевой функции двойственной задачи.
3. Если A – матрица коэффициентов исходной задачи, то транспонированная матрица A^T будет матрицей коэффициентов двойственной задачи.
4. В задаче на максимум все ограничения имеют знак $\leq$ $(=)$, а в задаче на минимум все ограничения имеют знак $\geq$.
5. Число переменных в двойственной задаче равно числу ограничений в исходной задаче. Каждому ограничению исходной задачи соответствует переменная двойственной задачи. Если ограничение исходной задачи имеет знак ($\geq$), то соответствующая переменная двойственной задачи неотрицательна. Если ограничение имеет знак (=), то соответствующая переменная двойственной задачи может принимать положительные и отрицательные значения и наоборот.

В матричном виде двойственные задачи, заданные в симметричной форме, имеют вид

<u>Прямая задача</u>	<u>Двойственная задача</u>	
$\max\limits_{x}\left(c^T, x\right)=\left(c^T, x^*\right)$	$\min\limits_{y}\left(b^T, y\right)=\left(b^T, y^*\right)$	
$A \cdot x \leq b,$	$A^T \cdot y \geq c,$	(6.1)
$x \geq 0,$	$y \geq 0,$	

где $c = (c_1, c_2, ..., c_n)^T$, $x = (x_1, x_2, ..., x_n)^T$, $A = \|a_{ij}\|_{mxn}$,

$b = (b_1, b_2, ..., b_n)^T$, $y = (y_1, y_2, ..., y_n)^T$,

$x^* = \left(x_1^*, x_2^*, ..., x_n^*\right)^T$, $y^* = \left(y_1^*, y_2^*, ..., y_n^*\right)^T$ – *оптимальные решения задач.*

Переменные $(y_1, y_2, ..., y_n)$ называются двойственными (или объективно обусловленными) оценками.

Связь двойственных задач представлена в табл. 6.1

Таблица 6.1 – Связь двойственных задач

Прямая задача	Двойственная задача
$\max \sum_{j=1}^{n} c_j x_j$	$\min \sum_{i=1}^{m} b_i y_i$
$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1, m_1}$	$y_i \ge 0, \quad i = \overline{1, m_1}$
$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{m_1 + 1, m}$	y_i – *любые*, $\quad i = \overline{m + 1, m_1}$
$x_j \ge 0, \quad j = \overline{1, n_1}$	$\sum_{i=1}^{m} a_{ij} y_i \ge c_j, \quad j = \overline{1, n_1}$
x_j – *любые*, $\quad j = \overline{n_1 + 1, n}$	$\sum_{i=1}^{m} a_{ij} y_i = c_j, \quad j = \overline{n_1 + 1, n}$

Экономическая интерпретация задач (6.1) следующая. Вектор x – это вектор выпускаемой продукции, y – двойственные оценки ресурсов прямой задачи. Левая часть ограничений двойственной задачи представляет собой оценку затрат на единицу выпускаемой продукции в двойственных ценах.

Прямая задача представляет собой задачу на определение плана, обеспечивающего максимальный выпуск продукции при заданных ценах реализации и ограничениях на ресурсы. Двойственная задача – определение таких оценок ресурсов, в которых стоимость имеющихся ресурсов минимальна, а затраты на производство единицы продукции не меньше цен реализации продукции.

Сформулируем важные теоремы линейного программирования.

Первая теорема двойственности. Для взаимно двойственных задач имеет место один из трех случаев:

1. Если существует решение одной задачи, то существует решение и второй задачи. Значения целевых функций на

оптимальных решениях обеих задач равны $\left(c^T, x^*\right) = \left(b^T, y^*\right)$ и на множестве допустимых значений обеих задач выполняется неравенство $\left(c^T, x\right) \le \left(b^T, y\right)$. Если на допустимых решениях обеих задач целевые функции равны, то решения оптимальны.

2. Если решение одной задачи неограниченно, то другая задача несовместна.

3. Обе задачи несовместны

Вторая теорема двойственности. Оптимальные решения двойственных задач удовлетворяют соотношениям

$$
\begin{aligned}
&y_i \cdot \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) = 0, \quad i = \overline{1, m}, \\
&x_j \cdot \left(\sum_{i=1}^{m} a_{ij} y_i - c_j \right) = 0, \quad j = \overline{1, n}.
\end{aligned}
\qquad (6.2)
$$

Имеет место и обратное свойство: если допустимые значения переменных x_j, y_i удовлетворяют соотношениям (6.2), то они являются оптимальными решениями обеих задач.

Экономический смысл второй теоремы двойственности состоит в следующем: 1) если оптимальная оценка i-го ресурса не равна нулю $\left(y_i^* > 0\right)$, то в оптимальном плане этот ресурс используется полностью $\sum_{j=1}^{n} a_{ij} x_j^* = b_i$;

2) если в оптимальном плане ресурс не используется полностью $\sum_{j=1}^{n} a_{ij} x_j^* < b_i$, то его оценка равна нулю $\left(y_i^* = 0\right)$;

3) если j-й продукт входит в оптимальный план $\left(x_j^* > 0\right)$, то в оптимальных оценках ресурсов он неубыточен $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$;

4) если j -й продукт в оптимальных оценках ресурсов убыточен $\sum_{i=1}^{m} a_{ij} y_i^* > c_j$, то он не входит в оптимальный план $\left(x_j^* = 0\right)$.

Влияние изменения параметров исходной задачи на значение целевой функции

Запишем функцию Лагранжа для задачи ЛП:

$$L(x,y)=\sum_{j=1}^{n}c_j x_j+\sum_{i=1}^{m}y_i\left(b_i-\sum_{j=1}^{n}a_{ij}x_j\right) \tag{6.3}$$

Перегруппировав слагаемые, получим

$$L(x,y)=\sum_{i=1}^{nm}b_i y_i+\sum_{j=1}^{n}x_j\left(c_j-\sum_{i=1}^{m}a_{ij}y_i\right) \tag{6.4}$$

Из первой и второй теорем двойственности следует равенство

$$L(x^*,y^*)=\sum_{j=1}^{n}c_j x_j^*=\sum_{i=1}^{m}b_i y_i^* \tag{6.5}$$

Дифференцируя (6.4), получим три соотношения:

$$\frac{\partial L(x^*,y^*)}{\partial b_i}=y_i^*,$$

$$\frac{\partial L(x^*,y^*)}{\partial x_j}=c_j-\sum_{i=1}^{m}a_{ij}y_i^*, \tag{6.6}$$

$$\frac{\partial L(x^*,y^*)}{\partial a_{ij}}=-x_j^* y_i^*.$$

Из соотношений (6.5) и (6.6) можно сделать несколько выводов.

1. Вектор $y^*=\left(y_1^*, y_2^*, ..., y_m^*\right)$ дает градиент целевой функции в оптимальном плане, при изменении правых частей ограничений исходной задачи (запасов ресурсов) на единицу. Отсюда следует, что изменение значения целевой функции в оптимальном плане при небольшом изменении объемов имеющихся ресурсов будет приближенно равно

$$\Delta\left(\sum_{j=1}^{n}c_j x_j^*\right)=\sum_{i=1}^{m}y_i^*\Delta b_i \tag{6.7}$$

Следовательно, незначительное изменение запасов недефицитных ресурсов не изменит целевую функцию, т.к. для них $y_i^*=0$.

2. Величины $d_j = \left(c_j - \sum_{i=1}^{m} a_{ij} y_i \right)$ дают изменение (градиент) целевой функции исходной задачи при изменении выпуска продукции на единицу в оптимальном плане. Из (6.2) следует, что для продуктов, вошедших в оптимальный план, $d_j = 0$, а для не вошедших в оптимальный план продуктов - $d_j < 0$.

Отсюда следует, что небольшое изменение оптимального плана по вошедшим в план продуктам не даст изменения целевой функции при тех же ресурсах, а выпуск небольшого количества убыточного продукта (который не вошел в оптимальный план) приведет к снижению значения целевой функции на величину

$$\Delta Z_j = \left(c_j - \sum_{i=1}^{m} a_{ij} y_i \right) \Delta x_j^* \qquad (6.8)$$

3. Величины $(-x_j^*, y_i^*)$ образуют градиент целевой функции относительно изменения технологических коэффициентов a_{ij}. Уменьшение этих коэффициентов ведет к росту целевой функции. Максимальный рост целевая функция получает при уменьшении значений технологических коэффициентов, при этом достигается

$$\max_{ij} x_j^* y_i^* \qquad (6.9)$$

Совместное решение двойственных задач

Решение двойственной задачи получается по значению F-строки последней симплексной таблицы прямой задачи следующим образом.

1. Устанавливается соответствие между переменными двойственных задач.
2. Значения F-строки свободных X-переменных приравниваются соответствующим значениям Y-переменных, остальные значения Y-переменных приравниваются к нулю.

Пример 6.1 Найти решение двойственной задачи по решению прямой.

Прямая задача	Двойственная задача
$f = 7.5x_1 + 3x_2 + 6x_3 + 12x_4 \to \max$	$g = 3400y_1 + 1200y_2 + 3000y_3 \to \min$
$2x_1 + x_2 + 0.5x_3 + 4x_4 \le 3400\ (+x_5) <-> y_1$	$2y_1 + y_2 + 3y_3 \ge 7.5(-y_4) <-> x_1$
$x_1 + 5x_2 + 3x_3 \qquad \le 1200(+x_6) <-> y_2$	$y_1 + 5y_2 \ge 3(-y_5) <-> x_2$
$3x_1 + \quad + 6x_3 + x_4 \quad \le 3000(+x_7) <-> y_3$	$0.5y_1 + 3y_2 + 6y_3 \ge 6(-y_6) <-> x_3$
$x_i \ge 0, \quad i = \overline{1,7}$	$4y_1 + y_3 \ge 12(-y_7) <-> x_4$
	$y_i \ge 0, \quad i = \overline{1,7}$

Таблица 6.2 –Таблица соответствия переменных

x_1	x_2	x_3	x_4	x_5	x_6	x_7
y_4	y_5	y_6	y_7	y_1	y_2	y_3

Заключительная симплекс-таблица.

Таблица 6.3 – Заключительная симплекс-таблица

БП	СЧ	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_4	799	0.46	0	0	1	0.25	-0.05	0
x_2	20	-0.05	1	0	0	0.03	0.19	-0.1
x_3	367	0.43	0	1	0	-0.04	0.01	0.17
f	11850	4	0	0	0	2.815	0.04	0.75
		y_4	y_5	y_6	y_7	y_1	y_2	y_3

Таблица 6.3 – Оптимальные планы двойственных задач

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7
f	4	0	0	0	2.815	0.04	0.75
y	y_4	y_5	y_6	y_7	y_1	y_2	y_3

Оптимальный план прямой задачи

$(x_1^* = x_5^* = x_7^* = x_6^* = 0, x_2^* = 20, x_3^* = 367, x_4^* = 799)$.

Оптимальный план двойственной задачи.

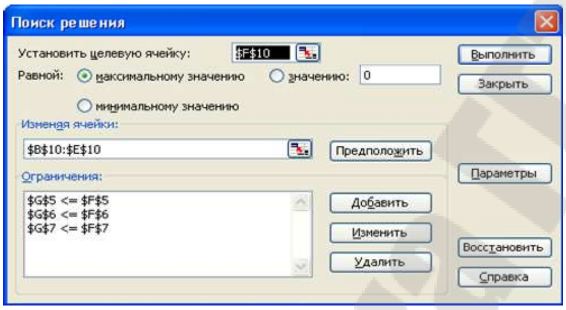
Значение целевой функции прямой и двойственной задач равно $f^* = g^* = 11850$.

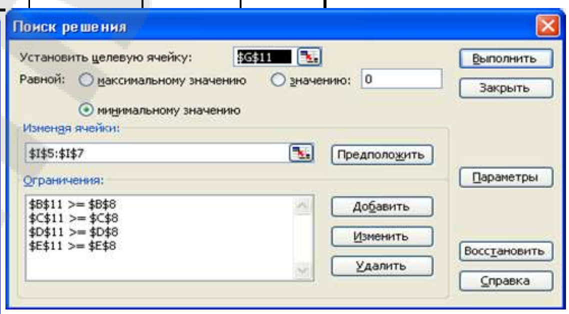
Экономический смысл оптимального решения двойственных задач представлен в таблице 6.2.

Из таблицы видно, что все ресурсы используются полностью, т.е. являются дефицитными, их остатки равны нулю. Оценки этих ресурсов равны соответственно $y_1^* = 2.815, y_2^* = 0.04, y_3^* = 0.75$. В оптимальном плане первый продукт не выпускается $x_1^* = 0$, он является убыточным, т.к. превышение затрат над ценой у него равно $y_4^* = 4$. Продукты второй, третий и четвертый выпускаются в оптимальном плане $x_2^* = 20, x_3^* = 367, x_4^* = 799$ и являются неубыточными, превышение затрат над ценой у них. При реализации оптимального плана предприятие получит максимально возможную прибыль, равную 11850 ден. ед.

Решения прямой и двойственной задач можно получить с помощью функции «ПОИСК РЕШЕНИЯ» в системе EXCEL.

	Детали							
Оборудование	1		2		Полезный фонд времени,	Ограничения прямой задачи	План двойственной задачи	Оптимальный план
	Технологические способы							
	1	2	3	4				
Фрезерное	2	1	0,5	4	3400	3400	y1	2,813278
Токарное	1	5	3	0	1200	1200	y2	0,0373444
Сварочное	3	0	6	1	3000	3000	y3	0,746888
Прибыль,руб.	7,5	3	6	12				
План прямой задачи	x1	x2	x3	x4	max f			
Оптимальный план	0	19,917	366,805	799,17	11850,62241			
Ограничения двойственной задачи	7,904564	3	6	12	min g	11850,62241		





ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Сформулируйте общую задачу математического программирования.
2. Как общую задачу линейного программирования привести к канонмческому виду.
3. Как решить графически ЗЛП с двумя переменными?
4. Что такое симплексное преобразование?
5. Как найти начальный опорный план?
6. Как найти оптимальное решение ЗЛП?
7. Как найти решение двойственной задачи по решению прямой?

ТЕМА 6. ТРАНСПОРТНЫЕ ЗАДАЧИ

Лекция 7. Постановка и типы транспортной задачи. Определение исходного опорного плана: правило «северо-западного угла, правило «северо-западного угла, способ аппроксимации Фогеля. Метод потенциалов.

Оптимальные назначения или проблема выбора. Венгерский метод. Оптимальное закрепление за станками операций по обработке деталей. Задачи размещения с учетом транспортных и производственных затрат.

Постановка и типы транспортной задачи

Транспортная задача в общем виде формулируется так.

Пусть имеется m пунктов отправления (или пунктов производства) $A_1, A_2, ..., A_m$ и n пунктов назначения (или пунктов потребления) $B_1, B_2, ..., B_n$. Обозначим ресурсы груза в i –м пункте отправления через a_i ,$\left(i = \overline{1, m}\right)$, а потребность каждого j- го пункта потребления через b_j, $\left(j = \overline{1, n}\right)$. Заданы стоимости перевозки единицы от каждого i –ого пункта отправления до каждого j- го пункта потребления. Требуется определить, какое количество груза $x_{ij} \geq 0$ необходимо перевезти от каждого i –ого пункта отправления до каждого j- го пункта потребления, чтобы:

- ввезти грузы всех поставщиков;
- удовлетворить всех потребителей;
- достигнуть экстремума целевой функции.

Транспортную задачу условно будем называть таблицей, в которой указаны поставщики, потребители, запасы и потребности грузов, тарифы перевозок единицы груза от i-того поставщика к j-тому потребителю (табл. 7.1).

Матрица $\left\| c_{ij} \right\|_{mxn}$ называется *матрицей тарифов* (издержек или транспортных расходов). *Планом* транспортной задачи называется матрица $X = \left\| x_{ij} \right\|_{mxn}$, где каждое число x_{ij} обозначает количество единиц груза, которое надо доставить из i-го пункта отправления в j-й пункт назначения.

Общие суммарные затраты, связанные с реализацией плана перевозок, можно представить целевой функцией:

$$f = c_{11}x_{11} + c_{12}x_{12} + ... + c_{1n}x_{1n} + c_{21}x_{21} + c_{22}x_{22} + ... + c_{2n}x_{2n} + ... + \\ + c_{m1}x_{m1} + c_{m2}x_{m2} + ... + c_{mn}x_{mn}. \quad (7.1)$$

Таблица 7.1 – Транспортная задача

Поставщик	Потребитель				Запас груза
	B_1	B_2	…	B_n	
A_1	c_{11} x_{11}	c_{12} x_{12}	…	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	…	c_{2n} x_{2n}	a_2
…	…	…	…	…	…
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	…	c_{mn} x_{mn}	a_m
Потребность в грузе	b_1	b_2	…	b_n	

Переменные x_{ij} должны удовлетворять ограничениям по запасам, по потребителям и условиям неотрицательности. В математической записи условие задачи можно представить так:

$$\left.\begin{aligned} &\sum_{j=1}^{n} x_{ij} = a_i \quad (i = 1, ..., m); \\ &\sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1, ..., n); \end{aligned}\right\} \quad (7.2)$$

$$x_{ij} \geq 0 \quad (i = 1, ..., m; \; j = 1, ..., n). \quad (7.3)$$

Таким образом, математически транспортная задача ставится следующим образом. Дана система ограничений (7.2) при условии (7.3) и целевая функция (7.1). Требуется среди множества решений системы найти такое неотрицательное решение (план) перевозок, которое минимизирует целевую функцию (7.1).

Система ограничений (7.11)-(7.3) содержит $m+n$ уравнений с $m \times n$ переменными. Предполагается, что суммарные запасы равны суммарным потребностям, т.е.

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \, . \quad (7.4)$$

Различают два типа транспортных задач: *закрытые,* в которых суммарный объем груза поставщиков равен суммарному спросу потребителей, т.е. (7.4) и *открытые*, в которых суммарная производственная мощность поставщиков превышает спрос потребителей или в которых спрос потребителей больше фактической суммарной мощности поставщиков, т.е.

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j \quad \textit{или} \quad \sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j \ .$$

Равенство (7.4) является необходимым и достаточным условием совместности ограничений задачи.

Для транспортной задачи важное значение имеет теорема о ранге матрицы.

Теорема. *Ранг матрицы транспортной задачи на единицу меньше числа уравнений, т.е. r(A)=m+n-1.*

С учетом этой теоремы в каждой матрице перевозок опорный план должен содержать не более m+n-1 занятых клеток, а остальные свободные.

Определение исходного опорного плана

Существуют различные способы построения опорного плана. Рассмотрим правило «северо-западного угла», правило «минимального элемента» и способ аппроксимации Фогеля.

Правило «северо-западного угла»

Будем заполнять таблицу, начиная с левого верхнего (северо-западного) угла, двигаясь далее по строке вправо или по столбцу вниз. Занесем в клетку (1;1) меньшее из чисел a_1 и b_1, т.е. $x_{11}=\min(a_1;b_1)$. Если $a_1>b_1$, то $x_{11}=b_1$, и первый столбец «закрыт» для заполнения остальных его клеток, т.е. $x_{i1}=0$ для $i=2, \ldots, m$ (потребности первого потребителя удовлетворены полностью).

Двигаясь далее по первой строке, записываем в соседнюю клетку (1;2) меньшее из чисел a_1-b_1, b_2, т.е. $x_{12}=\min(a_1-b_1, b_2)$.

Если $b_1>a_1$, то аналогично «закрывается» первая строка, т.е. $x_{1k}=0$ для $k=2,\ldots, n$. Переходим к заполнению соседней клетки (2;1), куда заносим $x_{21}=\min(a_2, b_1-a_1)$.

Заполнив клетку (1;2) или (2;1), переходим к заполнению следующей третьей клетки либо по второй строке, либо по второму столбцу. Будем продолжать этот процесс до тех пор, пока на каком-то этапе не исчерпаются ресурсы a_m и потребности b_n. Последняя заполненная клетка окажется лежащей в последнем n-м столбце и в последней m- ой строке.

Пример 7.1**.** Определить исходное решение по правилу «северо-западного угла» следующей транспортной задачи (табл. 7.2).

Таблица 7.2 – Исходные данные

a_i \ b_k	40	25	20	50
60	5	4	1	2
40	4	2	6	3
35	7	3	5	4

Решение. В клетку (1;1) помещаем x_{11}=min (60; 40)= 40. Спрос первого потребителя удовлетворен полностью, первый столбец из расчета исключается. Остаток груза от первого поставщика (60-40=20 ед.) с учетом потребности второго потребителя помещаем в клетку (1;2), т.е. x_{12}= min (20;25)=20. В этом случае запас первого поставщика исчерпан, переходим к распределению груза второго поставщика. В клетку (2;2) помещаем необходимое количество груза, т.е. x_{22}= min(40; 5)= 5. В этом случае спрос второго потребителя удовлетворен полностью. Второй столбец в расчет не принимается, остаток груза от второго поставщика с учетом спроса третьего потребителя помещаем в клетку (2;3), т.е. x_{23}= min(35; 20)= 20. Спрос третьего потребителя удовлетворен полностью. Остаток груза второго поставщика с учетом спроса четвертого потребителя помещаем в клетку (2;4), т.е. x_{24}= min(15; 50)=5. В этом случае запас второго поставщика исчерпан. Переходим к распределению запаса у третьего поставщика. В клетку (3;4) помещаем необходимое количество груза x_{34}= min (35;35)=35. Получили полное распределение запасов груза.

Если число занятых клеток не удовлетворяет условию m+n-1, то план перевозок называют *вырожденным*. В данном случае получен невырожденный план

$$X_1 = \begin{pmatrix} 40 & 20 & 0 & 0 \\ 0 & 5 & 20 & 15 \\ 0 & 0 & 0 & 35 \end{pmatrix}$$

Этому плану соответствует значение целевой функции $f_1 = 5 \cdot 40 + 4 \cdot 20 + 2 \cdot 5 + 6 \cdot 20 + 3 \cdot 15 + 4 \cdot 35 = 595$.

Правило «минимальног о элемента»

Сущность состоит в следующем. Распределение груза у поставщиков начинается с клетки, которой соответствует наименьший тариф c_{ij} из всей матрицы тарифов. В клетку с наименьшим тарифом помещают наименьшее из чисел a_i или b_j. Затем из рассмотрения исключают строку, соответствующую поставщику, запасы которого полностью израсходованы, или столбец, соответствующий потребителю, спрос которого полностью удовлетворен. Может оказаться, что следует исключить строку и столбец одновременно, если полностью израсходованы запасы поставщика и полностью удовлетворен спрос потребителя. Далее из оставшихся строк (столбцов) таблицы снова выбирают наименьший тариф и процесс распределения запасов продолжают до тех пор, пока все они не будут распределены, а спрос удовлетворен.

Пример 7.2. Определить исходное опорное решение по правилу «минимального элемента» транспортной задачи, представленной в табл. 7.2, и сравнить значения целевых функций, полученных по правилам «минимального элемента» и «северо-западного угла».

Решение. Наименьший тариф имеем для клетки (1;3), c_{13}=1. Поэтому в данную клетку помещаем количество груза x_{13}= min (60; 20)=20 и исключаем из дальнейшего рассмотрения третий столбец. Снова находим в таблице наименьший тариф c_{14}=c_{22}= 2. Поместим, например, необходимое количество груза в клетку (1;4), x_{14} =min(60-20; 50)= 40. Исключается из дальнейшего рассмотрения первая строка. Далее в клетку (2;2) помещаем количество груза x_{22}=min(40; 25)=25 и исключаем из дальнейшего рассмотрения второй столбец. В оставшейся матрице тарифов минимальным будет тариф для клеток (2; 4), (3; 2)., c_{24}=c_{32}= 3. Однако второй столбец уже исключен из рассмотрения, поэтому загружаем клетку (2;4), x_{24}=min(40-25; 50-40)=10, и исключаем из дальнейшего рассмотрения четвертый столбец. Новый наименьший тариф для оставшихся тарифов будет c_{21}=4. В клетку (2;1) помещаем количество груза x_{21}=min(15-10; 40)= 5. Исключаем из дальнейшего рассмотрения вторую клетку.

На последнем этапе выберем единственный невычеркнутый первый столбец и единственную третью строку, для которых имеем общую свободную клетку (3;1) с тарифом c_{31}=7. В нее помещаем остаток груза x_{31}=min(35; 40-5)=35. В результате полного распределения грузов получим исходное опорное решение

$$X_2 = \begin{pmatrix} 0 & 0 & 20 & 40 \\ 5 & 25 & 0 & 10 \\ 35 & 0 & 0 & 0 \end{pmatrix},$$

для которого значение целевой функции f_2=415.

Сравнивая значения целевых функций для опорных планов, полученных по правилам «северо-западного угла» и «минимального элемента», замечаем, что транспортные издержки разные, а именно: f_1=595; f_2=415.

Способ аппроксимации Фогеля. В большинстве случаев этот способ дает опорный план, самый близкий к оптимальному.

В основе способа аппроксимации Фогеля лежит концепция штрафов, взимаемых за выбор не самого оптимального с точки зрения транспортных издержек маршрута. Штраф по каждой строке и по каждому столбцу определяется из анализа маршрутов с различными показателями издержек (как разность двух различных уровней транспортных издержек). Первой заполняется клетка матрицы (таблицы), в которой фиксируется самый крупный штраф. После заполнения клетки штрафы пересчитываются и так до тех пор, пока все ресурсы не будут распределены.

Алгоритм метода включает следующие основные этапы.

1. Вычисление разностей в каждой строке и в каждом столбце матрицы между наименьшей стоимостью и ближайшей к не величине. Разности по строкам записываются справа в столбце разностей, разности по столбцам – внизу в строке разностей.
2. Поиск из всех разностей как по строкам, так и по столбцам, максимальной. Обведем максимальную разность рамкой.

Размещение в клетку, где находится наименьшая стоимость, максимально возможного количества ресурсов.

3. Вычисление разностей по столбцам и строкам, не принимая во внимание стоимости в клетках, имеющих ресурсы, и клетках с точкой (исключенную строку или столбец), и определение максимальной разности в строке или столбце.
4. Поиск минимального элемента в строке или столбце с максимальной разностью и размещение в данную клетку максимального количества ресурса, возвращение к этапу 4) и т.д.

Пример7.3

	B1	B2	B3	B4	a_i	Столбцы разностей				
A1	5 ●	4 ●	1 20	2 40	60	1	2	-	-	
A2	4 40	2 ●	6 ●	3 ●	40	1	1	1	-	
A3	7 ●	3 25	5 ●	4 10	35	1	1	1		
b_j	40	25	20	50						
Строки	1	1	4	1						
разностей	1	1	-	1						
	3	1	-	1						
	-	-	-	-						

Метод потенциалов

Для решения транспортной задачи методом потенциалов необходимо:

- получить опорный план перевозок по одному из изложенных правил;
- вычислить потенциал u_i и v_j, соответственно, поставщиков и потребителей;
- вычислить сумму потенциалов (косвенные тарифы) для свободных клеток $u_i + v_j = c'_{ij}$;
- проверить разность $s_{ij} = c_{ij} - c'_{ij} = c_{ij} - (u_i + v_j)$.

Если все разности свободных клеток $s_{ij} \geq 0$, то полученный план оптимальный. Если хотя бы одна разность $s_{ij}<0$, то в число занятых клеток вводят переменную x_{ik}, для которой эта разность минимальна, и получают новый план перевозок. Процесс продолжают до тех пор, пока не будет определен план, для которого все разности $s_{ij} \geq 0$.

<u>Пример 7.4</u>**.** Решить транспортную задачу, представленную в табл. 7.2, методом потенциалов.

<u>Решение.</u> Пусть, например, исходное опорное решение получено по правилу «минимального элемента».

Итак, имеем опорное решение, представленное табл. 7.3.

Таблица 7.3– Опорное решение

	40	25	20	50	u_i
60	5	4	1 20	2 40	0
.40	+ 4 5	-2 25	6	3 10	1
35	7 - 35	3 +	5	4	4
v_j	3	1	1	2	

Полученный план невырожденный (см. табл. 7.3). Число занятых клеток удовлетворяет условию r=3+4-1=6. Для определения потенциалов поставщиков и потребителей будем иметь шесть уравнений:

$u_1+v_3=1$; $u_1+v_4=2$; $u_2+v_1=4$;
$u_2+v_2=2$; $u_2+v_4=3$; $u_3+v_1=7$.

Поскольку число уравнений на единицу меньше числа неизвестных, одно из неизвестных оказывается свободным и может принимать любое числовое значение. Положим, например, u_1=0. Тогда все остальные потенциалы для данного опорного решения определяются однозначно, а именно: u_1=0; v_4=2; u_2=1; v_2=1; v_1=3; v_3=1; u_3=4.

Найденные потенциалы поставщиков и потребителей указаны справа и внизу (табл. 7.3). Расчет их можно производить непосредственно в таблице, не решая систему уравнений. Очевидно, неизвестный потенциал поставщика (потребителя) определяется разностью между тарифом занятой клетки и известным потенциалом поставщика (потребителя) – строки (столбца), на пересечении которых расположена занятая клетка.

Определим разности для свободных клеток s_{ij} по формуле $s_{ij}=c_{ij}-(u_i+v_j)$. Получаем, s_{11}=5-(0+3)=1; s_{12}=4- (0+1)=3; s_{23}=6- (1+1)=4; s_{32}=3-(4+1)=-2; s_{33}=5-(4+1)=0; s_{34}=4-(4+2)=-2.

Полученный план не является оптимальным, так как среди разности s_{ij} имеются отрицательные: $s_{32}=s_{34}$=-2. Получены две перспективные (потенциальные) клетки с одинаковой оценкой. Загрузим, например, клетку (3;2). В таком случае для нее строится замкнутый цикл, который включает клетки: (3;2), (3;1), (2;1), (2;2). В свободной клетке указанного цикла условно запишем знак плюс, в следующей клетке (по часовой стрелке или против часовой стрелке) – минус, а в следующей плюс и т.д. В отрицательных вершинах цикла выбираем наименьшее количество груза min(25;35)=25 ед., т.е. x_{32}=25 ед. Чтобы общий баланс цикла не изменился, необходимо в клетки цикла со знаком плюс прибавить 25 ед. груза, а из клеток со знаком минус вычесть столько же единиц. Новый опорный план приведен в табл. 7.4.

Он невырожденный. Потенциалы поставщиков и потребителей представлены в таблице. Найдем разности s_{ij} для свободных клеток: s_{11}=5-(0+3)=2; s_{22}=2-(1+1)=0; s_{12}=4- (0+1)=3; s_{23}=6- (1+1)=4; s_{33}=5-(4+1)=0; s_{34}=4-(4+2)=-2.

План, представленный табл. 7.4., еще не оптимальный, разность s_{34}= -2. За счет загрузки клетки (3;4), план можно улучшить. Замкнутый цикл для клетки (3;4) виден из табл. 7.4.

После смещения 10 ед. груза получаем новый опорный план (табл. 7.5). Он вырожденный. Число занятых клеток не удовлетворяет условию r =3+4-1=6. Число занятых клеток равно 5.

Таблица 7.4 –Новый опорный план

	40	25	20	50	u_i
60	5	4	1 20	2 40	0
40	+ 4 30	-2	6	- 3 10	1
35	7 - 10	3 25	5	+ 4	4
v_j	3	1	1	2	

Таблица 7.5 – Оптимальный план

	40	25	20	50	
60	5	4	1 20	2 40	0
40	4 40	-2	6	-3	1
35	7	3 25	5	4 10	4
	3	1	1	2	

Поместим нуль, например в клетку (2;2) и будем считать эту клетку занятой. Очевидно, загрузка данной клетки не приводит к образованию замкнутого цикла из занятых клеток.

Потенциалы поставщиков и потребителей для плана, представленного в табл. 7.5, определены в этой же таблице.

Полученный план (см. табл. 7.5) является оптимальным, так как он не содержит ни одной отрицательной разности s_{ij}. Оптимальный план имеет вид:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 0 & 20 & 40 \\ 40 & 0 & 0 & 0 \\ 0 & 25 & 0 & 10 \end{pmatrix}.$$

Следует отметить, что полученный оптимальный план не единственный. Так, возьмем разность, равную нулю для клетки (2;4). За счет загрузки этой клетки, значение целевой функции для плана не изменится (табл. 7.5), однако структура плана будет иная.

Оптимальные назначения или проблема выбора

Рассмотрим задачу нахождения наилучшего способа назначения людей на работу, при котором будет достигнуто максимальная суммарная производительность. Задача может быть интерпретирована в различных вариантах. Например, под видами работ, можно понимать различные изделия, строительные площадки и т.д. Работы могут выполняться станками, механизмами, приспособлениями.

Рассмотрим простейший вариант задачи о назначениях.

Пусть имеется n видов работ $A_1, \ldots, A_n$ и m механизмов $B_1, \ldots, B_m$, производительность которых c_{ij}. Требуется определить оптимальное назначение людей на работы, при котором будет достигнута максимальная суммарная производительность.

При такой интерпретации задачи важно, чтобы одновременно на каждый вид работы было назначено только одно лицо.

Рассмотрим случай, когда количество работ равно числу лиц, занятых их выполнением, т.е. $n=m$. Положим,

$$x_{ij} = \begin{cases} 1, \text{ } если \ i-ое \ лицо \ назначено \ j-ую \ работу \\ 0, \text{ } в \ противном \ случае \end{cases} \tag{7.5}$$

Так как каждое лицо можно назначать только на один вид работы, то

$$\sum_{j=1}^{n} x_{ij} = 1 \quad (i = 1, \ldots, m) \tag{7.6}$$

Так как каждая работа предназначена только для одного лица, то

$$\sum_{i=1}^{n} x_{ij} = 1 \quad (j = 1, \ldots, n) \tag{7.7}$$

Суммарная производительность выразится так:

$$f = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{7.8}$$

Итак, математически задача о назначениях формулируется в следующем виде. Даны системы ограничений (7.6) и (7.7) при условии (7.5). Требуется найти переменные x_{ij}, удовлетворяющие (7.6) и (7.7) и обращающие целевую функцию (7.8) в максимум.

Решать задачу о назначениях можно теми же методами, какими решается транспортная задача.

Венгерский метод

Прежде чем приступить к алгоритму, необходимо привести подготовительный этап. Заменяем задачу максимума (7.8) задачей минимума.

$$f' = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}{}' x_{ij}, \quad \text{где } c'_{ij} = l_j - c_{ij}$$

$$l_j = \max\{c_{ij}\} \quad i = 1, ..., n.$$

Основной принцип венгерского метода – оптимальность решения задачи о назначении не нарушается при уменьшении (увеличении) строки (столбца) на одну и ту же величину d_i (d_j).

Решение считается оптимальным, если все измененные искусственные затраты $c_{ij}{}' \geq 0\ (i, j = 1, ..., n)$ и можно отыскать такой набор x_{ij}, что $F' = \sum_{i=1}^{n} \sum_{j=1}^{n} c'_{ij} x_{ij} = \min$.

Алгоритм метода включает следующие основные этапы.

1. Получение нулей в каждой строке, для чего необходимо найти наименьший элемент в каждой строке d_i и вычесть его из всех ее элементов. Получаем новую таблицу. Аналогично делается для каждого столбца. Получаем новую таблицу.
2. Поиск оптимального решения. Для поиска оптимального решения необходимо рассмотреть сначала одну из строк таблицы, имеющую меньше нулей. Отметить точкой один из нулей этой строки и зачеркнуть все остальные нули этой строки и того столбца, в которых находится этот нуль. Аналогичные операции последовательно проводят для всех строк. Если назначения, которые получены при всех нулях, отмеченных точкой, являются полными (т.е. число нулей, отмеченных точкой, равно n), то решение является оптимальным. В противном случае следует переходить к следующему этапу.
3. Поиск минимального набора строк и столбцов, содержащих нули. Для этого необходимо отметить точкой:

   а) все строки, в которых не имеется ни одного отмеченного точкой нуля;

   б) все столбцы, содержащие перечеркнутый нуль, хотя бы в одной из отмеченных точкой строк;

   в) все строки, содержащие отмеченные точкой нули, хотя бы в одном из отмеченных точкой столбцов.

Действия б) и в) повторяются поочередно до тех пор, пока есть что отмечать. После этого необходимо зачеркнуть каждую непомеченную строку и каждый помеченный столбец. Цель этого этапа – провести минимальное число горизонтальных и вертикальных прямых, пересекающих по крайне мере, один раз все нули.

4. Перестановка некоторых нулей. Взять наименьшее число из тех клеток, через которые не проведены прямые. Вычесть его из каждого числа не вычеркнутых столбцов и прибавить к каждому числу вычеркнутых строк. Получим новую таблицу.

Эта операция не изменяет оптимального решения, после чего весь цикл начинается с этапа 2 и продолжается до получения оптимального решения.

Пример.7.5 Пусть для монтажа четырех объектов (n=4) требуется четыре крана. Из отчетных данных известно, какое время необходимо каждому крану A_i для монтажа объекта B_j. Нужно так распределить краны по объектам, чтобы суммарное время на монтаж этих объектов было минимально. В нашем случае c_{ij} – это затраты времени A_i – того крана при монтаже объекта B_j. Исходные данные представлены в табл. 7.6.

Таблица 7.6 –Исходные данные

A_i	B_j				a_i	d_i
	B_1	B_2	B_3	B_4		
A_1	3	7	5	8	1	3
A_2	2	4	4	5	1	2
A_3	4	7	2	8	1	2
A_4	9	7	3	8	1	3
b_j	1	1	1	1		

где d_i – минимальный элемент строки.

Решение

1) Получение нулей в каждой строке табл. 7.7.

Таблица 7.7 – Приведение по строкам

A_i	B_j				a_i
	B_1	B_2	B_3	B_4	
A_1	0	4	2	5	1
A_2	0	2	2	3	1
A_3	2	5	0	6	1
A_4	6	4	0	5	1
b_j	1	1	1	1	
d_j	0	2	0	3	

Получение нулей в каждом столбце табл. 7.8

Таблица 7.8 – Приведение по столбцам

A_i	B_j				a_i
	B_1	B_2	$B_3\bullet$	B_4	
A_1	$0\bullet$	2	2	2	1
A_2	$\cancel{0}$	$0\bullet$	2	$\cancel{0}$	1
$A_3\bullet$	2	3	$0\bullet$	3	1
$A_4\bullet$	6	2	$\cancel{0}$	2	1
b_j	1	1	1	1	

2) Поиск оптимального решения. Если число $0\bullet = n$, то решение оптимально. У нас нет, значит переходим к следующему пункту.
3) Поиск минимального набора строк и столбцов, содержащих нули.

а) отметить строки, в которых не имеется ни одного отмеченного точкой нуля – строка 4;

б) все столбцы, содержащие перечеркнутый нуль, хотя бы в одном из отмеченных точкой строк – столбец 3;

в) все строки, содержащие отмеченные точкой нули, хотя бы в одном из отмеченных точкой столбцов – строка 3;

Действия б) и в) повторяются поочередно до тех пор, пока есть что отмечать.

4) Перестановка некоторых нулей. Взять наименьшее число из тех клеток, которые не проведены прямые – у нас это число 2. Вычесть его из каждого числа невычеркнутых столбцов и прибавить к каждому числу вычеркнутых строк (табл.7.9.).

В нашем случае число нулей, отмеченных точкой, равно 4. Следовательно, назначение является полным, а план – оптимальным.

$$Z = c_{11} + c_{22} + c_{33} + c_{44} = 3 + 4 + 2 + 8 = 17$$

Таблица 7.9 – Оптимальная решение

A_i	B_j				a_i
	B_1	B_2	B_3	B_4	
A_1	$0\bullet$	2	4	2	1
A_2	$\cancel{0}$	$0\bullet$	4	$\cancel{0}$	1
A_3	$\cancel{0}$	1	$0\bullet$	1	1
A_4	4	$\cancel{0}$	$\cancel{0}$	$0\bullet$	1
b_j	1	1	1	1	

Оптимальное закрепление за станками операций по обработке деталей

Пусть на предприятии имеется m видов станков, максимальное время работы которых соответственно $a_i \ (i=1, 2, ..., \ m)$ часов. Каждый из станков может выполнять n видов операций. Суммарное время выполнения каждой операции соответственно равно $b_j \ (j=1, 2, ..., n)$ часов. Известно производительность (C_{ij}) i –го станка при выполнении j-ой операции. Определить, сколько времени и на какой операции нужно использовать каждый из станков, чтобы обработать максимальное количество деталей.

Для составления математической модели обозначим через $x_{ij} \ (i=1, 2, ..., m; \ j=1, 2, ..., n)$ время, которое i-й станок должен работать на j-й операции. Тогда количество деталей, обработанных на i-м станке, равно $C_{ij}x_{ij}$. Количество деталей, обработанных на всех станках, можно выразить функций $Z=\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}x_{ij}$. Так как максимально возможное время работы i-го станка ограничено (значением a_i), то получим $\sum_{j=1}^{n}x_{ij}=a_i, \quad i=1, 2, ..., m$, если максимальное время работы станков используется полностью, или $\sum_{j=1}^{n}x_{ij}\le a_i, \quad i=1, 2, ..., m$, если это время используется неполностью.

С другой стороны, время отведенное на j-ю операцию, равно b_j часов, поэтому $\sum_{i=1}^{m}x_{ij}=b_j, \quad j=1, 2, ..., n$. Из условия следует, что общее время станков, т.е. $\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\sum_{i=1}^{m}a_i$, и времени, необходимого на выполнение всех операций, т.е. $\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\sum_{j=1}^{n}b_j$. Отсюда следует, что $\sum_{i=1}^{m}a_i=\sum_{i=1}^{n}b_j$.

Таким образом, необходимо найти максимальное значение линейной функции $Z=\sum_{i=1}^{m}\sum_{j=1}^{n}C_{ij}x_{ij}$ при ограничениях

$$\sum_{j=1}^{n}x_{ij}=a_i,\, i=1, 2, \ ..., m, \quad x_{ij}\geq 0,$$

$$\sum_{i=1}^{m}x_{ij}=b_j, \quad j=1, 2, \ ..., n, \quad (i=1, 2, \ ..., m; j=1, 2, \ ..., n).$$

Задачи размещения с учетом транспортных и производственных затрат

При решении такого рода задач используется открытая модель транспортной задачи.

Задача 7.1 Несколько предприятий выпускают однородную продукцию, объем производства которой необходимо сократить и перейти к выпуску другой продукции, причем себестоимость единицы продукции на каждом предприятии различна. Определить, на каких предприятиях следует провести сокращение, чтобы суммарные расходы на производство и транспортировку продукции, выпускаемой после сокращения, были минимальными.

Пусть $(C'_{ij})(i=1, 2, ..., m;\ j=1, 2, ..., n)$ - матрица стоимостей перевозок единицы продукции от i- го поставщика к j-му потребителю, а C_i - стоимость производства единицы продукции i-го поставщика. Тогда, очевидно $(C_{ij})=(C'_{ij}+C_i)$ - матрица стоимостей рассматриваемой задачи.

Уменьшим на необходимую величину объем потребностей реальных потребителей и введем фиктивным потребителя, объем потребностей которого равен производственному сокращению. В результате избыток продукции будет закреплен за фиктивным потребителем. Предприятия, прикрепленные к фиктивному потребителю, и должны сократить объем производства.

Задача 7.2 Для удовлетворения потребностей в некоторой однородной продукции необходимо разместить предприятия таким образом, чтобы суммарные затраты на доставку сырья, производство продукции и ее транспортировку были минимальными.

Задачу решим в два этапа. Определим сначала затраты на транспортировку сырья и производство продукции в намеченных пунктах производства, число которых выберем заведомо большим, чем количество предприятий. Для этого построим открытую модель, в которой поставщики – возможные пункты добычи сырья, а потребители – намеченные пункты производства. После этого вновь построим открытую модель, причем поставщиками будут возможные пункты производства, а потребителями – возможные пункты потребления готовой продукции, которая решается по критерию минимума суммарных затрат на производство, транспортировку сырья и готовой продукции. В результате находим наиболее выгодный вариант размещения предприятий по производству продукции с точки зрения общих затрат на транспортировку и производство.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Какие типы транспортных задач бывают?
2. Как определить исходный опорный плана по правилу «северо-западного угла»?
3. Как определить исходный опорный плана по правилу «минимального элемента»?
4. В чем заключается способ аппроксимации Фогеля?
5. В чем суть метода потенциалов?
6. Сформулируйте постановку задачи о назначениях?
7. Как найти осуществить проблему выбора с помощью венгерского метода?
8. Как осуществить оптимальное закрепление за станками операций по обработке деталей?
9. Приведите примеры открытой модели транспортной задачи.

ТЕМА 7. ДИСКРЕТНАЯ ОПТИМИЗАЦИЯ

Лекция 8. Постановка задачи целочисленного линейного программирования. Классификация математических моделей дискретного программирования. Метод отсечения. Алгоритм Р. Гомори решения задачи целочисленного программирования.

Постановка задачи целочисленного программирования

На практике часто встречаются задачи с требованием целочисленности всех или части переменных, получивших соответственно название *задач целочисленного* или *частично целочисленного программирования*. В целочисленном программировании переменные изменяются не непрерывно, а принимают некоторые фиксированные (целые) значения.

Задача целочисленного линейного программирования (ЗЦЛП) ставится так: найти план $\boldsymbol{x}^*=(x_1^*, \ldots, x_n^*)$, при котором максимизируется (минимизируется) значение целевой функции

$$f = \sum_{j=1}^{n} c_j x_j \qquad (8.1)$$

при ограничениях

$$\sum_{j=1}^{n} a_{ij} x_j \{\le, =, \ge\} a_{i0} \quad (i = 1, \ldots, m), \qquad (8.2)$$

$$x_j \ge 0 \quad (j = 1, \ldots, n), \qquad (8.3)$$

$$x_j - \text{целые числа.} \qquad (8.4)$$

Классификация математических моделей дискретного программирования

Методы целочисленной оптимизации можно разделить на три основных группы:

а) методы отсечения;

б) комбинаторные методы;

в) приближенные методы.

Задачи целочисленного программирования подразделяются на линейные и нелинейные, статические и динамические, детерминированные и стохастические и т.д.

Как мы уже знаем, экстремум для задач линейного программирования достигается в крайней точке выпуклого множества или в точке, которая является выпуклой комбинацией крайних точек. Для целочисленного программирования точкой оптимума может быть любая точка области допустимых решений. Поэтому методы решения задач линейного программирования в рассмотренном ранее виде не пригодны.

Для решения задач целочисленного (особенно нелинейного) программирования используются *комбинаторные методы.* Это методы направленного частичного перебора допустимых планов.

Трудности машинной реализации точных методов решения задач целочисленного программирования привели к появлению различного рода *приближенных методов*, построенных на использовании особенностей конкретной задачи. Здесь наметилось два направления: 1) разработка детерминированных эвристических алгоритмов, использующих специфику задачи; 2) применение случайного поиска в сочетании с локальной оптимизацией.

Метод отсечения

Сущность *методов отсечения* состоит в том, что сначала задача решается без условия целочисленности. Если полученный план будет целочисленным, то задача решена. В противном случае к ограничениям задачи добавляется новое ограничение, обладающее следующими свойствами:

- оно должно быть линейным;
- должно отсекать найденный оптимальный нецелочисленный план;
- не должно отсекать ни одного целочисленного плана.

Дополнительное ограничение, обладающее указанными свойствами, называется *правильным отсечением*.

Затем полученная задача решается с учетом нового ограничения. После этого в случае необходимости добавляется еще одно ограничение и т.д.

Геометрически добавление каждого линейного ограничения отвечает проведению гиперплоскости, которая отсекает от многогранника планов некоторую его часть вместе с оптимальной точкой с нецелыми координатами, но не затрагивает ни одной из целочисленных точек этого многогранника.

Алгоритм Р. Гомори решения задачи целочисленного программирования

Р. Гомори разработал алгоритм решения ЗЦЛП (8.1)-(8.4), использующий симплексную процедуру. Гомори предложил простой способ построения правильного отсечения.

Предположим, что в результате решения симплекс-методом канонической ЗЛП (8.1)-(8.3) получен оптимальный план $(b_{10}, \ldots, b_{k0}, \ldots, b_{m0}, \ldots, 0, \ldots, 0)$, в котором, например, b_{k0} – нецелая компонента. В таком случае можно доказать, что неравенство

$$\{b_{k,m+1}\}x_{m+1} + \ldots + \{b_{k,n}\}x_n >= \{b_{k0}\}, \qquad (8.5)$$

сформированное по x_k –строке, обладает всеми свойствами правильного отсечения.

В неравенстве (8.5) присутствует символ {…}, означающий дробную часть числа. Дробной частью {a} числа a называют разность между этим числом и его целой частью [a], т.е. наибольшим целым, не превосходящим a.

Для решения ЗЦЛП (8.1)-(8.4) можно использовать следующий алгоритм.

1. Симплекс-методом решить ЗЛП (8.1)-(8.3). Если все компоненты оптимального плана целые, то он является оптимальным и для ЗЦЛП (8.1)-(8.4). Если ЗЛП (8.1)-(8.3) неразрешима, то и ЗЦЛП (8.1)-(8.4) также неразрешима. Если среди компонент оптимального плана есть нецелые, то перейти к п. 2.
2. Среди нецелых компонент выбрать компоненту с наибольшей дробной частью и по соответствующей строке симплекс- таблицы, содержащей нецелочисленный оптимальный план, сформировать правильное отсечение (8.5).
3. Неравенство (8.5) введением дополнительной неотрицательной целочисленной переменной преобразовать в эквивалентное уравнение и включить его в симплекс-таблицу с нецелочисленным оптимальным планом.
4. Полученную расширенную ЗЛП решить симплекс-методом. Если полученный оптимальный план будет целочисленным, то ЗЦЛП (8.1)-(8.4) решена. В противном случае вернуться к п. 2 алгоритма.

Если задача разрешима в целых числах, то после конечного числа итераций оптимальный целочисленный план будет найден.

Если в процессе решения появится строка с нецелым свободным членом и целыми остальными коэффициентами, то соответствующее уравнение не имеет решения в целых числах. В таком случае и данная задача не имеет целочисленного оптимального плана.

Пример 8.1 Найти максимальное значение целевой функции $f = 2x_1 + 3x_2$ при ограничениях

$$3x_1 + 5x_2 \le 60;$$
$$3x_1 + 4x_2 \le 34;$$
$$x_2 \le 8;$$
$$x_1 \ge 0, x_2 \ge 0,\ x_1, x_2 - \text{целые числа}.$$

Решение. Канонический вид задачи:

$$\max : f(x) = 2x_1 + 3x_2$$
$$3x_1 + 5x_2 + x_3 = 60$$
$$3x_1 + 4x_2 + x_4 = 34$$
$$x_2 + x_5 = 8$$

Начальный базисный план

$$x_3 = 60 - (3x_1 + 5x_2)$$
$$x_4 = 34 - (3x_1 + 4x_2)$$
$$x_5 = 8 - x_2$$
$$БП = (x_3, x_4, x_5), \quad СП = (x_1, x_2)$$
$$X_{нач} = (0, 0, 60, 34, 8)$$

Решим задачу симплекс –методом (см табл.8.1-8.3).

Таблица 8.1 –Первая симплекс -таблица

БП	СЧ	-X1	-X2	-X3	-X4	-X5
X3	60	3	5	1	0	0
X4	34	3	4	0	1	0
X5	8	0	$\boxed{1}$	0	0	1 ⇒
F(x)	0	-2	-3	0	0	0

⇑

$$\max\{|-2|, |-3|\} = 3;$$

$$\min\left\{\frac{60}{5}, \frac{34}{4}, \frac{8}{1}\right\} = 8$$

Таблица 8.2 –Вторая симплекс -таблица

БП	СЧ	-X1	-X2	-X3	-X4	-X5
X3	20	3	0	1	0	-5
X4	2	$\boxed{3}$	0	0	1	-4 ⇒
X2	8	0	1	0	0	1
F(x)	24	-2	0	0	0	3
		⇑				

$$\max\{|-2|\} = 2$$

$$\min\left\{\frac{20}{3}, \frac{2}{3}\right\} = \frac{2}{3}$$

Таблица 8.3 –Третья симплекс -таблица

БП	СЧ	-X1	-X2	-X3	-X4	-X5
X3	18	0	0	1	-1	-1
X1	2/3	1	0	0	1/3	-4/3
X2	8	0	1	0	0	1
F(x)	$25\frac{1}{3}$	0	0	0	2	1/3

Решая задачу симплекс- методом, получаем оптимальный план (табл.3): $\boldsymbol{x}^*$=(2/3; 8; 18; 0; 0) со значением целевой функции $f_{\max} = 25\frac{1}{3}$. Однако полученный план не удовлетворяет условию целочисленности. По строке переменной x_1, получившей нецелочисленное значение в оптимальном плане, составляем дополнительное ограничение:

$$\left\{\frac{1}{3}\right\}x_4 + \left\{-\frac{4}{3}\right\}x_5 \geq \left\{\frac{2}{3}\right\} \quad \text{или } \frac{1}{3}x_4 + \frac{2}{3}x_5 \geq \frac{2}{3}.$$

Вводя дополнительную целочисленную переменную $x_6 \geq 0$, получаем

$$\frac{1}{3}x_4 + \frac{2}{3}x_5 - x_6 = \frac{2}{3}$$

$$0 = \frac{2}{3} - \left(\frac{1}{3}x_4 + \frac{2}{3}x_5 - x_6\right)$$

Составляем симплекс-таблицу расширенной задачи (табл. 8.4).

Таблица 8.4 – Расширенная симплекс таблица

БП	СЧ	-X1	-X2	-X3	-X4	-X5	-X6
X3	18	0	0	1	-1	-1	0
X1	2/3	1	0	0	1/3	-4/3	0
X2	8	0	1	0	0	1	0
0	2/3	0	0	0	1/3	$\boxed{2/3}$	-1 ⇒
F(x)	$25\frac{1}{3}$	0	0	0	2	1/3	0

$$\min\left\{\frac{\frac{2}{3}}{\frac{2}{3}}, \frac{8}{1}\right\} = 1$$

Таблица 8.5 – Вторая расширенная симплекс - таблица

БП	СЧ	-X1	-X2	-X3	-X4	-X5	X6
X3	19	0	0	1	-1/2	0	-2/3
X1	2	1	0	0	1	0	-2
X2	7	0	1	0	-1/2	0	3/2
X5	1	0	0	0	2	$\boxed{1}$	-3/2
F(x)	25	0	0	0	1/2	0	1/2

Повторив процесс решения симплексным методом применительно к расширенной задаче, получим оптимальный целочисленный план (табл. 5) x^* = (2; 7; 19; 0; 1; 0), f_{max}=25.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Дайте постановку задачи целочисленного программирования.
2. Приведите классификацию математических моделей дискретного программирования.
3. В чем суть метода отсечений?
4. Что такое правильное отсечение?
5. В чем суть алгоритма Р. Гомори?

ТЕМА 8. ИСПОЛЬЗОВАНИЕ МЕТОДА ВЕТВЕЙ И ГРАНИЦ ПРИ РЕШЕНИИ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

Лекция 9. Суть метода ветвей и границ. Задача целочисленного (частично целочисленного) программирования. Задача о коммивояжере. Решение задачи коммивояжера методом ветвей и границ (алгоритм Литтла). Решение задачи коммивояжера на максимум Задача о коммивояжере с заданным началом и минимальным путем. Решение задачи коммивояжера методом «ближайшего соседа».

Сущность метода ветвей и границ

Метод ветвей и границ – один из наиболее эффективных решений различных задач комбинаторного типа.

Пусть задана конечная область, в которой требуется найти набор переменных $X \in \Omega$, доставляющих максимум целевой функции $f(X)$. Во многих случаях можно определить верхнюю границу $f(X)$ на множестве планов Ω. Обозначим ее через $\varphi(\Omega)$, *т.е.* $f(X) \le \varphi(\Omega)$. Разобъем Ω на несколько непересекающихся подмножеств Ω_1, Ω_2, …,$\Omega_к$, т.е. разветвим множество Ω. На каждом из подмножеств оценим сверху функцию $f(X)$, т.е. определим числа $\varphi(\Omega_1)$, $\varphi(\Omega_2)$, …,$\varphi(\Omega_k)$. Если удастся найти некоторый набор переменных $X^0 \in \Omega_p$, для которого $f(X^0) = \varphi(\Omega_p) \ge \varphi(\Omega_i)$, $i = \overline{1,k}$, то X^0 – оптимальное решение исходной задачи. Если такой план не обнаружен, то продолжим процесс разбиения, начиная с подмножества с самой высокой оценкой и т.д. Так как множество всех допустимых планов предполагается конечным, то после конечного числа шагов будет найден оптимальный план. В случае задачи линейного целочисленного программирования верхнюю границу находят решением соответствующей задачи с отброшенным условием целочисленности. Ветвление производится последовательным включением в условия дополнительных ограничений.

Задача целочисленного (частично целочисленного) программирования. Задача о коммивояжере

Коммивояжер должен посетить каждый из *n* городов только один раз и вернуться в исходный пункт. Его маршрут должен минимизировать суммарную длину пройденного пути.

Для некоторых пар (*i, j*) непосредственный переход от *i* к *j* может быть запрещен. В этом случае элемент матрицы *C* полагается равным бесконечности. В других случаях требуют, чтобы определенная дуга обязательно входила в маршрут.

На каждом шаге описываемого алгоритма задача включает *n* городов, причем из *n* шагов маршрута *k* шагов могут быть уже установлены и нужно выбрать оптимальным образом из оставшихся *n-k* шагов.

Формализуем задачу. Введем неизвестные величины:

$$x_{ij} = \begin{cases} 1, & \text{если коммивояжер из города } i \text{ переезжает в город } j; \\ 0 & \text{в противном случае.} \end{cases}$$

Пусть $C = \|c_{ij}\|$ матрица расстояний между городами.

Тогда $F = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$ (9.1)

есть длина маршрута для выбранного плана переездов. На матрицу $X = \|x_{ij}\|$ неизвестных должны быть наложены ограничения:

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = \overline{1, n}, \qquad (9.2)$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = \overline{1, n}, \qquad (10.3)$$

Система ограничений (9.2) обеспечивает построение маршрута, при котором коммивояжер въезжает в каждый город только раз, а система ограничений (9.3) – маршрута, когда он выезжает из каждого города только раз. Эти ограничения недостаточны, так как среди допускаемых ими решений имеются маршруты, не образующие полный цикл, включающий все города.

Устранение подциклов достигается при добавлении системы ограничений:

$$u_i - u_j + nx_{ij} \le n-1, \quad i, j = \overline{1, n}, \quad i \ne j \quad , \qquad (9.4)$$

где переменные *u* могут принимать произвольные вещественные значения.

Задача о коммивояжере в математической постановке эквивалентна задаче упорядочения конечного числа работ на машине при учете потерь от переналадок.

Для задачи упорядочения каждая работа соответствует городу, а длительность настройки – расстоянию между городами. Матрица *C* может быть и несимметричной, так как длительность настройки при переходе от работы i к работе j, как правило, отлична от длительности настройки при переходе от j к i.

Решение задачи коммивояжера методом ветвей и границ (алгоритм Литтла)

Алгоритм Литтла для решения задачи о коммивояжере можно сформулировать в виде следующих правил.

1. Находим в каждой строке матрицы $C = \left\| c_{ij} \right\|$ минимальный элемент $u_i = \min\limits_j c_{ij}$ и вычитаем его из всех элементов соответствующей строки. Получим матрицу, приведенную по строкам, с элементами $c_{ij}' = c_{ij} - \min\limits_j c_{ij}$.
2. Если в матрице C', приведенной по строкам, окажутся столбцы, не содержащие нуля, то ее приводим по столбцам. Для этого в каждом столбце матрицы C' выбираем минимальный элемент v_j, j=1, …, n и вычитаем его из всех элементов соответствующего столбца. Получим матрицу $C'' = \left\| c_{ij} - \min\limits_j c_{ij} - \min\limits_i c_{ij}' \right\|$, каждая строка и каждый столбец которой содержит хотя бы один нуль. Такая матрица называется приведенной по строкам и столбцам.
3. Суммируем элементы u_i и v_j, получим константу приведения

$\gamma = \sum\limits_{i=1}^{n} u_i + \sum\limits_{j=1}^{n} v_j$, которая будет нижней границей множества всех допустимых гамильтоновых контуров, т.е. $\gamma = \varphi(\Omega^0) \le f(X)$.

Примечание. Контур – это конечный путь, у которого начальная вершина совпадает с конечной. При этом контур называется элементарным (гамильтоновым), если все его вершины различны, за исключением начальной и конечной, которые совпадают).

4. Находим степени нулей для приведенной по строкам и столбцам матрицы. Для этого мысленно нули в матрице заменяем на ∞ и находим сумму минимальных элементов строки и столбца, соответствующих этому нулю. Записываем ее в правом верхнем углу клетки:

$$\theta_{ij} = \min_{j' \neq j} c_{ij'} + \min_{i' \neq i} c_{i'j}$$

5. Выбираем дугу (i_0, j_0), для которой степень нулевого элемента достигает максимального значения

$$\theta_{i_0 j_0} = \max_{c''_{ij}} \theta_{ij} .$$

6. Разбиваем множество всех гамильтоновых контуров Ω^0 на два подмножества $\Omega^1_{i_0 j_0}$ *и* $\Omega^1_{\overline{i_0 j_0}}$. Подмножество $\Omega^1_{i_0 j_0}$ гамильтоновых контуров содержит дугу (i_0, j_0), а $\Omega^1_{\overline{i_0 j_0}}$ - ее не содержит. Для получения матрицы контуров $\Omega^1_{i_0 j_0}$, включающих дугу (i_0, j_0), вычеркиваем в матрице C'' строку i_0 и столбец j_0. Чтобы не допустить образования негамильтонова контура, заменяем симметричный элемент (j_0, i_0) на ∞.

7. Приводим матрицу гамильтоновых контуров $\Omega^1_{i_0 j_0}$. Пусть $h^1_{i_0 j_0}$ - константа ее приведения. Тогда нижняя граница множества $\Omega^1_{i_0 j_0}$ определяется так: $\varphi(\Omega^1_{i_0 j_0}) = \gamma + h^1_{i_0 j_0}$.
8. Находим множество гамильтоновых контуров $\Omega^1_{\overline{i_0 j_0}}$, не включающих дугу ($i_0$, j_0). Исключение дуги (i_0, j_0) достигается заменой элемента $c''_{i_0 j_0}$ в матрице C'' *на* ∞.
9. Делаем приведение матрицы гамильтоновых контуров $\Omega^1_{\overline{i_0 j_0}}$. Пусть $h^1_{\overline{i_0 j_0}}$ - константа ее приведения. Нижняя граница множества $\Omega^1_{\overline{i_0 j_0}}$ установится так: $\varphi(\Omega^1_{\overline{i_0 j_0}}) = \gamma + h^1_{\overline{i_0 j_0}}$.
10. Сравниваем нижние границы подмножества гамильтоновых контуров $\Omega^1_{i_0 j_0}$ и $\Omega^1_{\overline{i_0 j_0}}$. Если $\varphi(\Omega^1_{i_0 j_0}) < \varphi(\Omega^1_{\overline{i_0 j_0}})$, то дальнейшему ветвлению в первую очередь подлежит подмножество $\Omega^1_{i_0 j_0}$. Если $\varphi(\Omega^1_{i_0 j_0}) > \varphi(\Omega^1_{\overline{i_0 j_0}})$, то разбиению подлежит подмножество $\Omega^1_{\overline{i_0 j_0}}$ и т.д. Процесс разбиения множеств на подмножества сопровождается построением дерева ветвления.

11. Если в результате ветвления получаем размерность сокращенной матрицы 2x2, то определяем полученный ветвлением гамильтонов контур и его длину (рекорд).
12. Сравниваем длину гамильтонова контура с нижними границами оборванных ветвей. Если длина контура не превышает их нижних границ, то задача решена. В противном случае развиваем ветви подмножеств с нижней границей, меньшей полученного контура, до тех пор, пока не получим маршрут с меньшей длиной или не убедимся, что такого не существует.

Пример 9.1 Решить задачу о коммивояжере для матрицы расстояний, представленной в табл. 9.1.

Таблица 9.1 – Матрица расстояний Ω^0

i\j	1	2	3	4	5	u_i
1	∞	30	40	15	6	6
2	10	∞	18	7	9	7
3	20	30	∞	1	10	1
4	25	10	35	∞	5	5
5	9	8	7	6	∞	6

Решение

1. Справа к табл. 9.1 присоединим столбец u_i, в котором записываем минимальные элементы строк. Вычитаем элементы u_i из соответствующих элементов матрицы C, получим матрицу приведенную по строкам (табл. 9.2).

Таблица 10.2 – Матрица приведенная по строкам

i\ j	1	2	3	4	5
1	∞	24	34	9	0
2	3	∞	11	0	2
3	19	29	∞	0	9
4	20	5	30	∞	0
5	3	2	1	0	∞
v_j	3	2	1	0	0

2. Внизу матрицы табл. 9.2 Присоединяем строку v_j, в которой записываем минимальные элементы столбцов. Вычитаем элементы v_j из соответствующих столбцов матрицы C' (табл. 9.3).

Таблица 10.3 –Матрица, приведенная по строкам и столбцам $\widetilde{\Omega}^0$

i\j	1	2	3	4	5
1	∞	22	33	9	0^9
2	0^0	∞	10	0^0	2
3	16	27	∞	0^9	9
4	17	3	29	∞	0^3
5	0^0	0^3	0^{10}	0^0	∞

3. Вычислим константу приведения

$$\gamma = \sum_{i=1}^{5} u_i + \sum_{j=1}^{5} v_j = 25 + 6 = 31.$$

Нижней границей множества всех гамильтоновых контуров будет число $\gamma = 31 = \varphi(\Omega^0) \le f(X)$.

4. Находим степени нулей полностью приведенной матрицы табл. 3. Для этого мысленно заменяем в ней нули на ∞ и устанавливаем сумму минимальных элементов соответствующей строки и столбца. Степени нулей записаны в правых верхних углах клеток, для которых c_{ij}=0.
5. Определяем максимальную степень нуля. Она равна 10 и соответствует клетке (5; 3). Таким образом, претендентом на включение в гамильтонов контур является дуга (5, 3).
6. Разбиваем множество всех гамильтоновых контуров Ω^0 на два Ω^1_{53} и $\Omega^1_{\overline{53}}$. Матрицу Ω^1_{53} с дугой (5, 3) получаем из табл. 9.3 путем вычеркивания строки 5 и столбца 3. Чтобы не допустить образования негамильтонова контура, заменяем элемент (3, 5) на ∞ (табл. 9.4).

Таблица 9.4 – Матрица Ω^1_{53} с дугой (5, 3)

i\j	1	2	4	5	u_i
1	∞	22	9	0	0
2	0	∞	0	2	0
3	16	27	0	∞	0
4	17	3	∞	0	0
v_j	0	3	0	0	0

7. Матрицу гамильтоновых контуров $\Omega^1_{\overline{53}}$ получим из табл. 9.3 путем замены элемента c_{53}'' на ∞ (табл. 9.5).

Таблица 9.5 – Матрица гамильтоновых контуров $\Omega^1_{\overline{53}}$

i\j	1	2	3	4	5	u_i
1	∞	22	33	9	0	
2	0	∞	10	0	2	
3	16	27	∞	0^9	9	
4	17	3	29	∞	0	
5	0	0	∞	0	∞	
v_j			10			

8. Делаем дополнительное приведение матрицы контуров Ω^1_{53}: h^1_{53}=3 (табл. 9.6). Нижняя граница множества Ω^1_{53} равна $\varphi(\Omega^1_{53})$=31+3=34.

9. Находим константу приведения для множества контуров $\Omega^1_{\overline{53}}$: $h^1_{\overline{53}}$=10. Следовательно, нижняя граница $\varphi(\Omega^1_{\overline{53}})$=31+10=41.

Таблица 9.6 – Матрица, приведенная по строкам и столбцам Ω^1_{53}

i\j	1	2	4	5
1	∞	19	9	0^9
2	0^{16}	∞	0^0	2
3	16	24	0^{16}	∞
4	17	0^{19}	∞	0^0

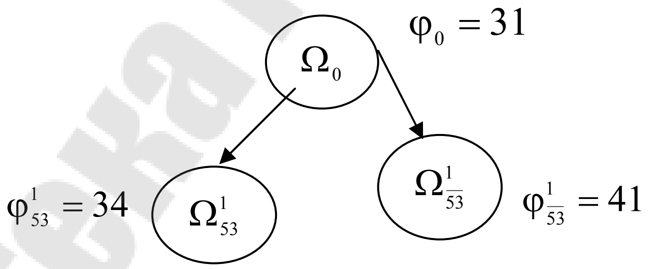


Рисунок 9.1 – Ветвление по дуге (5, 3)

10. Сравниваем нижние границы подмножеств Ω^1_{53} и $\Omega^1_{\overline{53}}$. Так как $\varphi(\Omega^1_{53})$=34< $\varphi(\Omega^1_{\overline{53}})$=41, то дальнейшему ветвлению подвергаем множество Ω^1_{53} (табл. 9.4). На рис. 9.1 представлено ветвление по дуге (5, 3). Переходим к ветвлению подмножества Ω^1_{53}. Его приведенная матрица представлена в табл. 9.6. Узнаем

степени нулей этой матрицы (табл. 9.6). Претендентом на включение в гамильтонов контур будет дуга (4, 2). Разбиваем множество Ω^1_{53} на два подмножества Ω^2_{42} и $\Omega^1_{\overline{42}}$ (табл. 9.7 и 9.8).

Определяем константы приведения этих матриц: $h^2_{42} = 0,\quad h^2_{\overline{42}} = 19$. Следовательно, $\varphi(\Omega^2_{42}) = 34,\ \varphi(\Omega^2_{\overline{42}}) = 53$. На рис. 9.2 представлено ветвление с использованием дуги (4, 2). Так как $\varphi(\Omega^2_{42}) < \varphi(\Omega^2_{\overline{42}})$, то ветвлению подлежит подмножество Ω^2_{42} (табл. 9.7).

Таблица 9.7 – $\widetilde{\Omega}^2_{42}$

i\j	1	4	5	u_i	
1	∞	9	0^{11}	0	
2	0^{18}	∞	2	0	
3	16	0^{25}	∞	0	
v_j	0	0	0	0	

Таблица 10.8 – $\Omega^1_{\overline{42}}$

i\j	1	2	4	5	u_i
1	∞	19	9	0	0
2	0	∞	0	2	0
3	16	24	0	∞	0
4	17	∞	∞	0	0
v_j	0	19	0	0	

Вычислим степени нулей матрицы. Претендентом к включению в гамильтонов контур станет дуга (3, 4). Разбиваем множество Ω^2_{42} на подмножества Ω^3_{34} и $\Omega^3_{\overline{34}}$ (табл. 9.9. и 9.10.). Очевидно, $\varphi(\Omega^3_{34}) = 34,\ \varphi(\Omega^3_{\overline{34}}) = 34 + 25 = 59$. Следовательно, ветвлению нужно подвергнуть подмножество Ω^3_{34}. Но его матрица имеет размерность 2x2. Поэтому в гамильтонов контур следует включить дуги, соответствующие в матрице подмножества Ω^3_{34} нулевым элементам, т.е. дуги (1,5), (2,1).

Таблица 9.9 – $\widetilde{\Omega}^3_{34}$

i\j	1	5	u_i
1	∞	0	0
2	0	2	0
v_j	0	0	

Таблица 9.10 – $\Omega^3_{\overline{34}}$

i\j	1	4	5	u_i
1	∞	9	0	0
2	0	∞	2	0
4	17	∞	∞	16
v_j	0	9	0	

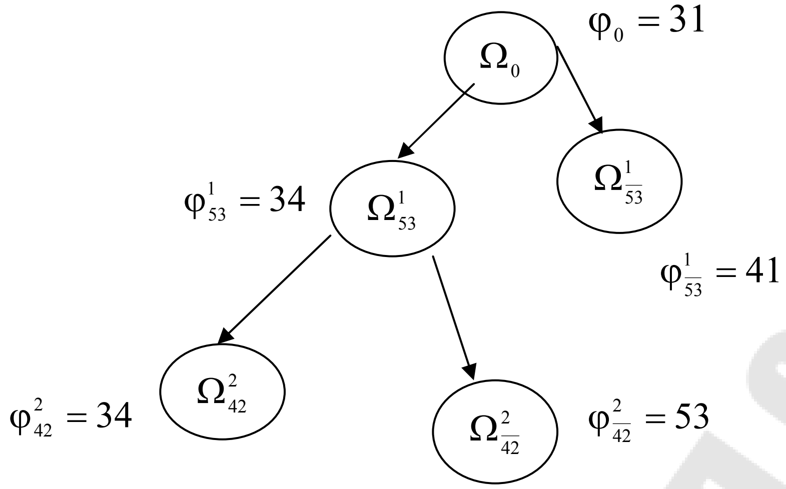


Рисунок 9.2 – Ветвление по дуге (4,2)

На рис. 9.3 представлено дерево ветвлений. Определим полученный гамильтонов контур. В него вошли дуги {(5,3), (4,2), (3,4), (1,5), (2,1)}. Длина контура равна $c_{53}+c_{34}+c_{15}+c_{21}+c_{42}=7+10+1+6+10=34$. Так как границы оборванных ветвей больше длины контура 5-3-4-2-1-5, то этот контур имеет наименьшую длину. На рис. 9.4 представлен замкнутый маршрут минимальной длины.

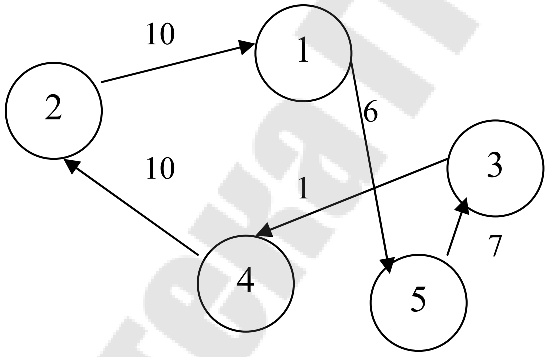


Рисунок 9.4 – Замкнутый маршрут минимальной длины

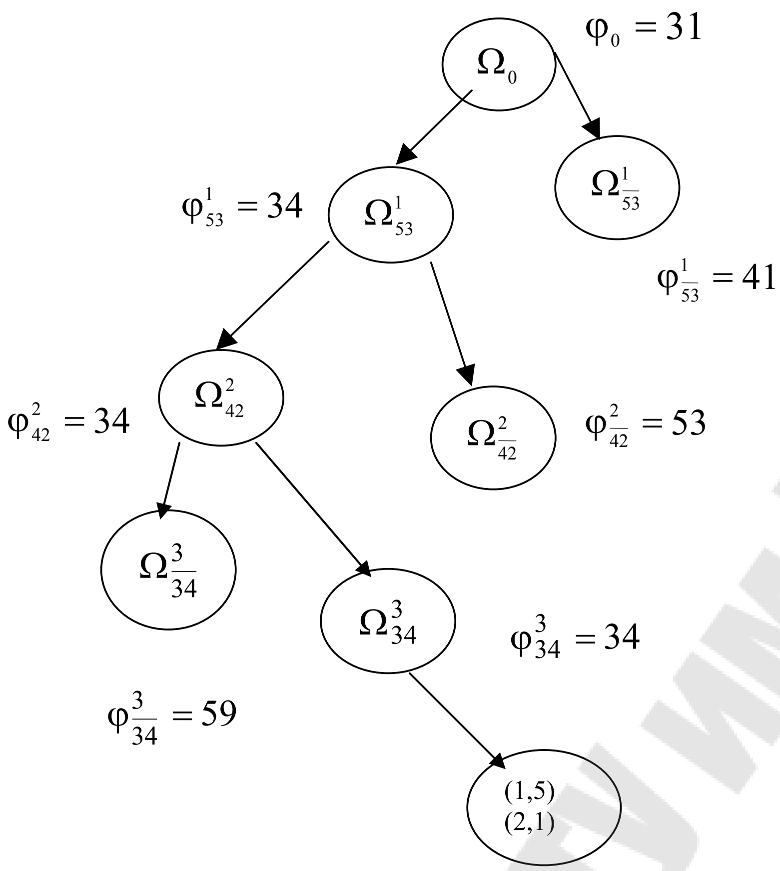


Рисунок 9.3 – Дерево ветвлений

З а м е ч а н и е. Более точно отыскать нижнюю границу можно решением задачи о назначениях. Если при этом получится замкнутый маршрут, то он имеет минимальную длину.

Решение задачи коммивояжера на максимум

В случае решения задачи на max ее легко свести к алгоритму Литтла для решения задачи коммивояжера min. Для этого поступают так. Найдем максимальный элемент в каждом столбце

$$l_j = \max_i c_{ij} .$$

Построим матрицу

$$\widetilde{c}_{ij} = \max_i c_{ij} - c_{ij} .$$

Как и в задаче о назначениях, $\widetilde{F}=\sum_{j=1}^{n} l_j - F$, т.е. $\widetilde{F}$ достигает min для того же плана, для которого F достигает максимума.

Прежде всего, припишем символ ∞ всем элементам матрицы, для которых в графе нет дуги. Затем решим задачу коммивояжера на минимум с матрицей $\widetilde{C}=\left\|\widetilde{c}_{ij}\right\|=\left\|\max_i c_{ij}-c_{ij}\right\|$.

Задача о коммивояжере с заданным началом и минимальным путем

В ряде случаев начальная вершина графа задана. Коммивояжеру нужно обойти все города, отправляясь из наперед заданного города, побывав в каждом из них только один раз, но не обязательно возвращаться обратно. При этом нужно затратить минимальное время. Путь проходит через все вершины графа, но он не замкнут. Если его начало не совпадает с концом, то такой путь называется гамильтоновым. Таким образом, задача сводится к нахождению гамильтонова пути с фиксированным началом.

Для ее решения можно воспользоваться алгоритмом Литтла. На главную диагональ следует поставить символ ∞. В столбце же, соответствующем заданному началу, заменить все элементы, кроме диагонального, нулями. Далее нужно действовать по алгоритму Литтла. Если же нужно найти минимальный гамильтонов путь с нефиксированным началом, то следует ввести нулевую вершину s, соединив ее с остальными вершинами графа дугами с нулевой длиной. Найти алгоритмом Литтла минимальный гамильтонов контур полученного графа. Исключив затем вершину s, получим искомый путь. Если не существует гамильтонова пути, исходящего из заданной вершины, то при вычислениях нижняя граница одного из подмножества окажется равной ∞.

Решение задачи коммивояжера методом «ближайшего соседа»

Существует несколько точных методов решения задачи коммивояжера: ветвей и границ, динамического программирования. Методом ветвей и границ с использованием современных ЭВМ можно решать задачи коммивояжера для $n \le 40$, динамического программирования – для $n \le 17$

В связи с малой эффективностью точных методов получили широкое распространение эвристические методы. В настоящее время существует более ста приближенных методов решения задачи коммивояжера, среди которых прост метод ближайшего соседа. Он реализует требование включать в искомый замкнутый контур вершину, ближайшую к только что найденной. Алгоритм «ближайшего соседа» состоит в последовательном добавлении к начальной вершине ближайшей к ней и т.д. Метод очень прост, однако степень приближения к оптимальному решению зависит от выбора начальной точки. Поэтому алгоритм целесообразно применять, начиная с каждой вершины, и затем выбрать замкнутый контур, наименьшей длины. Отметим, что если ближайший сосед для некоторой вершины уже вошел в контур, то берется следующая по близости вершина и т.д.

На обширном статистическом материале показано, что с увеличением n ошибка решения убывает. Поэтому при $n \leq 40$ можно применять точные методы, при n>40 – приближенные типа ближайшего соседа.

Пример 9.2 – Решить задачу методом «ближайшего соседа»

	1	2	3	4	5
1	∞	30	40	15	6
2	10	∞	18	7	9
3	20	30	∞	1	10
4	25	10	35	∞	5
5	9	8	7	6	∞

Начнем с первой вершины. Ближайшей к ней является пятая вершина: $\min c_{ij}=6$. Процесс нахождения минимального контура целесообразно сопровождать построением дерева ветвления. Если ближайшая вершина уже попала в контур, то блокируем ее и переходим к следующей по степени близости. Ближайшей к пятой вершине является четвертая, а к четвертой – пятая. Но она уже вошла в контур, поэтому блокируем ее и находим следующую по близости вершину – вторую, и т.д. дерево ветвлений, начиная с первой вершины, представлено на рисунке 9.5а.

Для начала с других вершин соответствующие деревья представлены на рис. 9.5б-д.

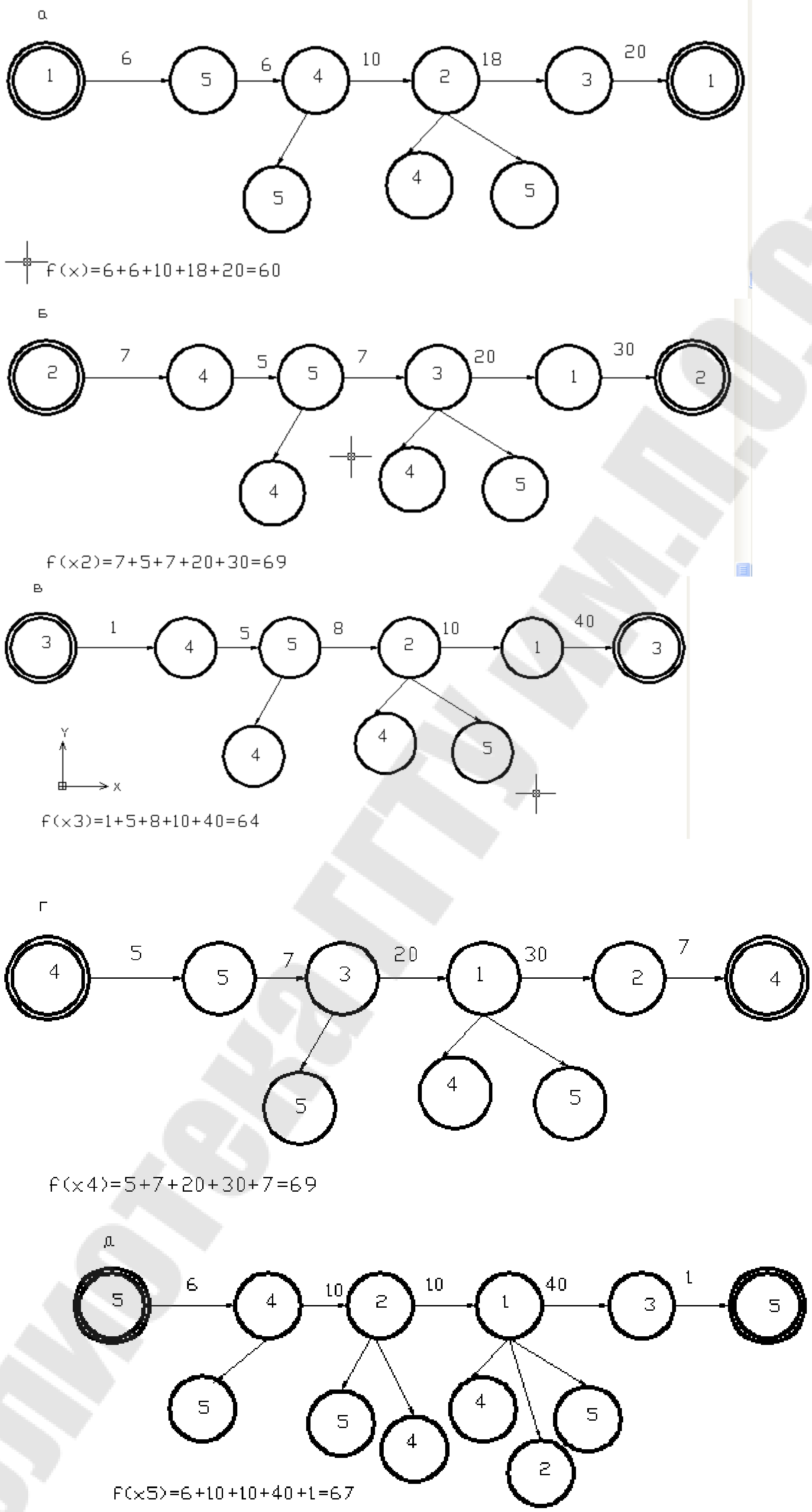


Рисунок 9.5 –Деревья ветвлений «ближайшего соседа»

В результате получим $\min\{f(x), f(x2),f(x3),f(x4),f(x5)\}=\min\{60,69,64,69,67\}=60$.

Таким образом, минимальным контуром, найденным способом «ближайшего соседа» является контур $Г_1=\{1-5-4-2-4-1\}$, $f(Г_1\}=60$. Выше было найдено точное решение методом ветвей и границ: $Г^*=\{1-5-3-4-2-1\}$, $f(Г^*)=34$.

Как видно, ошибка значительна.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. В чем суть метода ветвей и границ?
2. Сформулируйте задачу о коммивояжере.
3. Как решить задачу коммивояжера алгоритм Литтла?
4. Как решить задачу о коммивояжера на максимум?
5. Как решить задачу о коммивояжере с заданным началом и минимальным путем?
6. Как решить задачу о коммивояжера методом «ближайшего соседа»?

ТЕМА 9. ИСПОЛЬЗОВАНИЕ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ ПРИ РЕШЕНИИ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

Лекция 10. Использование динамического программирования при решении технологических задач. Простейшие задачи динамического программирования. Функциональные уравнения Беллмана. Решение задачи коммивояжера методом динамического программирования. Перевозка грузов с минимальными затратами. Оптимальное распределение денежных средств между предприятиями. Оптимальная политика замены оборудования.

Функциональные уравнения Беллмана

Пусть некоторый управляемый процесс находится в первоначальном состоянии $\boldsymbol{S_0}$, которое характеризуется вектором $\boldsymbol{S_0}=(S_{01};\ \ldots;\ S_{0n})$. Множество всех первоначальных состояний определяет $\widetilde{S}_0$. С течением времени процесс меняется и переходит в конечное состояние $\boldsymbol{S_N}=(S_{N1};\ \ldots;\ S_{Nn})$. Множество всех конечных состояний $\widetilde{S}_N$. Процесс перехода из состояния $\boldsymbol{S_0}$ в $\boldsymbol{S_N}$ распадается на N этапов. Причем если процесс находится на i-м этапе в состоянии $\boldsymbol{S_i}$, то состояние его $\boldsymbol{S_{i+1}}$ на следующем (i+1)-м этапе определяется не только вектором состояний $\boldsymbol{S_i}=(S_{i1};\ \ldots;\ S_{in})$, но и решением $\boldsymbol{U_i}$, которое принято на i-м этапе. Исходя из этого, вектор состояний следующего этапа можно представить как $\boldsymbol{S_{i+1}}=W(\boldsymbol{S_i},\ \boldsymbol{U_i})$. Решение же на каждом этапе выбирается из множества $\widetilde{U}_i$ возможных решений и определяет значение целевой функции $f(\boldsymbol{S_i})$. Целевую функцию $f(\boldsymbol{S})$ представим в виде суммы функций $W(\boldsymbol{S_i},\ \boldsymbol{U_i})$, значения которых получаются на каждом этапе при переходе из состояния $\boldsymbol{S_i}$ в состояние $\boldsymbol{S_{i+1}}$:

$$f(\mathbf{S}) = \sum_{i=0}^{N-1} W(\mathbf{S_i}, \mathbf{U_i})\,. \qquad (10.1)$$

Тогда задача динамического программирования состоит в том, чтобы из множества возможных решений $\widetilde{U}$ найти такое решение $\mathbf{U^*}$, которое позволит перевести процесс из первоначального состояния $\boldsymbol{S_0}$ в конечное состояние $\boldsymbol{S_N}$ так, что целевая функция (10.1) будет принимать экстремальное значение при выполнении условий $\boldsymbol{S}_{i+1}=W(\boldsymbol{S}_i,\ \boldsymbol{U}_i)$; $\boldsymbol{S}_i \in \widetilde{S}_i$; $\boldsymbol{U}_i \in \widetilde{U}_i$ (i= 1, …, N-1).

При решении задача динамического программирования разбивается на ряд более простых задач (естественным образом или искусственно). На каждом этапе решается одна из этих задач, причем выбор оптимального решения производиться с учетом будущего, т.е., оптимизируя процесс на каждом этапе, нельзя забывать о всех последующих этапах.

Последний этап (N=1)-й не зависит от будущего, поэтому на данном этапе выбирают решение, позволяющее получить экстремальное значение целевой функции $f_1(S_1)$. Чтобы найти оптимальное решение на N-м этапе, необходимо знать состояние системы на (N-1)-м этапе. Так как состояние процесса на (N-1)-м этапе неизвестно, делают различные предположения о возможных состояниях процесса $\boldsymbol{S}_{N-1,1}$, …, $\boldsymbol{S}_{N-1,m}$ на этом этапе, и для каждого предположения выбирают оптимальное решение $\boldsymbol{U}^*_{N,1}$, …,$\boldsymbol{U}^*_{N,m}$ на N-м этапе. Таким образом, оптимальное решение на N-м этапе найдено. Затем, учитывая уже полученное оптимальное решение на этапе, получают оптимальное решение на (N-1)-м этапе и т.д. В результате приходят к первоначальному состоянию $\boldsymbol{S}_0$ процесса.

Для первого этапа предположений о возможных состояниях процесса не делают, так как, состояние $\boldsymbol{S}_0$ известно. Оптимальное решение первого этапа находят исходя из полученного оптимального решения на втором этапе. Оптимальное решение для всего процесса определяют, просматривая полученные оптимальные решения на всех этапах (от $\boldsymbol{S}_0$ до $\boldsymbol{S}_N$).

Основным методом решения задач динамического программирования является метод функциональных уравнений Беллмана Для каждой конкретной задачи динамического программирования функциональное уравнение имеет свой вид, характеризующийся видом функции W и величинами S, U. Функциональное уравнение задачи динамического программирования можно записать и в общем виде. Так как решение задач динамического программирования проводится от конца к началу, функциональное уравнение для последнего этапа (N=1) имеет вид

$$f_1(\mathbf{S}_{N-1}) = \underset{\mathbf{U}_{N-1}}{extr}\left[W_{N-1}(\mathbf{S}_{N-1};\mathbf{U}_{N-1}) + f_0(\mathbf{S}_N)\right], \qquad (10.2)$$

где $f_0(\boldsymbol{S}_N)$– экстремальное значение целевой функции за нуль этапов, начиная с конечного состояния $\boldsymbol{S}_N \in \widetilde{\mathbf{S}}_N$. Так как за пределами конечного состояния $\boldsymbol{S}_N$ процесс не рассматривается, то $f_0(\boldsymbol{S}_N)$=0.

Функциональное уравнение для N-го этапа запишется так

$$f_N(\mathbf{S}_0) = \underset{\mathbf{U}_0}{extr}\left[W_0(\mathbf{S}_0;\mathbf{U}_0) + f_{N-1}(\mathbf{S}_1)\right], \qquad (10.3)$$

где $f_N(\boldsymbol{S_0})$ является экстремальным значением критерия (целевой функции) за все N этапов процесса, начиная с состояния $\boldsymbol{S_0}$; $f_{N-1}(\boldsymbol{S_1})$ – экстремальное значение критерия за (N-1) этапов, начиная с состояния $\boldsymbol{S_0}$; $W_0(\boldsymbol{S_0}; \boldsymbol{U_0})$ – величина критерия, полученная на N-м этапе, начиная с состояния $\boldsymbol{S_0}$, в результате принятого решения $\boldsymbol{U_0}$.

Задача динамического программирования методом функциональных уравнений Беллмана решается в такой последовательности:

- записывают функциональное уравнение (10.2) для конечного состояния процесса (этапа N=1). Так как $f_0(\boldsymbol{S_N})$=0, то уравнение будет иметь вид $f_1(\mathbf{S}_{N-1}) = \underset{\mathbf{U}_{N-1}}{extr}\left[W_{N-1}(\mathbf{S}_{N-1};\mathbf{U}_{N-1})\right]$.

  Рассматривают набор фиксированных состояний $\boldsymbol{S}_{N-1} \in \widetilde{\mathrm{S}}_{\mathrm{N-1}}$ и решений $\boldsymbol{U}_{N-1} \in \widetilde{\mathrm{U}}_{\mathrm{N-1}}$ и отвечающих им значений W_{N-1}. Среди решений выбирают такое $\mathbf{U}^*_{N-1}$, которое обеспечивает extr$W_{N-1}(\boldsymbol{S}^*_{N-1}; \boldsymbol{U}^*_{N-1})$;
- переходят к следующему, предпоследнему этапу (N=2). Функциональное уравнение для данного этапа записывают в виде $f_2(\mathbf{S}_{N-2}) = \underset{\mathbf{U}_{N-2}}{extr}\left[W_{N-2}(\mathbf{S}_{N-2};\mathbf{U}_{N-2}) + f_1(\mathbf{S}_{N-1})\right]$. Для каждого возможного состояния $\boldsymbol{S}_{N-2}$ находят значение W_{N-2} в зависимости от допустимого решения $\boldsymbol{U}_{N-2}$. Затем сравнивают суммы $W_{N-2}+f_1$ (в которых учтены полученные значения $f_1(\boldsymbol{S}_{N-1})$ последнего этапа) и определяют экстремальную сумму для каждого состояния $\boldsymbol{S}_{N-2}$ и соответствующее условное оптимальное решение $\boldsymbol{U}^*_{N-2}$, т.е. определяют $\boldsymbol{U}^*_{N-2}$, при котором функция $f_2(\boldsymbol{S}^*_{N-2})$ принимает экстремальное значение. Аналогично переходят к следующим этапам (N=3, N=4 и т.д.) до этапа N=N;
- записывают функциональное уравнение (10.3) для первого этапа N=N. На данном этапе предположения о возможных состояниях процесса не делают, так как первоначальное состояние известно $\boldsymbol{S_0}$. Для этого состояния следует определить оптимальное решение с учетом всех полученных условно-оптимальных решений второго этапа;
- проходят весь процесс в прямом направлении от $\boldsymbol{S_0}$ до $\boldsymbol{S_N}$ и определяют оптимальное решение $\boldsymbol{U^*}$ для всего процесса (всей

задачи), которое придает целевой функции $f_N(\boldsymbol{U^*}_N)$ экстремальное значение.

Решение задачи коммивояжера методом динамического программирования

Решение задачи коммивояжера методом динамического программирования было предложено Беллманом, Хелдом и Кэрпом. Опубликованный ими алгоритм является более общим, чем метод ветвей и границ, и применим к более широкому классу задач упорядочивания. Однако трудности, связанные с объемом и временем вычислений при этом алгоритме, появляются для задач менее громоздких, чем при решении методом ветвей и границ.

Перейдем к изложению метода. Предположим, что решается задача коммивояжера для n городов с матрицей расстояний *C*. Без потери общности выберем некоторый город s_0 в качестве исходного. После этого разобьем все множество городов на четыре непересекающиеся подмножества:

1. $\{s_0\}$ - множество, состоящее только из одного исходного города;
2. $\{s_i\}$ - множество, состоящее только из одного города (не исходного);
3. $\{S_k\}$- множество, состоящее из к городов, за исключением $\{s_0\}$и $\{s_i\}$;
4. $\{S_{n-k-2}\}$- множество оставшихся (n-k-2) городов.

Пусть далее, известен оптимальный порядок прохождения городов, начинающийся и заканчивающийся городом $\{s_0\}$. Тогда можно выбрать город $\{s_0\}$ и подмножество $\{S_k\}$, состоящее из к городов, таким образом, что этот оптимальный маршрут начинается в $\{s_0\}$ и проходит через множество $\{S_{n-k-2}\}$, затем через $\{s_i\}$, после чего, пройдя множество городов $\{S_k\}$, оканчивается в $\{s_i\}$.

Теперь рассмотрим только ту часть маршрута, которая связывает $\{s_i\}$ и $\{s_0\}$ с промежуточным посещением городов $\{S_k\}$. Заметим, что для этого участка нам известен наикратчайший путь. Если бы это не так, то, не изменяя части маршрута до города $\{s_i\}$ можно было бы найти лучший путь его завершения и, следовательно, наикратчайший путь целиком. Но это невозможно, поскольку противоречит исходному предположению, что оптимальный маршрут известен.

Пусть $f(s_i;\{S_k\})$- длина наикратчайшего пути от $\{s_i\}$ к $\{s_0\}$ с промежуточным посещением множества $\{S_k\}$. Заметим, что при *к*=0 $f(s_i;\{\theta\}) = c_{i0}$ есть элемент матрицы *С*. Если *k*=*n*-1 и $\{s_i\}$ совпадает с началом движения, то $f(s_0;\{S_{n-1}\})$ является длиной оптимального маршрута исходной задачи. Идея метода динамического программирования состоит в том, чтобы, начиная с *k*=0, шаг за шагом увеличивать *k*. При этом, начав с s_i маршрут проходит в обратном порядке до s_0 и тем самым находится оптимальное решение.

Для рассматриваемой задачи основное функциональное уравнение динамического программирования имеет вид

$$f(s_i;\{S_k\} = \min_{c_j \in \{C_k\}} \left[c_{ij} + f(s_j;\{S_k\} - \{s_j\})\right].$$

Это уравнение показывает, что для того, чтобы найти лучший путь, начинающийся в s_i и завершающийся в s_0 с промежуточным посещением *к* городов, нужно выбрать наикратчайший из *k* возможных путей, начинающихся переходом из s_i в один из *k* и затем проходящий кратчайшим образом в s_0 с промежуточным посещением *(k-1)* других. Любой из этих *k* вариантов, в свою очередь, представляет собой наикратчайший из (*k-1*) возможных путей в соответствии с приведенным ранее уравнением. В конце концов достигается точка, в которой правая часть уравнения, представляет собой просто элемент *С*.

<u>Пример 10.1</u> Для данной матрицы переналадки станков решить задачу о коммивояжере с помощью функциональных уравнений

	1	2	3	4	5
1	∞	30	40	15	6
2	10	∞	18	7	9
3	20	30	∞	1	10
4	25	10	35	∞	5
5	9	8	7	6	∞

<u>Решение.</u>

На нулевом шаге ищется решение для четырех вариантов при *к*=0.

0-й шаг. *к*=0

<u>$f(s_1,\varnothing)=c_{1,5}=6$</u>

$f(s_2,\varnothing)=c_{2,5}=9$

$f(s_3,\varnothing)=c_{3,5}=10$

$f(s_4,\varnothing)=c_{4,5}=5$

На первом шаге при k=1 решения выражаются через известные решения при k=0.

1-й шаг. k=1

$f(s_{1,}\{s_2\})=c_{1,2}+f(s_2,\varnothing)=30+9=39$

$f(s_{1,}\{s_3\})=c_{1,3}+f(s_3,\varnothing)=40+10=50$

$f(s_{1,}\{s_4\})=c_{1,4}+f(s_4,\varnothing)=15+5=20$

$\underline{f(s_{2,}\{s_1\})}=\underline{\mathbf{c_{2,1}+f(s_1,\varnothing)=}}10+6=16$

$f(s_{2,}\{s_3\})=c_{2,3}+f(s_3,\varnothing)=18+10=28$

$f(s_{2,}\{s_4\})=c_{2,4}+f(s_4,\varnothing)=7+5=12$

$f(s_{3,}\{s_1\})=c_{3,1}+f(s_1,\varnothing)=20+6=26$

$f(s_{3,}\{s_2\})=c_{3,2}+f(s_2,\varnothing)=30+9=39$

$f(s_{3,}\{s_4\})=c_{3,4}+f(s_4,\varnothing)=1+5=6$

$f(s_{4,}\{s_1\})=c_{4,1}+f(s_1,\varnothing)=25+6=31$

$f(s_{4,}\{s_3\})=c_{4,3}+f(s_3,\varnothing)=35+10=45$

$f(s_{4,}\{s_2\})=c_{4,2}+f(s_2,\varnothing)=10+9=19$

На втором шаге решения при k=2 выражаются через известные решения при k=1.

2-й шаг. k=2

$f(s_{1,}\{s_2,s_3\})=min[c_{1,2}+f(s_{2,}\{s_3\}),\ c_{1,3}+f(s_{3,}\{s_2\})]=min[30+28,40+39]=58$

$f(s_{1,}\{s_2,s_4\})=min[c_{1,2}+f(s_{2,}\{s_4\}),\ c_{1,4}+f(s_{4,}\{s_2\})]=min[30+12,15+19]=34$

$f(s_{1,}\{s_3,s_4\})=min[c_{1,3}+f(s_{3,}\{s_4\}),\ c_{1,4}+f(s_{4,}\{s_3\})]=min[40+6,15+45]=46$

$f(s_{2,}\{s_1,s_3\})=min[c_{2,1}+f(s_{1,}\{s_3\}),\ c_{2,3}+f(s_{3,}\{s_1\})]=min[10+50,18+26]=44$

$f(s_{2,}\{s_1,s_4\})=min[c_{2,1}+f(s_{1,}\{s_4\}),\ c_{2,4}+f(s_{4,}\{s_1\})]=min[10+20,7+31]=30$

$f(s_{2,}\{s_3,s_4\})=min[c_{2,3}+f(s_{3,}\{s_4\}),\ c_{2,4}+f(s_{4,}\{s_3\})]=min[18+6,7+45]=24$

$f(s_{3,}\{s_1,s_2\})=min[c_{3,1}+f(s_{1,}\{s_2\}),\ c_{3,2}+f(s_{2,}\{s_1\})]=min[20+39,30+16]=46$

$f(s_{3,}\{s_1,s_4\})=min[c_{3,1}+f(s_{1,}\{s_4\}),\ c_{3,4}+f(s_{4,}\{s_1\})]=min[20+20,1+31]=32$

$f(s_{3,}\{s_2,s_4\})=min[c_{3,2}+f(s_{2,}\{s_4\}),\ c_{3,4}+f(s_{4,}\{s_2\})]=min[30+12,1+19]=20$

$\underline{f(s_{4,}\{s_1,s_2\})=}min[c_{4,1}+f(s_{1,}\{s_2\}),\ \underline{\mathbf{c_{4,2}+f(s_{2,}\{s_1\})}}]=min[25+39,10+16]=26$

$f(s_{4,}\{s_1,s_3\})=min[c_{4,1}+f(s_{1,}\{s_3\}),\ c_{4,3}+f(s_{3,}\{s_1\})]=min[25+50,35+26]=61$

$f(s_{4,}\{s_2,s_3\})=min[c_{4,2}+f(s_{2,}\{s_3\}),\ c_{4,3}+f(s_{3,}\{s_2\})]=min[10+28,35+39]=38$

Переходим к третьему шагу, использующему каждое из решений второго шага.

Шаг третьей ($к$=3).

$f(s_{1,}\{s_2,s_{3,}\ s_4\})=min(c_{1,2}+f(s_{2,}\{s_3,s_4\});c_{1,3}+f(s_{3,}\{s_2,s_4\});c_{1,4}+$
$f(s_{4,}\{s_2,s_3\}))=min(30+24;\ 40+20;15+38)=53$

$f(s_{2,}\{s_1,s_{3,}\ s_4\})=min(c_{2,1}+f(s_{1,}\{s_3,s_4\});c_{2,3}+f(s_{3,}\{s_1,s_4\});c_{2,4}+$
$f(s_{4,}\{s_1,s_3\}))=min(10+46;\ 18+32;7+61)=50$
$\underline{f(s_{3,}\{s_1,s_{2,}\ s_4\})=}min(c_{3,1}+f(s_{1,}\{s_2,s_4\});c_{3,2}+f(s_{2,}\{s_1,s_4\});\ \underline{\mathbf{c_{3,4}+}}$
$\underline{\mathbf{f(s_{4,}\{s_1,s_2\}}}))=min(20+34;\ 30+30;1+26)=27$

$f(s_{4,}\{s_1,s_{2,}\ s_3\})=min(c_{4,1}+f(s_{1,}\{s_2,s_3\});c_{4,2}+f(s_{2,}\{s_1,s_3\});\ c_{4,3}+$
$f(s_{3,}\{s_1,s_2\}))=min(25+58;\ 10+44;\ 35+46)=54$

Наконец, на четвертом шаге получаем решение исходной задачи.

Шаг четвертый ($к$=4).
$f(s_{5,}\{s_{1,}\ \ s_2,s_{3,}\ \ s_4\})=min(c_{5,1}+\ f(s_{1,}\{s_2,s_{3,}\ \ s_4\});\ \ c_{5,2}+\ f(s_{2,}\{s_1,s_{3,}\ \ s_4\});\ \ \underline{\mathbf{c_{5,3}+}}$
$\underline{\mathbf{f(s_{3,}\{s_1,s_{2,}\ s_4\}}});\ c_{5,4}+f(s_{4,}\{s_1,s_{2,}\ s_3\}))=min(9+53;50+8;\ 27+7;54+6)=34$

Итак, значение целевой функции $f(x)$=34. Сам маршрут найдем, если будем двигаться в обратном направлении с четвертого шаге к нулевому, проходя по минимальному значению (выделено жирным подчеркиванием). На рис. 10.1 представлен гамильнонов контур данной задачи.

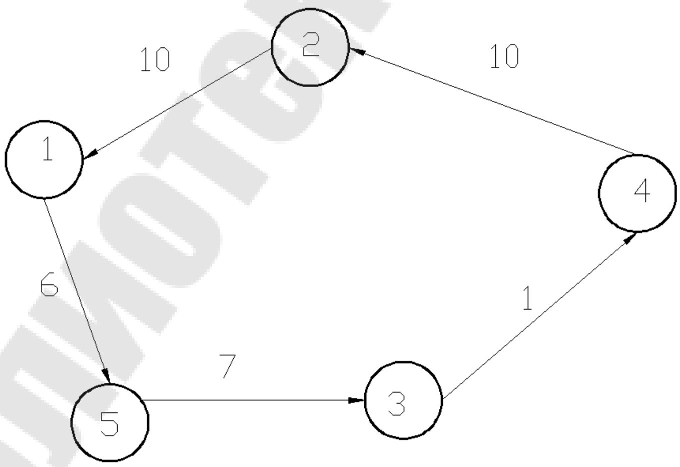


Рисунок 10.1 – Гамильнонов контур

Перевозка грузов с минимальными затратами

Пример 10.2. Требуется перевести груз из пункта 1 в пункт 14. На рис. 10.2 показана сеть дорог и стоимости перевозки единицы груза между отдельными пунктами сети (проставлены у соответствующих ребер). Необходимо определить маршрут доставки груза из пункта 1 в пункт 14, которому соответствуют наименьшие затраты.

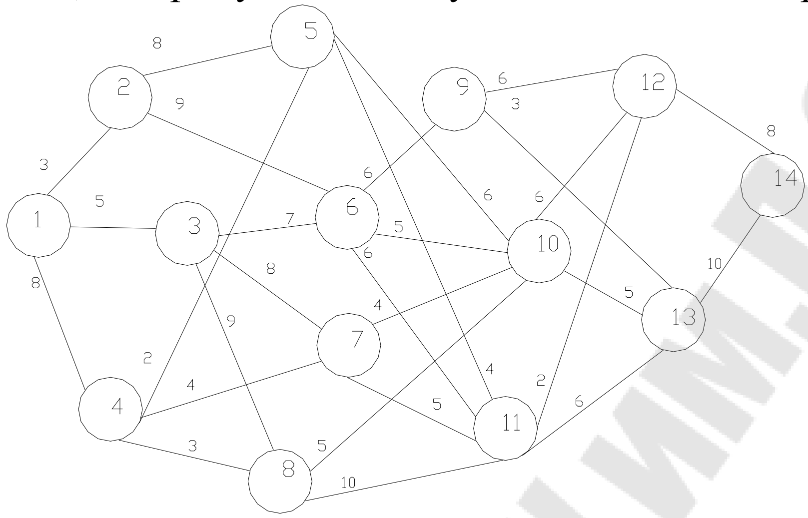


Рисунок 10.2 – Сеть дорог

Решение. Весь процесс доставки груза из пункта 1 в пункт 14 разбиваем на этапы.

- на первом этапе транспорт с грузом из пункта 1 перемещается непосредственно в пункты 2, 3, 4;
- на втором этапе из пунктов 2; 3; 4 – в пункты 5; 6; 7; 8;
- на третьем этапе – из пунктов 5; 6; 7; 8 – в пункты 9; 10; 11;
- на четвертом этапе – из пунктов 9; 10; 11 – в пункты 12; 13;
- на пятом этапе – из пунктов 12; 13 – в пункт 14.

Введем обозначения: N – номер этапа (N =1; …;5); i – пункт, из которого осуществляются перевозки (i =1, …, 13); j – пункт, в который доставляется груз (j = 2, …, 14); c_{ij} – стоимость перевозки груза из пункта i в пункт j; $f_N(i)$ минимальные затраты на перевозку груза из пункта i в конечный, если до конечного пункта осталось N этапов.

Запишем функциональное уравнение для последнего этапа (N=1):

$f_1(i)$= min [$c_{ij}+f_0(14)$]= min(c_{ij}).

В пункт 14 груз может быть доставлен из пункта 12, или из пункта 13, поэтому вычисляем: $f_1(12)= c_{12}+f_0(14)=8+0=8$ и $f_1(13)= c_{13}+f_0(14)= 10+0= 10$.

Для предпоследнего этапа (N=2) функциональное уравнение имеет вид: $f_2(2)= \min\ [c_{ij}+f_1(j)]$. Рассматриваем все возможные исходы третьего этапа (груз оказался в одном из пунктов 9; 10;11). Предположим, что груз оказался в пункте 9. Из этого пункта далее можно следовать либо через пункт 12, либо через пункт 13, и $f_2(9)=\min[c_{9;12}+f_1(12);\ c_{9;13}+f_1(13)]= \min(6+8;\ 3+10)=13$. Значит, условно оптимальный путь из пункта 9 в пункт 14 проходит через пункт 13, при этом минимальные затраты составляют 13. Если груз оказался в пункте 10, то $f_2(10)= \min[c_{10;12}+f_1(12);\ c_{10;13}+f_1(13)]=\min(6+8;\ 5+10)=14$, условно – оптимальный маршрут для пунктов 10-12-14 с минимальными затратами, равными 14 ед. Для пункта 11, получаем $f_2(11)= \min[c_{11;12}+f_1(12);\ c_{11;13}+f_1(13)]= \min(2+8;6+10)=10$, т.е. для этого пункта условно-оптимальным будет маршрут 11-12-14 и затраты составят 10 ед.

Переходим к третьему этапу (N=3). Чтобы получить условно-оптимальное решение на этом этапе, рассматриваем все возможные исходы четвертого этапа (после этого этапа груз может оказаться в одном из пунктов 5; 6; 7; 8). Для пункта 5 получаем $f_3(5)=\min[c_{5j}+f_2(j)]=\min(6+14;\ 4+10)=14$, условно- оптимальный маршрут для этого пункта будет 5-11-12-14. Для пункта 6 записываем $f_3(6)=\min[c_{6j}+f_2(j)]=\min(6+13;\ 5+14;\ 6+10)=16$, и находим условно-оптимальный маршрут 6-11-12-14. Для пункта 7 получаем $f_3(7)=\min[c_{7j}+f_2(j)]=\min(4+14;\ 5+10)=15$, а условно-оптимальный маршрут будет 7-10-12-14. Так как $f_3(8)=\min[c_{8j}+f_2(j)]= \min(5+14;10+10)=19$, то условно-оптимальный маршрут для пункта 8 будет 8-10-12-14, а минимальные затраты составляют 19 ед.

Рассматриваем четвертый этап (N=4). После первого этапа (N=5) груз мог оказаться только в одном из пунктов 2; 3; 4, т.е. $f_4(2)=\min(8+14;\ 9+16)=22$; $f_4(3)=\min(7+16;8+15;9+19)=23$; $f_4(4)=\min(2+14;\ 4+15;\ 3+19)=16$. Это показывает, что для пункта 2 оптимальным будет маршрут 2-5-11-12-14 (затраты 22 ед.), для пункта 4 маршрут –4-5-11-12-14 (затраты 16 ед.), для пункта 3 маршрут – 3-6-10-12-14 (затраты 23 ед.).

Наконец, переходим к первому этапу (N=5). Вычисляем $f_5(1)=\min(3+22;5+23;\ 8+16)=24$. Оптимальным в этом случае будет маршрут 1-4-5-11-12-14, при этом затраты окажутся минимальными (24 ед.).

Найденное решение позволяет установить не только оптимальный маршрут доставки груза из пункта 1 в пункт 14, но и всю структуру оптимальных маршрутов относительно конечного пункта для данной сети дорог (рис. 10.3).

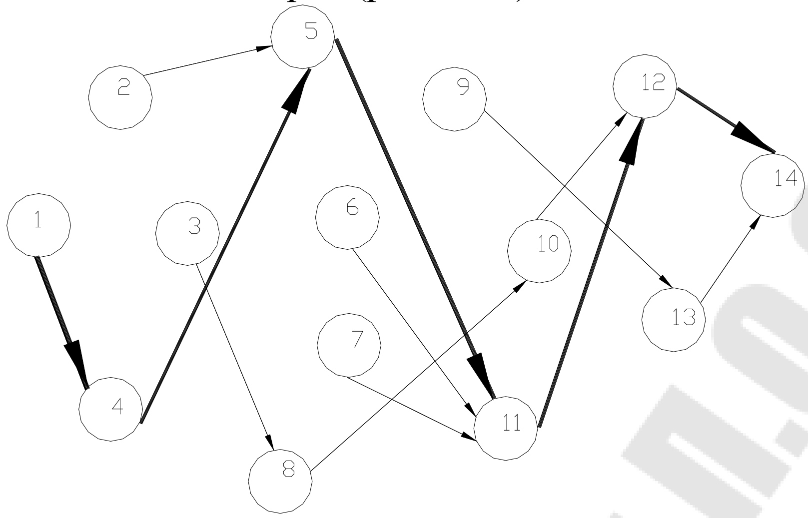


Рисунок 10.3 – Кратчайший маршрут доставки груза

Распределение денежных средств между предприятиями

<u>Пример 10.3</u> Пусть между четырьмя предприятиями распределяются 100 тыс. руб. Значения r_i(x) прироста выпуска продукции на предприятиях в зависимости от выделенной суммы x приведены в табл. 10.1.

Таблица 10.1 – Значения r_i(x) прироста выпуска продукции

Средства *c*, тыс. руб.	Номер предприятия			
	1	2	3	4
	Прирост продукции на предприятиях r_i(x), тыс. руб.			
	$r_1(x)$	$r_2(x)$	$r_3(x)$	$r_4(x)$
20	10	12	11	16
40	31	26	36	37
60	42	36	45	46
80	62	54	60	63
100	76	78	77	80

В соответствии с вычислительной схемой метода динамического программирования условная оптимизация начинается с оптимизации одношагового процесса (последнего шага). В данной задаче этому будет отвечать случай выделения имеющихся средств на долю одного предприятия (пусть им будет, например, предприятия N=1). Для одного предприятия функциональное уравнение имеет вид:

$$f_1(y_{n-1}) = \max_{0 \le x \ge y_{n-1}} [r_1(x)] \qquad (10.4)$$

В соответствии с этим уравнением и в зависимости от начальной суммы c получаем с учетом табл. 10.1 значения $f_1(c)$ (табл. 10.2.).

Таблица 10.2 – Значения $f_1(c)$

$x_1^*(c)$	$f_1(c)$
20	10
40	31
60	42
80	62
100	76

Оптимизируем теперь двухшаговый процесс с учетом оптимизации одношагового. В условиях рассматриваемой задачи это означает, что средства вкладываются в два предприятия: N=1 и N=2. Функциональное уравнение для этого оптимизации:

$$f_2(c) = \max_{0 \le x \ge c} [r_2(x) + f_1(c-x)] \qquad (10.5)$$

На данном этапе надо найти значения функции (10.5) для всех допустимых комбинаций c и x. Для упрощения расчетов значения x будем принимать кратными 20 тыс. руб. Чтобы упорядочить вычисления и придать записям большую наглядность, составляем таблицу. Данному этапу расчетов соответствует табл. 10.3. Для каждого значения начальной суммы c распределяемых средств предусмотрена строка, а для каждого возможного значения распределяемой суммы x – столбец. Некоторые клетки таблицы останутся незаполненными, так как они соответствуют недопустимым сочетаниям c и x.

Таблица 10.3 – Результат двухшагового процесса

c \ x	0	20	40	60	80	100	$f_2(c)$	$x_2^*(c)$
20	0+10	12+0					12	20
40	0+31	12+10	26+0				31	0
60	0+42	12+31	26+10	36+0			43	20
80	0+62	12+42	26+31	36+10	54+0		62	0
100	0+72	12+62	26+42	36+31	54+10	78+0	78	100

В каждую клетку будем вписывать значения суммы $r_2(\text{x})+f_1(c\text{-}x)$. При этом первое слагаемое берем из табл. 10.1, а второе – из табл. 10.2. В двух последних столбцах табл. 10.3 проставлены: максимальный по строке прирост продукции (столбец $f_2(c)$) и соответствующая ей оптимальная сумма средств, выделенная второму предприятию (столбец $x_2^*(c)$).

Расчет значений $f_3(c)$, соответствующий оптимизации трехшагового процесса приведен в таблице 10.4. Здесь использована формула:

$$f_3(c) = \max_{0 \le x \ge c}[r_3(x) + f_2(c - x)].$$

Первые слагаемые табл. 10.4 взяты из табл. 10.1, вторые – из табл. 10.3.

Таблица 10.4 – Расчет значений оптимизации трехшагового процесса

c \ x	0	20	40	60	80	100	F_3(c)	x_3^*(c)
20	0+12	11+0					12	0
40	0+31	11+12	36+0				36	40
60	0+31	11+31	36+12	45+0			48	40
80	0+62	11+43	36+31	45+12	60+0		67	40
100	0+78	11+62	36+43	45+31	60+12	77+0	79	40

Аналогично находятся значения $f_4(c)$. Соответствующая таблица не приводится, рекомендуется ее составить самостоятельно, а результаты сравнить с данными табл. 10.5, полученной из расчетных таблиц, начиная с табл. 10.1.

Таблица 10.5 – Сводная таблица расчетов

c	x_1^*(c)	f_1(c)	x_2^*(c)	f_2(c)	x_3^*(c)	f_3(c)	x_4^*(c)	f_4(c)
0	0	0	0	0	0	0	0	0
20	20	10	20	12	0	12	20	16
40	40	31	0	31	40	36	40	37
60	60	42	20	43	40	48	20	52
80	80	62	0	62	40	67	40	73
100	100	76	100	78	40	79	40	85

Из последней строки табл. 10.5 видно, что наибольший прирост $f_4(c)$ продукции на четырех предприятиях при распределении между ними 100 тыс. руб. (с=100) составляет 85 тыс. руб. ($f_4(100)=85$). При этом четвертому предприятию должно быть выделено 40 тыс. руб. ($x_4^*(100)=40$), а остальным трем - 100-40=60 тыс. руб. Оптимальное распределение этих 60 тыс. руб. (c=60) между тремя предприятиями обеспечит общий прирост продукции на сумму 48 тыс. руб. ($f_3(60)=48$) при условии, что третьему предприятию будет выделено 40 тыс. руб. ($x_3^*(60)=40$), остальным двум 60-40=20 тыс. руб. Оставшиеся 20 тыс. руб. при оптимальном распределении между двумя предприятиями дадут прирост продукции на сумму 12 тыс. руб. ($f_2(20)=12$. При этом второму предприятию надо ассигновать 20 тыс. руб., а на долю первого останется 20-20=0 руб., т.е. $x_1^*=0$. Итак, максимальный прирост выпуска продукции на четырех предприятиях при распределении между ними 100 тыс. руб. составит 85 тыс. руб. и будет получен, если первому предприятию средств не выделять, второму выделить 20 тыс. руб., а третьему и четвертому – по 40 тыс. руб. Это основной результат решения задачи. Кроме него, табл. 4.5. содержит решения аналогичных задач по распределению различных сумм от 20 до 100 тыс. руб. между двумя, тремя и четырьмя предприятиями.

Оптимальная политика замены оборудования

Задача о замене оборудования (обновлении, восстановлении, перестройке) имеет важное значение. Рассмотрим ее в упрощенной постановке. Известно, что оборудование со временем изнашивается, стареет физически и морально. В процессе эксплуатации, как правило, падает его производительность и растут эксплуатационные расходы на текущий ремонт. Со временем возникает необходимость замены оборудования, так как его дальнейшая эксплуатация обходиться дороже, чем ремонт. Отсюда задача о замене может быть сформулирована так. В процессе работы оборудование дает ежегодно прибыль, требует эксплуатационных затрат и имеет остаточную стоимость. Эти характеристики зависят от возраста оборудования. В любом году можно оборудование можно сохранить, продать по остаточной цене и приобрести новое. В случае сохранения оборудования возрастают эксплуатационные расходы и снижается производительность. При замене нужны значительные дополнительные капитальные вложения. Задача состоит в определении оптимальной стратегии замен в плановом периоде, с тем чтобы суммарная прибыль за этот период была максимальной.

Для количественной формулировки задачи введем следующие обозначения: $r(t)$ – стоимость продукции, производимой за год на единице оборудования возраста t лет; $u(t)$ – расходы, связанные с эксплуатацией этого оборудования; s(t) – остаточная стоимость оборудования возраста t лет; p – покупная цена оборудования; T – продолжительность планового периода; t=0,1,2,…,T – номер текущего года.

__Решение.__ Чтобы решить задачу, применим принцип оптимальности Р. Беллмана. Рассмотрим интервалы (годы) планового периода в последовательности от конца к началу. Введем функцию условно-оптимальных значений функции цели Fk(t). Эта функция показывает максимальную прибыль, получаемую от оборудования возраста t лет за последние k лет планового периода. Здесь возраст оборудования рассматривается в направлении естественного хода времени. Например, t=0 соответствует использованию совершенно нового оборудования. Временные же шаги процесса нумеруются в обратном порядке. Например, при k=1 рассматривается последний год планового периода, при k=2 – последние два года и т.д., при k=T – последние T лет, т.е. весь плановый период. Направления изменения t и k показаны на рис. 10.4.

Рисунок 10.4 – Направления изменения t и k

В этой задаче систему составляет оборудование. Ее состояние характеризуется возрастом. Вектор управления это решение в момент t=0,1,2,…,T о сохранении или замене оборудования. Для нахождения

оптимальной политики замен следует проанализировать, согласно принципу оптимальности, процесс от конца к началу. Для этого сделаем предположение о состоянии оборудования на начало последнего года (k=1). Пусть оборудование имеет возраст t лет. В начале T-го года имеются две возможности: 1) сохранить оборудование на T-й год, тогда прибыль за последний год составит $r(t) – u(\mathrm{t})$; 2) продать оборудование по остаточной стоимости и купить новое, тогда прибыль за последний год будет равна $s(t) – \mathrm{p}+\mathrm{r}(0) – u(0)$, где $r(0)$ – стоимость продукции, выпущенной на новом оборудовании за первый год его ввода; $u(0)$ - эксплуатационные расходы в этом году. Здесь целесообразно разворачивать процесс от конца к началу (см. рис. 10.3). Для последнего года (k=1) оптимальной политикой с точки зрения всего процесса будет политика, обеспечивающая максимальную прибыль только за последний год. Учитывая значение прибыли при различном образе действия (замена – сохранение), приходим к выводу, что решение о замене оборудования возраста t лет следует принять в случае, когда прибыль от нового оборудования на последнем периоде больше, чем от старого, т.е. при условии
$s(t) - p + r(0) - u(0) > r(t) - u(t)$.

Если же $s(t) - p + r(0) - u(0) \le r(t) - u(t)$,то старое оборудование целесообразно сохранить.

Итак, для последнего года оптимальная политика и максимальная прибыль $F_1(t)$ находятся из условия

$$F_1(t) = \max_t \begin{cases} r(t) - u(t) & (\textit{сохранение}), \\ s(t) - p + r(0) - u(0) & (\textit{замена}). \end{cases}$$

Пусть k=2, т.е. рассмотрим прибыль за два последних года. Делаем предположение о возможном состоянии t оборудования на начало предпоследнего года. Если в начале этого года принять решение о сохранении оборудования, то к концу года будет получена прибыль $r(t)$-$u(t)$. На начало последнего года оборудование перейдет в состояние t+1, и при оптимальной политике в последнем году оно принесет прибыль, равную $F_1(t+1)$. Таким образом, общая прибыль за два года составит $r(\mathrm{t})$-$u(\mathrm{t})$+$F_1(t+1)$. Если же в начале предпоследнего года будет принято решение о замене оборудования, то прибыль за предпоследний год составит $s(t)$-p+$r(0)$-$u(0)$. Поскольку приобретено новое оборудование, на начало последнего года оно будет в состоянии t=1. Следовательно, общая прибыль за последние два года при оптимальной политике в последнем году составит $s(t)-p+r(0)-u(0)+F_1(1)$.

Условно-оптимальной в последние два года будет политика, доставляющая максимальную прибыль:

$$F_2(t)=\max_t\begin{cases} r(t)-u(t)+F_1(t+1) & (\textit{сохранение}),\\ s(t)-p+r(0)-u(0)+F_1(1) & (\textit{замена}).\end{cases}$$

Аналогично находим выражения для условно-оптимальной прибыли за три последних года, четыре и т.д. Общее функциональное уравнение примет вид

$$F_k(t)=\max_t\begin{cases} r(t)-u(t)+F_{k-1}(t+1) & (\textit{сохранение}),\\ s(t)-p+p(0)-u(0)+F_{k-1}(1) & (\textit{замена}).\end{cases}$$

При k=T получим max$Z$=$F_T(t_0)$, причем

$$F_T(t_0)=\max_t\begin{cases} r(t_0)-u(t_0)+F_{T-1}(t_0+1) & (\textit{сохранение}),\\ s(t_0)-p+r(0)-u(0)+F_{T-1}(1) & (\textit{замена}).\end{cases}$$

Таким образом, разворачивая весь процесс от конца к началу, получаем, что максимальная прибыль за плановый период T составит $F_T(t_0)$. Так как начальное состояние t_0 известно, из выражения для $F_T(t_0)$ находим оптимальное решение в начале первого года, потом вытекающее из него оптимальное решение для второго года и т.д.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Что является основным методом решения задач динамического программирования?
2. В какой последовательности решается задача динамического программирования?
3. Суть решения задачи коммивояжера методом динамического программирования?
4. Как решается задача «Перевозка грузов с минимальными затратами»?
5. Как решается задача «Распределение денежных средств между предприятиями»?

РАЗДЕЛ 4. ОПЕРАТИВНО – КАЛЕНДАРНОЕ ПЛАНИРОВАНИЕ В ТЕХНОЛОГИЧЕСКИХ СИСТЕМАХ НА ОСНОВЕ ТЕОРИИ РАСПИСАНИЙ

ТЕМА 10. ЗАДАЧИ ТЕОРИИ РАСПИСАНИЙ И АЛГОРИТМЫ ИХ РЕШЕНИЯ

Лекция 11. Анализ задач теории расписаний. Классификация задач теории расписаний. Формы представления расписаний. Задачи теории расписаний с одним обслуживающим устройством. Постановка задачи и критерии эффективности. Алгоритмы решения задач с одним станком (обслуживающим прибором).

Анализ задач теории расписаний

Теория расписаний занимается изучением специфических моделей, имеющих место в оперативно-календарном планировании, и разработкой математических методов построения наилучших в смысле заданного критерия эффективности календарных планов. Модели, изучаемые в теории расписаний, могут быть отнесены к классу детерминированных. Это значит, что обстановка в которой приходится вырабатывать наилучшее решение, характеризуется полной определенностью.

Задачи теории расписаний могут быть условно разделены на три группы:

- задачи согласования (задачи сетевого планирования и управления);
- задачи распределения (задачи балансирования сборочной линии, задача о назначениях и др.);
- задачи упорядочения.

Задачи упорядочения носят самый общий характер. Они возникают повсюду, где существует возможность выбора той или иной очередности выполнения работ: при распределении работ на производстве, составлении расписания приземления самолетов, обслуживании клиентов в банках, формировании очередности выполнения программ вычислительным центром и т.д.

Формальная схема процесса упорядочения примерно такова:

- α есть совокупность последствий при выполнении сначала задачи *A,* а затем *B,*

- β есть совокупность последствий ври выполнении сначала задачи *B,* а затем *A,*
- если α предпочтительное, чем β, то выбирается последовательность: «*A*, затем *B*».

В дальнейшем рассматриваются задачи упорядочения при условии, что решены все вопросы, относящиеся к тому, что и каким образом должно быть выполнено. При этом предполагается, что не существует зависимости между характером этих решений и устанавливаемым порядком. Кроме того, предполагается следующее:

1. Подлежащие выполнению работы определены и известны полностью.
2. Однозначно определены устройства, выделяемые для выполнения заданных работ.
3. Задана совокупность всех элементарных действий, связанных с выполнением каждой из работ, и ограничений, налагаемых на порядок их выполнения.

Существует определенное различие между упорядочением и составлением расписания. Упорядочение подразумевает формирование очередности операций, выполняемых одной машиной, в то время как составление расписания означает задание последовательности действий нескольким машинам. Однако такое различие не вносит существенной четкости в постановку задач, поэтому оба термина будут рассматриваться как синонимы.

Основным понятием теории расписаний является понятие *операции.* Операцию можно рассматривать как элементарную задачу, подлежащую выполнению.

Каждая операция характеризуется:

- индексом принадлежности к определенной *работе;*
- индексом принадлежности к определенной *машине;*
- числом, представляющим собой *длительность* операции.

По первому индексу все множество операций разбивается на полную систему непересекающихся подмножеств, называемых *работами.*

Разбиение исходного множества по второму индексу приводит к взаимно непересекающимся подмножествам операций, относящихся к определенным *машинам.*

Для каждой работы задается последовательность составляющих ее операций (определяемая технологическим процессом).

Машиной будем называть устройство, способное выполнить все, что связано с некоторой операцией; *системой обслуживания* - множество всех машин, используемых для выполнения некоторого подмножества операций. Совокупность машин, работ (операций) и дисциплин назначения операций соответствующим машинам назовем *процессом обслуживания.*

С другой стороны, расписание может рассматриваться как задача *упорядочения операций, выполняемых каждой машиной.*

Почти вся теория, разработанная в настоящее время, относится к весьма ограниченному числу моделей *простого процесса обслуживания.* Под последним понимается процесс, для которого существенны следующие ограничения.

1. Каждая машина может быть назначена в любой момент времени. Машины не могут выходить из строя и, следовательно, быть недоступными из-за неисправности или ремонта. Они не могут быть недоступны также из-за пересменок, типичных для производства.

Каждая машина формально представляет собой интервал (0, T), где T есть произвольно большое число.

2. Работы представляют собой строго упорядоченные последовательности операций. Из отдельных операций этих последовательностей не может формироваться никакая новая последовательность.
3. Каждая операция выполняется только одной машиной.
4. Существует только по одной машине каждого типа.
5. Отсутствуют прерывания операций. Это означает, что если некоторая машина начала выполнять операцию, то эта операция должна быть завершена прежде, чем на этой же машине начнется выполнение какой-либо другой операции.
6. Интервалы выполнения последовательных операций одной и той же работы не пересекаются. Одновременно не может выполняться двух операций одной и той же работы.
7. В каждый момент времени машина может выполнять не более одной операции.

Перечисленные ограничения, с одной стороны, упрощают формализацию, а с другой, делают ее более абстрактной. Последнее приводит к тому, что модель становится неадекватной большинству практических случаев, где требуется ослабление одного или нескольких из приведенных ограничений. Тем не менее, такая модель сохраняет в основном структуру большинства практических задач, а при ее исследовании вырабатывается интуиция, полезная в многочисленных приложениях. Во всяком случае в настоящее время описанная формализация применяется в большинстве исследований. Если говорить о более сложных моделях, то для них пока существуют лишь отдельные результаты.

Классификация задач теории расписаний

Задача теории расписаний считается заданной, если определены:

- подлежащие выполнению работы и операции;
- количество и типы машин, выполняющих операции;
- порядок прохождения машин;
- критерии оценки расписаний.

Задачи теории расписаний различаются числом выполняемых в системе работ, характером поступления их в систему и порядком участия отдельных машин в выполнении конкретной работы. В зависимости от характера поступления работ различают два вида задач: *статические* и *динамические.* В статических задачах, если система свободна, в нее одновременно поступает определенное число работ. После этого новые работы не поступают и расписание составляется для вполне определенного и известного заранее числа работ. В динамических задачах выполнение работ происходит непрерывно. Работы поступают в систему в некоторые моменты времени, которые можно предсказать только в статистическом смысле. Поэтому моменты будущих поступлений не определены. Можно ожидать, что упорядочение в динамических и статических задачах потребует совершенно различных методов решения.

Порядок выполнения машинами операций одной работы определяет, является ли система машин *конвейерной.* В конвейерной системе последовательность прохождения машин одинакова для каждой из работ. Согласно принятой терминологии это означает, что существует такая нумерация машин, что для одной и той же работы номер машины, выполняющей операцию x, меньше номера машины, выполняющей операцию y, если x предшествует y. В противоположность этому существуют системы со *случайным порядком выполнения работ машинами.* Здесь отсутствует описанная направленность прохождения машин, и в этом случае любая операция любой работы может выполняться равновероятно любой машиной.

В качестве примера задачи упорядочения рассмотрим одну из задач оперативно-календарного планирования работы производственного участка, обеспечивающего выпуск некоторого количества деталей различных типов. Для каждого типа деталей предполагается известными технологическая последовательность обработки деталей на станках и время обработки каждой детали на каждом из станков. Требуется принять решения, направленные на эффективную организацию работы участка, т.е. определить такой порядок запуска деталей в производство, при котором общее время пребывания их на обработке было бы минимальным.

Покажем, что общее время пребывания деталей на обработке (участке) зависит от порядка запуска деталей в производство.

Пример 11.1. Пусть участок состоит из двух станков №1 и №2. По плану должны быть изготовлены три различные детали, каждая из которых сначала обрабатывается на станке №1, затем на станке №2. Время обработки деталей на каждом из станков задано в табл. 11.1.

Таблица 11.1 – Исходные данные

Станок	Детали		
	Д1	Д2	Д3
№1	3	1	2
№2	2	4	1

Существуют шесть различных способов запуска деталей в производство. Рассмотрим и сравним два из них: (Д2, Д1, Д3) и (Д3, Д1, Д2).

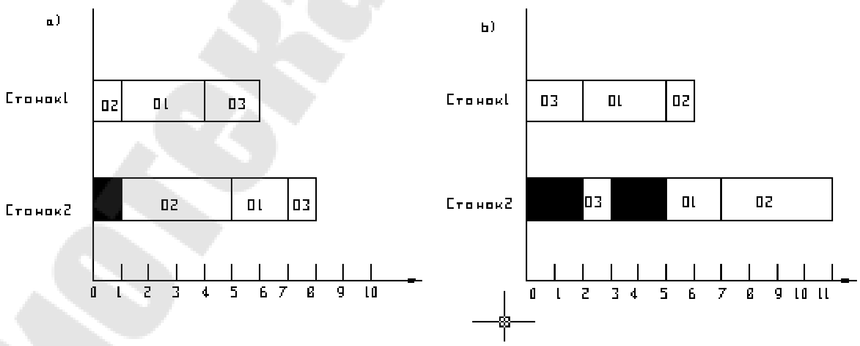


Рисунок 11. 1 – Два способа запуска деталей в производство

Так как на станке №2 деталь может обрабатываться лишь после того как она обработана на станке №1, график обработки деталей показан на рис. 11.1. Из рисунка 11.1 видно, что при первом способе запуска деталей в производство для выполнения всего плана требуется 8 ед. времени, в то время как при втором -11 ед., т.е. первый способ оформления работы является более выгодным.

Формы представления расписаний

Формы представления расписаний могут быть разделены на три группы: графические, аналитические, табличные.

Там, где требуется наглядность часто используются графические способы представления расписаний (графики Ганта, Гант-карты, планировочные графики, учетные графики и т.д.).

График-Ганта представляет упорядочивание занятости различных устройств, механизмов, рабочих мест и т.д. Каждая операция на графике изображается масштабированным отрезком прямой, соответствующим длительности выполнения операции. Моменты начала и конца операций изображаются ⎡ ⎤ соответственно.

Планировочный график – это схематическое изображены процесса, состоящего из ряда работ. При построении планировочного графика каждой работе отводится отдельная строка, в которой указывается названия работы, и проводится линия, соответствующая продолжительности работы в выбранном масштабе времени (табл. .

Таблица 11.2 – Пример планировочного графика

Работа	Недели							
	1	2	3	4	5	6	7	8
Ознакомление с заданием								
Выбор и обоснование метода								
Разработка блок-схемы алгоритма								
Написание программы								
Отладка программы								
Оформление отчета								

Ленточный график служит для изображения выполнения операций, составляющих одну работу. Так как операции во времени не пересекаются, на графике они изображаются последовательно с учетом времени, необходимого для их выполнения (табл. 11.2).

Таблица 11.2 – Пример ленточного графика

Работа	Время, мин						
	5	10	15	20	25	30	35
Токарная обработка							
Сверление отверстий							
Шлифование							
Окраска							
Сушка							

Среди аналитических способов задания расписания можно выделить способ задания с помощью вектор- функции времени. Для рис.11.1 достаточно задать вектор-функцию $S(t)=\left(S_1(t),S_2(t)\right)$, где $S_1(t)$ описывает загрузку станка №1, а $S_2(t)$т- станка №2. Если $S_i(t')=0, i=1,2$, то это означает, что станок *i* в момент времени t' не занят работой и простаивает. Если же $S_i(t')=R, i=1,2$, то это означает, что в момент времени t'на станке *i* выполняется операция *R*. Так, расписание, представленное на рис.11.1а в виде вектор-функции может быть записано следующим образом.

$$S(t)=\begin{pmatrix} S_1(t)=\begin{cases} 2, \text{ при } 0\le t<1, \\ 1, \text{ при } 1\le t<4, \\ 3, \text{ при } 4\le t<6, \\ 0, \text{ в остальных случаях} \end{cases}, \\ S_2(t)=\begin{cases} 2, \text{ при } 1\le t<5, \\ 1, \text{ при } 5\le t<7, \\ 3, \text{ при } 7\le t<8, \\ 0, \text{ в остальных случаях.} \end{cases} \end{pmatrix}$$

Для выработки оптимальных решений в задачах теории расписаний широко применяются различные методы исследования операций комбинаторные, методы математического программирования: методы динамического и дискретного программирования и т.д.

Задачи теории расписаний с одним обслуживающим устройством. Постановка задачи и критерии эффективности

Дадим постановку задачи с одним обслуживающим устройством в терминах «машина-работа (операция)». Пусть для выполнения на одной машине одновременно поступает множество $N=\{1,\ 2, ..., n\}$ работ. Также предположим, что продолжительность выполнения каждой работы на машине известно, и обозначим ее через $t_i,\ i \in N$. Задача построения расписания выполнения множества работ одной машиной состоит в том, чтобы определить такой порядок выполнения работ, при котором некоторый заданный критерий эффективности будет принимать оптимальное решение.

Введем следующие обозначения, необходимые в дальнейшем:

$\underline{t_i}$ - время начала выполнения работы $i \in N$;

$\overline{t}_i$ - время окончания выполнения работы $i \in N$;

d_i- директивное время, в течение которого должно быть завершено выполнение работы $i \in N$;

α_i- штраф за ожидание работы i в единицу времени (или некоторая стоимость, связанная с выполняемой работы) до момента начала ее обработки;

β_i - стоимость выполнения работы $i \in N$.

Время начала и окончания выполнения работ связано следующей зависимостью: $\overline{t_i} = \underline{t_i} + t_i$.

Предполагается, что перерывы в выполнении работ не допускаются. Это значит, что если машина приступила к выполнению некоторой работы, то она продолжает выполнять ее до тех пор, пока не закончит. Время T, необходимое для выполнения всех работ множества N, не зависит от порядка выполнения работ и равно сумме времен выполнения всех работ $T = \sum\limits_{i \in N} t_i$.

Можно записать, что задержка z_i (превышение директивного срока пребывания в системе) работы i составляет

$z_i = \max(0, \overline{t_i} - d_i), \quad i \in N$.

Таким образом, критерий, позволяющий вычислять максимальный штраф, связанный с опозданием в выполнении работ, имеет вид: $\Phi_1 = \max\limits_{i \in N}(\alpha_i \cdot z_i)$.

Часто встречается критерий $\Phi_2 = \sum\limits_{i \in N} \alpha_i \cdot \underline{t_i}$, представляющий сумму штрафов, связанных с ожиданием работ в системе, которую необходимо минимизировать.

Рассмотрим задачу минимизации суммы связанных средств на производственном участке. Стоимость, связанная с выполняемой работой, после ее выполнения будет равна $\gamma_i = \alpha_i + \beta_i$. Время, в течение которого выполненная работа i будет находиться на участке, ожидая окончания всех работ, составит $T-\overline{t_i}$

Можно записать критерий $\Phi_3=\sum_{i\in N}\overline{\gamma_i}\cdot \underline{t_i}$, который можно использовать для минимизации суммы связанных средств, где $\overline{\gamma_i}=-\gamma_i$

Алгоритмы решения задач с одним обслуживающим прибором

Рассмотрим задачу построения оптимального расписания по критерию Φ_2. Для него может быть сформулировано следующее правило.

1. Для всех работ вычислить отношения $\dfrac{t_i}{\alpha_i}$.
2. Упорядочить работы по возрастанию этого отношения.

Учитывая, что $\overline{\gamma_i}=-\gamma_i$, получим правило для построения расписания, минимизирующего сумму связанных средств по критерию Φ_3:

1. Для всех работ вычислить отношения $\dfrac{t_i}{\gamma_i}$.
2. Упорядочить работы по убыванию этого отношения.

Приведем алгоритм, построения оптимального расписания по критерию Φ_1.

Предварительный шаг. Вычисляем время окончания всех работ T. Переходим к шагу1.

Шаг1. Среди всех неупорядоченных работ находим такую работу l, для которой

$$\alpha_l\cdot z_l=\min_i \alpha_i\cdot z_i,$$

где $z_i=\max(0,T-d_i)$. Здесь минимум по i вычисляется по множеству индексов только неупорядоченных работ. Переходим к шагу 2.

Шаг 2. Работу с номером l выполняем последней среди рассматриваемого множества. Исключаем работу l из рассмотрения. Если множество работ пусто, то задача решена. Если нет, то заменяем T на $T-t_l$ и переходим к Шагу1.

После повторения n раз Шага1 и Шага2 будет построено расписание, оптимальное по критерию Φ_1.

Пример 11.2. Определить оптимальный порядок обработки деталей, минимизирующий сумму штрафов, связанных с пролежанием деталей в ожидании обслуживания для исходных данных (табл. 11.2), приведенных в таблице и построить оптимальное расписание в виде кусочно непрерывной функции *S(t)*.

Таблица 11.2 – Исходные данные для примера 11.2

Характеристики	Номер детали					
	Д1	Д2	Д3	Д4	Д5	Д6
t_i	3	5	4	3	2	6
α_i	2	2	1	3	1	2

Для построения оптимального расписания в соответствии с критерием Φ_2 вычислим отношения $\frac{t_i}{\alpha_i}$ и пронумеруем детали в порядке возрастания отношения.

Результаты вычислений представлены в таблице 11.3 а график функции *S(t)* на рис 11.2.

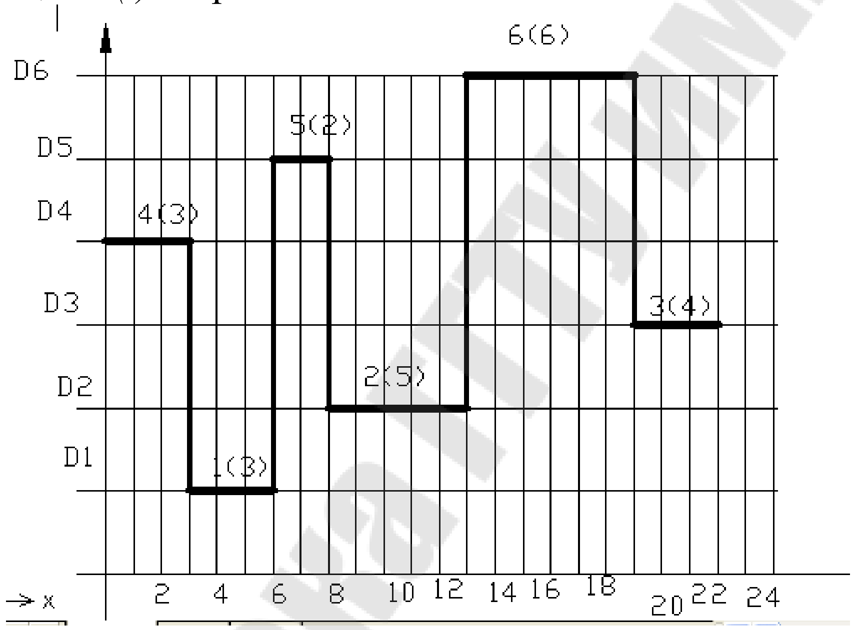


Рисунок 11.2 – Расписание в виде кусочно –непрерывной функции

Таблица 11.3 – Результаты вычислений

Параметры	Номер детали					
	Д1	Д2	Д3	Д4	Д5	Д6
$\frac{t_i}{\alpha_i}$	1.5	2.5	4	1	2	3
Оптимальный порядок	2	4	6	1	3	5

При этом значение критерия эффективности

$$\Phi_2 = \sum_{i \in N} \alpha_i \underline{t_i} = 3 \cdot 0 + 2 \cdot 3 + 1 \cdot 6 + 2 \cdot 8 + 2 \cdot 13 + 1 \cdot 19 = 73 .$$

$\underline{t_i}$ – *начало работы*

Лекция 12. Задача теории расписаний с двумя станками (последовательными обслуживающими устройствами - задача Джонсона для двух станков). Постановка задачи и алгоритм Джонсона. Смешанный вариант задачи Джонсона. Задача теории расписаний с тремя и более последовательными обслуживающими устройствами. Общее решение задачи Джонсона методом ветвей и границ.

Задача теории расписаний с двумя последовательными обслуживающими устройствами (Задача Джонсона). Постановка задачи и алгоритм Джонсона

Дадим формулировку в общем виде.

Имеется множество $N = \{1, 2, ..., n\}$ работ, которые должны быть выполнены на m машинах. Время работы i на машине j обозначим через $t_{i,j}$ $(i = 1 \div n, \quad j = 1 \div m)$, предполагая его заранее известным. Порядок выполнения операций, составляющих работу, может быть как одним и тем же, так и различным для разных работ. Задача построения расписания состоит в указании порядка в котором должны выполняться работы, чтобы суммарное время простоя всех машин было минимальным. При построении любого расписания в том числе и оптимального должны учитываться следующие условия:

1. В любой момент времени на машине не может выполняться больше одной работы.
2. Одна работа в фиксированный момент времени может занимать только одну машину.

Сначала рассмотрим случай, когда число машин равен двум $(M_1 \; и \; M_2)$. Каждая работа состоит из двух операций, которые выполняются сначала на первой машине, затем на второй. Время работы i на первой, а затем на второй машине равно $t_{i,1}$ и $t_{i,2}$ соответственно.

Считая, что порядок выполнения операций на первой и второй машине один и тот же, приведем следующий алгоритм построения оптимального расписания, который называется *алгоритмом Джонсона*.

Предварительный шаг. Записываем матрицу $\|t_{i,j}\|$ $(i = 1 \div n, \quad j = 1, 2)$ времени выполнения операций. Переходим к первому шагу.

Первый шаг. Выбираем в матрице $\|t_{i,j}\|$ минимальный элемент. Если он находится в первой строке (соответствующей первой машине), то данную работу выполняем первой, если во второй строке – то последней. Переходим к шагу два.

Второй шаг . Исключаем из рассмотрения время выполнения операции, относящееся к упорядоченной работе. Если множество элементов матрицы $\|t_{i,j}\|$ пусто, то задача решена. Если нет, переходим к первому шагу.

Таким образом, для построения оптимального расписания шаг1 и шаг2 должны быть повторены *n* раз. Если же случится, что $t_{i,1} = t_{i,2}$, то эта работа может быть упорядочена как по $t_{i,1}$, так и по $t_{i,2}$.

Описанный алгоритм минимизирует суммарное время простоя обеих машин.

Пример 12.1. Имеется пять работ, каждая из которых состоит из двух операций, которые выполняются сначала на первой, затем на второй машине. Время выполнения операций приведено в таблице 12.1.

Таблица 12.1 – Исходные данные

	P1	P2	P3	P4	P5
M1	5	3	4	1	2
M2	2	1	3	2	3

По алгоритму Джонсона оптимальный порядок выполнения работ (P4, P5, P3, P1, P2).

Для определения времени простоя второй машины построим график Ганта.

Из графика Ганта (рис. 12.1) нетрудно заметить, что если машины начинают выполнять работы одновременно, то время простоя составит 5 единиц. Для ликвидации простоя необходимо обеспечить дополнительный уровень работы.

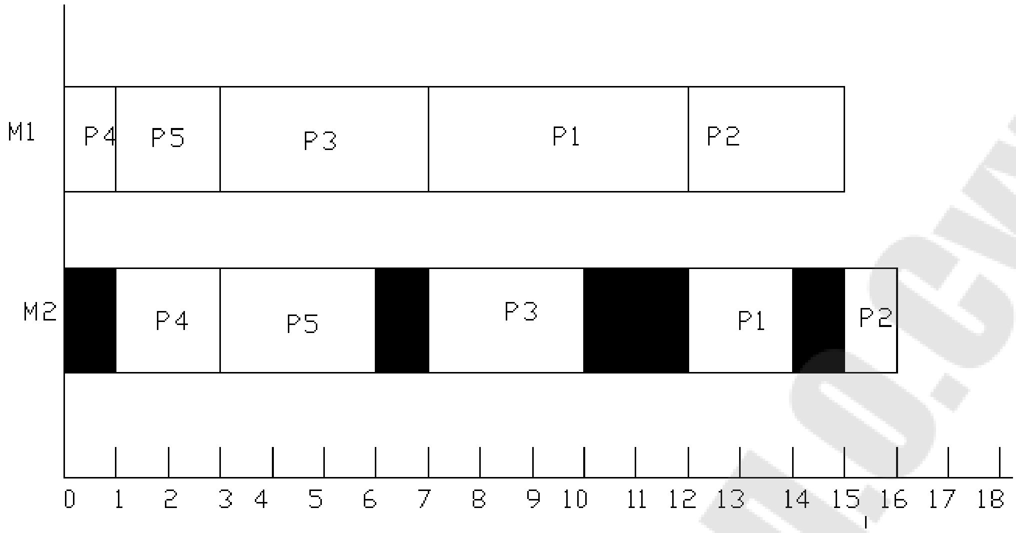


Рисунок 12.1 – График Ганта для двух машин

Смешанный вариант задачи Джонсона

Рассмотрим ситуацию, когда множество работ, подлежащих выполнению может быть разделено на 4 подмножества:

- N_1 - подмножество работ, которые выполняются только на первой машине;
- N_2 - подмножество работ, которые выполняются только на второй машине;
- $N_{1,2}$ - подмножества работ, которые выполняются сначала на первой, затем на второй машине;
- $N_{2,1}$ - подмножества работ, которые выполняются сначала на второй, затем на первой машине.

Для того чтобы оптимально упорядочить множество работ $N = N_1 \cup N_2 \cup N_{1,2} \cup N_{2,1}$ необходимо сначала определить оптимальный порядок выполнения работ, принадлежащих подмножествам $N_{1,2}$ и $N_{2,1}$, в соответствии с алгоритмом Джонсона. Чтобы суммарное время простоя было минимальным (в данном случая может простаивать и первая машина) необходимо:

- на первой машине выполнить сначала работы из подмножества $N_{1,2}$, затем из N_1, а потом – из $N_{2,1}$;
- на второй машине выполнить сначала работы из подмножества $N_{2,1}$, затем из N_2, а потом – из $N_{1,2}$.

Рассмотренный общий случай задачи Джонсона для двух машин часто называют *смешанным вариантом* задачи Джонсона.

Пример 12.2. Условие смешанного варианта задачи Джонсона для двух машин задается следующей таблицей 12.2.

Таблица 12.2 – Исодные данные для смешанной задачи Джонсона

N		$N_{1,2}$				$N_{2,1}$		N_1		N_2	
M\P	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}
M_1	3	1	2	4	1	3	2	5	2	0	0
M_2	4	3	1	2	5	2	1	0	0	2	1

здесь p_j - имя работы, M_i - имя машины.

По алгоритму Джонсона: оптимумы для $N_{1,2}$ и $N_{2,1}$, соответственно: (p_2,p_1,p_4,p_3) и (p_7,p_6,p_5). Оптимальные режимы работ:

- Машина №1: $(p_2,p_1,p_4,p_3,(p_8,p_9), p_7,p_6,p_5)$;
- Машина №2: $(p_7,p_6,p_5,(p_{10},p_{11}), p_2,p_1,p_4,p_3)$.

Записи (p_8,p_9) и (p_{10},p_{11}) означают, что порядок выполнения этих работ произвольный.

Задача теории расписаний с тремя и более последовательными обслуживающими устройствами

В общем виде точное решение задачи Джонсона найдено только для двух машин. Во всех остальных случаях приходится использовать специальные математические методы, которые весьма трудоемкие в вычислительном плане. Рассмотрим случай, когда число машин равно трем, предполагая, что порядок выполнения работ на машинах один и тот же.

Рассмотрим частные случаи, когда решение задачи может быть получено достаточно просто.

Перейдем к случаю трех машин.

Ограничимся здесь далеко не общей ситуацией; итак,

- m=3 - число машин;
- $N = \{1,2,...,n\}$ - множество работ;
- t_{ij} - время выполнения i-ой работы на j-ой машине;
- предполагается, что у всех работ одна и та же последовательность прохождения по машинам.

В этой ситуации справедливы нижеследующие два утверждения, которые приводятся без доказательства.

<u>Утверждение №1</u>. Пусть работы можно пронумеровать так, что окажутся выполненными одновременно следующие неравенства:

$t_{11} \le t_{21} \le \ldots \le t_{n1}$,

$t_{12} \le t_{22} \le \ldots \le t_{n2}$,

$t_{13} \le t_{23} \le \ldots \le t_{n3}$,

$t_{i1} \le t_{i2} \le t_{i3}, \forall i.$

Тогда суммарный простой машин будет минимальным при следующем порядке запуска работ на исполнение: 1—2—...—n.

<u>Утверждение №2</u>. Пусть для матрицы (t_{ij}) выполнено хотя бы одно из двух следующих условий:

$$\min_i t_{i1} \ge \max_i t_{i2}$$

$$\min_i t_{i3} \ge \max_i t_{i2}$$

Построим новую матрицу (τ_{ij}), в которой i=1, 2,..., n, j=1,2 и $\tau_{i1} = t_{i1} + t_{i2}$, $\tau_{i2} = t_{i2} + t_{i3}$, и будем считать ее матрицей времен задачи Джонсона для двух машин в последовательном варианте и множества тех же работ $N = \{1,2,...,n\}$. Тогда суммарное время простоя исходных трех машин при выполнении исходных работ будет минимальным, если эти работы направлять на исполнение в том порядке, который является оптимальным в задаче с матрицей (τ_{ij}).

Общее решение задачи Джонсона методом ветвей и границ

Пусть $N = \{1,2,...,n\}$ - множество работ, которые выполняются на m машинах, причем все работы проходят имеющиеся машины в одной и той же последовательности. Требуется найти такой порядок запуска работ на исполнение, при котором суммарное время простоя всех машин будет минимальным.

Очевидно, что имеется всего $n!$ таких порядков запуска на исполнение, каждый из которых представляет собой перестановку $(i_1 i_2 ... i_n)$ номеров работ $N = \{1,2,...,n\}$. Организацию полного перебора вариантов можно, в частности, провести методом ветвей и границ. Перед тем, как описать применение этого метода к поставленной задаче, заметим, что искомая последовательность работ, минимизирующая суммарное время простоя машин, является минимизирующей и общее время выполнения всех работ. Ниже мы будем искать эту последовательность работ, как обладающую именно этой последней характеристикой: минимальное время выполнения всех работ.

Для большей ясности мы будем далее предполагать, что $m = 3$.

Итак, пусть П - множество всех $n!$ последовательностей $(i_1 i_2 ... i_n)$. Пусть $T : \Pi \to R$ – функция, сопоставляющая каждому расписанию $(i_1 i_2 ... i_n)$ время завершения всех работ. Требуется найти минимум этой функции.

Согласно методу ветвей и границ, нужно построить оценочную функцию на множестве подмножеств множества П и систему разбиений этого множества, приводящую к рекорду.

Построим три вспомогательные функции на множестве Σ подмножеств множества П, которые обозначим, соответственно, $T_i : \sum \to R, i = 1, 2, 3$. Пусть символ $\Pi(i_1,...,i_k)$ обозначает множество всех расписаний, начинающихся с последовательности $(i_1,...,i_k)$. Очевидно, символ $\Pi(i_1,...,i_n)$ обозначает множество, состоящее из единственного элемента $(i_1,...,i_n)$.

Итак, пусть

$$T_1(\Pi) = T_2(\Pi) = T_3(\Pi) = 0;$$

$$T_1(\Pi(i_1)) = t_{i_1 1}; \quad T_2(\Pi(i_1)) = T_1(\Pi(i_1)) + t_{i_1 2}; \quad T_3(\Pi(i_1)) = T_2(\Pi(i_1)) + t_{i_1 3}.$$

Таким образом, $T_k(\Pi(i_1))$ - время окончания исполнения работы i_1 на k-ой машине. Далее определим функции T_k, $k = 1,2,3$, рекуррентно:

$$T_1\big(\Pi(i_1 ... i_k i_{k+1})\big) = T_1\big(\Pi(i_1 ... i_k)\big) + t_{i_{k+1} 1},$$

$$T_2\big(\Pi(i_1 ... i_k i_{k+1})\big) = \max\big\{T_2\big(\Pi(i_1 ... i_k)\big), T_1\big(\Pi(i_1 ... i_k i_{k+1})\big)\big\} + t_{i_{k+1} 2},$$

$$T_3\big(\Pi(i_1 ... i_k i_{k+1})\big) = \max\big\{T_3\big(\Pi(i_1 ... i_k)\big), T_2\big(\Pi(i_1 ... i_k i_{k+1})\big)\big\} + t_{i_{k+1} 3}.$$

Теперь можно построить сразу три оценочных функции для множеств $\Pi(i_1 i_2 ... i_k)$, которые будем обозначать так: Δ_1, Δ_2 и Δ_3. Вот их определение (мы будем обозначать для удобства записи $\Pi(i_1 i_2 ... i_k)$ через x и через $\bar{I}(x)$ те работы из множества $N = \{1,2,...,n\}$, которых нет среди работ $i_1,...,i_k$):

$$\Delta_3(x) = T_3(x) + \sum_{s \in \bar{I}(x)} t_{s3};$$

$$\Delta_2(x) = T_2(x) + \sum_{s\in\bar{I}(x)} t_{s2} + \min_{s\in\bar{I}(x)} t_{s3};$$

$$\Delta_1(x) = T_1(x) + \sum_{s\in\bar{I}(x)} t_{s1} + \min_{s\in\bar{I}(x)} (t_{s2} + t_{s3}).$$

Тот факт, что эти функции – оценочные, можно доказать; аналогично, можно доказать то же самое в отношении функции $\Delta = \max(\Delta_1, \Delta_2, \Delta_3)$; она – оценочная.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. На какие группы задачи теории расписаний могут быть условно разделены?
2. Чем характеризуется каждая операция?
3. Что является машиной?
4. Когда задача теории расписаний считается заданной?
5. Какие используются формы представления расписаний?
6. Дайте постановку задачи с одним обслуживающим устройством.
7. Алгоритм построения оптимального расписания по критерию, позволяющему вычислять максимальный штраф, связанный с опозданием в выполнении работ.
8. Алгоритм построения оптимального расписания по критерию представляющему сумму штрафов, связанных с ожиданием работ в системе, которую необходимо минимизировать.
9. Алгоритм построения оптимального расписания по критерию, который можно использовать для минимизации суммы связанных средств.
10. Дайте постановку задачи теории расписаний с двумя последовательными обслуживающими устройствами.
11. Алгоритм Джонсона.
12. Смешанный вариант задачи Джонсона.
13. Частные случаи решения задач теории расписаний с тремя обслуживающими устройствами.

РАЗДЕЛ 5. МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ НА ОСНОВЕ ТЕОРИИ ИГР

ТЕМА 11. ИСПОЛЬЗОВАНИЕ МАТРИЧНЫХ ИГР ПРИ РЕШЕНИИ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

Лекция 13.1. Некоторые основные понятия теории игр. Матричные игры. Решение матричных игр в чистых стратегиях. Решение матричных игр в смешанных стратегиях. Свойства оптимальных смешанных стратегий. Численные методы решения матричных игр. Связь теории игр с линейным программированием.

Некоторые основные понятия теории игр

В процессе целенаправленной человеческой деятельности возникают ситуации, в которых интересы отдельных лиц (участников, групп, сторон) либо прямо противоположны (антагонистичны), либо не будучи непримиримыми, все же не совпадают. Простейшими и наиболее наглядными примерами таких ситуаций являются спортивные игры, арбитражные споры и т. п. Здесь каждый из участников сознательно стремится добиться наилучшего результата за счет другого участника. Подобного рода ситуации встречаются и в различных сферах производственной деятельности.

Для указанных ситуаций (будем называть их, конфликтными), характерно, что эффективность решений, принимаемых в ходе конфликта каждой из сторон, существенно зависит от действий другой стороны. При этом ни одна из сторон не может полностью контролировать положение, так как и той и другой стороне решения приходится принимать в условиях неопределенности. Так, при определении объема выпуска продукции на одном предприятии нельзя не учитывать размеров выпуска аналогичной продукции на других предприятиях. В реальных условиях нередко возникают ситуации, в которых антагонизм отсутствует, но существуют противоположные тенденции. Так, для нормального функционирования производства необходимо наличие запасов разнообразных ресурсов, но с другой стороны стремление к чрезмерному увеличению этих запасов вызывает дополнительные затраты по их хранению. В приведенных примерах конфликтные ситуации возникают в результате сознательной деятельности людей. Однако на практике встречаются и такие неопределенности, которые порождаются не сознательным противодействием другой стороны, а недостаточной информированностью об условиях проведения планируемой операции.

Раздел математики, изучающий *конфликтные* ситуации, т. е. ситуации, в которых интересы участников (игроков) противоположны или не совпадают, называется *теорией игр.* Теория игр — это математическая теория конфликтных ситуаций, определяющая рекомендации по рациональному образу действий каждого из участников в ходе конфликтной ситуации, т. е. таких действий, которые обеспечивали бы ему наибольший выигрыш (наименьший проигрыш). Игровую схему можно придать многим ситуациям в экономике. Здесь выигрышем может быть эффективность использования дефицитных ресурсов, производственных фондов, величина прибыли, себестоимость и т. д.

Необходимо подчеркнуть, что методы и рекомендации теории игр разрабатываются применительно к таким конфликтным специфическим ситуациям, которые обладают свойством многократной повторяемости. Если конфликтная ситуация реализуется однократно или ограниченное число раз, то рекомендации теории игр теряют смысл. Чтобы проанализировать конфликтную ситуацию с помощью математических методов, ее необходимо упростить, учитывая лишь важнейшие факторы, существенно влияющие на ход конфликта. Отсюда *игра* — это упрощенная модель конфликтной ситуации, отличающаяся от реального конфликта тем, что ведется по определенным правилам. Поэтому можно сказать, что *игра* — совокупность правил, определяющих возможные действия участников игры. Суть игры в том, что каждый из участников принимает такие решения, которые, как он полагает, могут обеспечить ему наилучший результат (исход). *Исход* игры — это значение некоторой функции, называемой *функцией выигрыша (платежной функцией).*

Будем рассматривать только такие игры, в которых выигрыш выражается количественно: стоимостью, баллами и т. д. Величина выигрыша зависит от стратегии, применяемой игроком. *Стратегия*— это совокупность правил, однозначно определяющих последовательность действий игрока в каждой конкретной ситуации, складывающейся в процессе игры. Всякая игра состоит из отдельных партий. *Партией* называют каждый вариант проведения игры определенным образом. В свою очередь в партии игроки совершают определенные ходы. *Ход* — это выбор и реализация игроком одного из возможных вариантов поведения. Ходы бывают *личные* и *случайные.* При *личном ходе* игрок сознательно выбирает и реализирует ту или иную стратегию. Например, в шахматах каждый ход является личным. При *случайном ходе* выбор стратегии производится с использованием механизма случайного выбора, например, с применением таблицы случайных чисел.

Конфликтные ситуации, встречающиеся в практике, порождают различные виды игр. Классифицировать игры можно по разным признакам. Различают, например, игры по количеству игроков. В игре может участвовать любое конечное число игроков. Если при этом игроки объединяются, например, в две группы, преследующие противоположные цели, то имеет место *игра двух «лиц» (парная игра).* В зависимости от количества стратегий в игре они делятся на *конечные* и *бесконечные.* В зависимости от взаимоотношений участников различают игры *бескоалиционные* (участники не имеют права заключать соглашения), или *некооперативные,* и *коалиционные,* или *кооперативные.*

По характеру выигрышей игры делятся на игры с *нулевой суммой* и *ненулевой суммой.* В первых — общий капитал игроков не меняется, а лишь перераспределяется в ходе игры, в связи с чем сумма выигрышей равна нулю. В играх с ненулевой суммой сумма выигрышей участников отлична от нуля. Например, при проведении лотереи часть взноса игроков идет организатору лотереи.

По виду функции выигрыша игры делятся на; *матричные, биматричные, непрерывные, выпуклые, сепарабельные* и др. В матричных играх (при двух участниках) выигрыши первого игрока задаются матрицей, в биматричных выигрыши каждого игрока задаются своей матрицей. Другие типы таких игр различаются видом аналитического выражения функции выигрышей. По количеству ходов игры делятся на *одноходовые* (выигрыш распределяется после одного хода каждого игрока) и *многоходовые* (выигрыш распределяется после нескольких ходов). Многоходовые игры в свою очередь делятся на *позиционные, стохастические, дифференциальные* и др. В зависимости от состояния информации различают игры с полной и неполной информацией.

Матричные игры. Решение матричных игр в чистых стратегиях

Будем рассматривать игры, в которых у каждого из двух игроков A и B конечное число возможных действий — *чистых стратегий.* Допустим, что игрок A располагает m чистыми стратегиями A_1, ..., A_m, а игрок B — n чистыми стратегиями B_1, ...,B_n- Чтобы игра была полностью определенной, необходимо указать правило, сопоставляющее каждой паре чистых стратегий Ai и Bj число a_{ij} — выигрыш игрока A за счет игрока B или проигрыш игрока B. При a_{ij} <0 игрок A платит игроку B сумму | a_{ij} |. Если игра состоит только из личных ходов, то выбор пары чистых стратегий *(Ai, Bj)* единственным образом определяет исход (результат) игры. Если же в игре используются и случайные ходы, то исход, игры обусловлен средним значением выигрыша (математическим ожиданием). Таким образом, мы рассматриваем *парные игры с нулевой суммой,* в которых выигрыш одного игрока равен проигрышу другого.

Если известны значения a_{ij} выигрыша для каждой пары (A_i, Bj) стратегий, то можно составить *матрицу игры* — платежную матрицу (табл. 13.1). *Платежная матрица* — это табличная запись функции выигрыша. Описанные игры называют *матричными* или *прямоугольными.* Отдельная партия в такой игре реализуется следующим образом. Игрок A выбирает одну из строк платежной матрицы (одну из своих чистых стратегий). Не зная результата его выбора, игрок B выбирает один из столбцов (свою чистую стратегию). Элемент матрицы, стоящий на пересечении выбранных строки и столбца, определяет *выигрыш* игрока *A (проигрыш* игрока В).

Целью игроков является выбор наиболее выгодных стратегий, доставляющих игроку A максимальный выигрыш, а игроку B минимальный проигрыш. В теории игр исходят из предположения, что каждый игрок считает своего противника разумным и стремящимся помешать ему достичь наилучшего результата.

Стратегию игрока A называют *оптимальной,* если при ее применении выигрыш игрока A не уменьшается, какими бы стратегиями не пользовался игрок B. Оптимальной для игрока B называют стратегию, при использовании которой проигрыш игрока B не увеличивается, какие бы стратегии ни применял игрок A. С учетом этого игрок A анализирует матрицу выигрышей следующим образом: для каждой своей чистой стратегии А, - он определяет минимальное значение α_i,- выигрыша в зависимости от применяемых игроком *B* чистых стратегий B_j, *т.е.* $\alpha_i = \min a_{ij}$ (i=1, …, m)

Таблица 13.1 – Платежная матрица

Ai	*Bj*				αi
	B1	*B2*	…	*Bn*	
A1	a_{11}	a_{12}	…	a_{1n}	$\alpha 1$
A2	a_{21}	a_{22}		a_{2n}	$\alpha 2$
…	…	…	…	…	…
Am	a_{m1}	a_{m2}	…	a_{mn}	αm
βj	$\beta 1$	$\beta 2$	…	βn	

Затем по минимальным выигрышам α_i он отыскивает такую чистую стратегию A_0, при которой этот минимальный выигрыш будет максимальным, т. е. находит

$$\alpha = \max_i \alpha_i = \max_i \min_j a_{ij} \qquad (13.1)$$

Число α, определяемое по формуле (13.1), называется *нижней чистой ценой игры (максимином).* Оно показывает, какой минимальный выигрыш может получить игрок *A,* применяя свои чистые стратегии при любых действиях игрока *B*. Соответствующая стратегия A_0 игрока *A* называется *максиминной.*

Игрок *B* при наилучшем своем поведении максима*льно уменьшает проигрыш. Поэтому для каждой чистой стратегии* B_j он отыскивает $\beta_j = \max_i a_{ij}$ $(j=1,\cdots n)$ а затем по β_j находит такую свою стратегию B_{j0}, при которой его проигрыш будет минимальным, т. е.

$$\beta_j = \min_i \beta_j = \min_j \max_i a_{ij} \qquad (13.2)$$

Число β, определяемое по формуле (13.2), называется *верхней чистой ценой игры (минимаксом).* Оно показывает, какой максимальный проигрыш вследствие использования чистых стратегий может быть у игрока *B*. Соответствующая чистая стратегия B_{j0} игрока *B* называется *минимаксной.*

Таким образом, используя чистые стратегии, игрок *A* обеспечивает выигрыш не меньше α, а игрок *B* в результате применения своих чистых стратегий может не позволить игроку *A* выиграть больше, чем β. Принцип осторожности, диктующий игрокам выбор максиминной и минимаксной стратегий, называют *принципом минимакса.* Максиминную и минимаксную стратегии игроков для краткости иногда; называют *минимаксными.*

Пример 13.1. Найти максиминимаксную стратегии в игре с матрицей

$$\begin{matrix} 0 & 4 & -1 & 3 \\ 1 & 0 & 2 & 2 \\ 3 & 1 & -2 & -1 \end{matrix}$$

Таблица 13.2 – Поиск максиминимаксной стратегии

	B1	B2	B3	B4	α_i
A1	0	4	-1	3	-1
A2	1	0	2	2	0
A3	3	1	-2	-1	-2
β_j	3	4	2	3	

Решение. В последнем столбце табл. 13.2 выписаны минимальные по строкам элементы α_i. Наибольшим из них будет 0. Итак, максимин $\alpha = \max\limits_i \alpha_i = \max\limits_i \min\limits_j a_{ij}$ =0, а максиминной чистой стратегией игрока A является A_2.. Минимакс, а минимакс-$\beta_j = \min\limits_i \beta_j = \min\limits_j \max\limits_i a_{ij}$ =2, а минимаксной для игрока В является стратегия B_2

Теорема 1. В матричной игре нижняя чистая цена игры не превосходит верхней чистой цены игры, т. е. $\alpha \le \beta$

Если в матричной игре нижняя и верхняя чистые цены игры совпадают, т. е. $\alpha = \beta$ то говорят, что эта игра имеет *седловую точку* в чистых стратегиях и *чистую цену игры* $v = \alpha = \beta$

Обозначим i_0 и j_0 через номера чистых стратегий, при которых имеет место равенство $\alpha = \beta$. Пару чистых стратегий (A_{i0} ; B_{j0}) игроков A и B, при которых достигается это равенство, называют *седловой точкой матричной игры,* а элемент a_{ij} матрицы, стоящий на пересечении i-строки и j-столбца, — *седловым элементом платежной матрицы.*

Решение матричных игр в смешанных стратегиях

Если игра имеет седловую точку, то оптимальными для игроков будут их минимаксные стратегии, а чистой ценой игры — седловой элемент матрицы игры. Если игра седловой точки не имеет, то решение игры затрудняется.

Пример 13.2. Выяснить имеется ли решение в чистых стратегиях. Исходные данные представлены в таблице 13.3

Таблица 13.3 – Платежная матрица игры

	B1	B2	βj
A1	2	9	2
A2	6	3	3
αi	6	9	

Решение. Игрок *А* может выиграть не менее 3 единиц, а игрок *В* может ограничить свой проигрыш (выигрыш игрока *А)* 6 единицами. Область между числами 3 и 6 остается как бы нейтральной, и каждый игрок может попытаться улучшить свой результат за счет этой области. Как это сделать? Если игроки применяют свои минимаксные стратегии *А2* и *В1,* то игрок *А* выигрывает, а игрок *В* проигрывает 6. Это, конечно, устраивает игрока *А,* но не выгодно игроку *В.* Поэтому если игрок *В* заметит, что игрок *А* предпочитает стратегию *А2,* то он может использовать стратегию *В2* и уменьшить выигрыш игрока *А* до 3. Но если игрок *А* раскроет замысел игрока *В* и применит стратегию *А1* то он увеличит свой выигрыш до 9. В свою очередь, узнав об этом, игрок *В* может выбрать стратегию *В1* и понизит выигрыш игрока *А* до 2.

Из этих рассуждений следует, что игрокам надо так выбирать свои стратегии в очередной партии, чтобы противник не догадался о них. Этого можно добиться, если выбирать стратегию случайным образом, используя механизм; случайного выбора. Анализ игры без седловой точки показывает, что игрок А выиграет больше максимина α, получаемого им при максиминной стратегии, если в ходе игры будет пользоваться не одной, а несколькими чистыми стратегиями, т. е. применять смешанную стратегию. Аналогично игрок *В* проиграет меньше минимакса β, выплачиваемого им игроку *А* при минимаксной стратегии, если он будет использовать свою смешанную стратегию.

Таблица 13.3 – Общей случай матричной игры

	B1	…	Bn	p_i
A1	a_{11}		a_{1n}	p_1
…				
Am	a_{m1}		a_{mn}	p_m
q_j	q_1		q_n	

Обратимся к общему случаю матричной игры (табл. 13.3). Обозначим через $p_1, ..., p_m$ вероятности, с которыми игрок A использует чистые стратегии $A1.....A_m$. Ясно, что

$$p_i \geq 0, \quad \sum_{i=1}^{m} p_i = 1 \qquad (13.3)$$

Упорядоченное множество $p=(p_1; ...; p_m)$, элементы которого удовлетворяют условиям (13.3), полностью определяет характер игры игрока A и называется его *смешанной стратегией.* Таким образом, смешанной стратегией игрока A является полный набор вероятностей применения его чистых стратегий. Ясно, что механизм случайного выбора чистых стратегий, которым пользуется игрок A, обеспечивает ему множество смешанных стратегий. Любая его чистая стратегия Ai может рассматриваться как частный случай смешанной стратегии, *i-я* компонента которой равна 1, а остальные равны 0, т. е. p= (0; ...; 1; ...; 0).

Аналогично упорядоченное множество $q = (q_1, ...; q_n)$, элементы которого удовлетворяют соотношениями ($q \geq 0 \quad \sum_{j=1}^{n} q_j = 1$ j=1, ..., n) является *смешанной стратегией игрока B*. Игрок B также располагает множеством смешанных стратегий.

Итак, пусть игроки A и B применяют смешанные стратегии p и q. Это означает, что игрок A использует стратегию Ai с вероятностью p_i а игрок B — стратегию Bj с вероятностью q_j. Поскольку игроки выбирают свои чистые стратегии случайно и независимо друг от друга, то вероятность выбора комбинации стратегий (Ai; Bj) будет равна произведению вероятностей p_i и q_j, т. е. $p_i q_j$. При использовании смешанных стратегий игра приобретает случайный характер, случайной становится и величина выигрыша игрока A (проигрыша игрока B). Поэтому теперь можно вести речь лишь о средней величине (математическом ожидании) выигрыша. Ясно, что эта величина является функцией от смешанных стратегий p, q и определяется по формуле

$$f(p;q) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} p_i q_j$$

Функция $f(p; q)$ называется *платежной функцией* игры с матрицей ||aij||.

Игрок A, изменяя свои смешанные стратегии p, стремится максимизировать средний выигрыш $f(p; q)$, а игрок B, изменяя свои смешанные стратегии q, старается сделать этот его выигрыш как можно меньше. Учитывая разумность действий противника в условиях противоположности целей, для решения игры с точки зрения игрока A необходимо найти такие смешанные стратегии p и q, при которых ему обеспечивался бы средний выигрыш, равный $\min_q \max_p f(p;q)$ Эту величину назовем *верхней ценой игры* и обозначим через β, т. е. $\beta = \min_q \max_p f(p;q)$

Аналогичной должна быть ситуация и относительно игрока *B:* Нижняя цена игры α должна равняться $\max_p \min_q f(p;q)$, т.е. $\alpha = \max_p \min_q f(p;q)$

По аналогии с играми, имеющими седловые точки в чистых стратегиях, назовем *оптимальными* смешанные стратегии p^* и q^* игроков A и B, удовлетворяющие равенству

$$\min_q \max_p f(p;q) = \max_p \min_q f(p;q) = f(p^*, q^*) \qquad (13.4)$$

Величину $f(p^*;\ q^*)$, полученную по формуле (13.4), называют *ценой игры.* Обозначим ее через v. Итак, $v = f(p^*;\ q^*)$.

В дальнейшем будем пользоваться и другим, эквивалентным данному, определением оптимальных смешанных стратегий: p^* и q^* называются *оптимальными смешанными стратегиями* соответственно игроков A и B, если они образуют седловую точку для платежной функции $f(p;\ q)$, т. е. удовлетворяют неравенству.

$$f(p;q^*) \le f(p^*;q^*) \le f(p^*;q) \qquad (13.5)$$

Из неравенства (13.5) следует, что в седловой точке $(p^*;\ q^*)$ платежная функция $f(p;\ q)$ достигает максимума, по смешанным стратегиям p игрока A и минимума по смешанным стратегиями q игрока B.

Свойства оптимальных смешанных стратегий.

Теорема.2. В смешанных стратегиях любая конечная матричная игра имеет седловую точку.

Теорема.3. Для того чтобы смешанные стратегии $p^* = (p_1;\ \ldots;\ p_m)$ и $q^* = (q_1^*;\ \ldots;\ q_n)$ *были оптимальными для игроков A и B в игре с матрицей* $||a_{ij}||$ *и ценой v, необходимо и достаточно выполнение неравенств*

$$\sum_{i=1}^{m} a_{ij} p_i^* \geq v \qquad (13.6)$$

$$\sum_{j=1}^{n} a_{ij} q_j^* \geq v \qquad (13.7)$$

Итак, чтобы проверить является ли набор $(p^*; q^*; v)$ решением матричной игры надо найти неотрицательное решение($p_1 \dots p_m; q_1 \quad q_n;$ v) системы $n+m$ линейных неравенств (13.6)-(13.7) и уравнений (13.8)

$$\sum_{i=1}^{m} p_i = 1, \quad \sum_{j=1}^{n} q_j = 1 \qquad (13.8)$$

*Теорема 4. **Е**сли один из игроков придерживается своей оптимальной смешанной стратегии, то его выигрыш остается неизменным и равным цене игры независимо от того, какую стратегию применяет другой игрок, если только не выходит за пределы своих активных стратегий.*

Решение игры можно упростить, если выявить в матрице игры доминирование одних стратегий над другими, это позволит сократить размерность матрицы.

Если в матрице игры элементы *k*-й не меньше соответствующих элементов *s*-й строки, то выигрыш игрока *А* при стратегии *Ак* будет больше, чем при стратегии *As*. Говорят, что стратегия *Ак* доминирует над стратегией *As*. *Ак* –доминирующая, A_s – доминируемая. Аналогично и для стратегии В.

Теорема 5. Пусть I — игра, в матрице которой k-я стратегия игрока А доминирует над s-й, а I' — игра, матрица которой получена из матрицы игры I исключением s-и строки. Тогда: а) цена игры I' равна цене игры I; б) оптимальная смешанная стратегия q = (q*; ...; qn) игрока В в игре I' является также его оптимальной смешанной стратегией и в игре I; в) если p'** $(p_1^* \dots p_{s-1}^*\ p_{s+1; \dots}^* p_m^*)$ *оптимальная смешанная стратегия игрока А в игре I', то его смешанная стратегия p* =*$(p_1^*; \dots p_{s-1}^*; 0\ p_{s+1; \dots}^* p_m^*)$ *является оптимальной в игре I.*

Аналогичная теорема доказывается для случая доминирования стратегий игрока *В*.

Теорема 6. Пусть p* *и* q* *— оптимальные t игроков А и В в игре I с матрицей* $||a_{ij}||$ с ценой игры v. *Тогда* p* *и* q* *будут оптимальными и в игре I' с матрицей* $||baij + c||_{mxn}$ *(где b* > 0) *и ценой* $v' = bv + c$.

Численные методы решения матричных игр. Связь теории игр с линейным программированием

Пусть имеем игру размерности *m*x*n* с матрицей $\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$.

Обозначим через $p^* = (p_1;...;p_n)$, $q^* = (q_1;...;q_n)$ оптимальные смешанные стратегии игроков *A* и *B*. Стратегия p^* игрока *A* гарантирует ему выигрыш не меньше ν, независимо от выбора стратегии *Bj* игрока В. Это можно записать так:

$$\left.\begin{array}{l} a_{11}p_1 + a_{21}p_2 + ... + a_{m1}p_m \ge \nu \\ a_{12}p_1 + a_{22}p_2 + ... + a_{m2}p_m \ge \nu \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}p_1 + a_{2n}p_2 + ... + a_{mn}p_m \ge \nu \end{array}\right\}, \qquad (13.9)$$

где $p_1 + p_2 + ... + p_m = 1; \quad p_i \ge 0 \quad (i = 1..m)$

Аналогично стратегия q^* игрока *B* гарантирует ему проигрыш не больше ν, независимо от выбора стратегии *Ai* игрока *A*, т.е.

$$\left.\begin{array}{l} a_{11}q_1 + a_{12}q_2 + ... + a_{1n}q_n \le \nu \\ a_{21}q_1 + a_{22}q_2 + ... + a_{2n}q_n \le \nu \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}q_1 + a_{m2}q_2 + ... + a_{mn}q_n \le \nu \end{array}\right\}, \qquad (13.10)$$

где $q_1 + q_2 + ... + q_n = 1; \quad q_j \ge 0 \quad (j = 1..n)$.

Поскольку элементы платежной матрицы всегда можно сделать положительными (теорема), то и цена игры $\nu > 0$.

Преобразуем системы (13.9) и (13.10), разделив обе части каждого неравенства на положительное число ν и введем новые обозначения: $\frac{p_i}{\nu} = x_i, \quad \frac{q_j}{\nu} = y_j \quad (i = 1..m; \quad j = 1..n)$.

Получим

$$\left.\begin{array}{l} a_{11}x_1 + a_{21}x_2 + ... + a_{m1}x_m \ge 1 \\ a_{12}x_1 + a_{22}x_2 + ... + a_{m2}x_m \ge 1 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}x_1 + a_{2n}x_2 + ... + a_{mn}x_m \ge 1 \end{array}\right\}, \qquad (13.11)$$

где $x_1 + x_2 + ... + x_m = 1/\nu; \quad x_i \ge 0 \quad (i = 1..m) \quad (13.12)$

и

$$\left.\begin{array}{l} a_{11}y_1 + a_{12}y_2 + ... + a_{1n}y_n \le 1 \\ a_{21}y_1 + a_{22}y_2 + ... + a_{2n}y_n \le 1 \\ .... \\ a_{m1}y_1 + a_{m2}y_2 + ... + a_{mn}y_n \le 1 \end{array}\right\} \qquad (13.13)$$

где $y_1 + y_2 + ... + y_n = 1/\nu; \quad y_j \ge 0 \quad (j = 1..n)$ (13.14).

Так как игрок A стремится максимизировать цену игры ν, то обратная величина $\frac{1}{\nu}$ будет минимизироваться, поэтому оптимальная стратегия игрока A определится из задачи линейного программирования следующего вида: найти минимальное значение целевой функции $f(x) = x_1 + x_2 + ... + x_m$ при ограничениях (13.11) и (13.12)

Оптимальная смешанная стратегия игрока B определится решением задачи следующего вида: найти максимальное значение целевой функции $\tilde{f}(y) = y_1 + y_2 + ... + y_n$ при ограничениях (13.13), (13.14).

Решив пару двойственных задач графическим (для случая двух переменных) или симплексным методом, далее определим:

$$\nu = \frac{1}{\sum_{i=1}^{m} x_i^*}, \quad p_i = \frac{x_i^*}{\sum_{i=1}^{m} x_i^*}, \quad q_j^* = \frac{y_j^*}{\sum_{i=1}^{m} x_i^*} \quad (i = 1..m, \quad j = 1..n).$$

Проиллюстрируем решение матричной игры сведением ее к ЗЛП.

<u>Пример 13.3.</u> Два сельскохозяйственных предприятий A и B выделяют денежные средства на строительство трех объектов. С учетом особенностей вкладов в строительство и местных условий прибыль предприятия A в зависимости в зависимости от объемов финансирования выражается элементами матрицы $\begin{bmatrix} 50 & 15 & 20 \\ 25 & 40 & 30 \\ 10 & 30 & 60 \end{bmatrix}$.

Будем предполагать, что убыток предприятия B при этом равен прибыли предприятия A. Требуется найти оптимальные стратегии предприятий А и В.

<u>Решение</u>. Обозначим чистые стратегии предприятий A и B через A_1, A_2, A_3 и B_1, B_2, B_3 соответственно. Предположим, что предприятие A располагает общей суммой a тыс. ден. ед., отпускаемой на строительство трех объектов. Аналогично и предприятие B имеет сумму в b тыс. ден. ед., отпускаемых на строительство тех же трех объектов. Тогда чистая стратегия A_1 - это выделение a_1 тыс. ден. ед. предприятия А на строительство первого объекта; A_2 - это выделение a_2 тыс. ден. ед. предприятия А на строительство второго объекта; A_3 -

это выделение a_3 тыс. ден. ед. предприятия A на строительство третьего объекта. Общая сумма средств, выделяемых на строительство трех объектов, $a = a_1 + a_2 + a_3$. Аналогично определяются чистые стратегии и для предприятия B.

Проверим игру на наличие седловой точки

$\alpha = \max\limits_i \min\limits_j a_{ij} = 25, \quad \beta = \min\limits_j \max\limits_i a_{ij} = 40, \quad \alpha \neq \beta$, поэтому решение игры определеним в смешанных стратегиях. Цена игры ν заключена между нижней α и верхней β ценами, т.е. $25 \le \nu \le 40$. Составим ЗЛП для каждого игрока.

Для игрока A: найти минимальное значение функции

$$f = x_1 + x_2 + x_3$$

при ограничениях

$$\left.\begin{aligned} 50x_1 + 25x_2 + 10x_3 &\ge 1 \\ 15x_1 + 40x_2 + 30x_3 &\ge 1 \\ 20x_1 + 30x_2 + 60x_3 &\ge 1 \end{aligned}\right\}$$

$$x_i \ge 0, \quad (i = 1, 2, 3)$$

Для игрока B: найти максимальное значение функции

$$\tilde{f} = y_1 + y_2 + y_3$$

при ограничениях

$$\left.\begin{aligned} 50y_1 + 15y_2 + 20y_3 &\le 1 \\ 25y_1 + 40y_2 + 30y_3 &\le 1 \\ 10y_1 + 30y_2 + 60y_3 &\le 1 \end{aligned}\right\}$$

$$y_j \ge 0, \quad (j = 1, 2, 3)$$

Вводя вспомогательные переменные $x_4 \ge 0, x_5 \ge 0, x_6 \ge 0$ для исходной задачи и $y_4 \ge 0, y_5 \ge 0, y_6 \ge 0$ для двойственной задачи, модели задач преобразуем к канонической форме. Соответствие между переменными пары двойственных задач будут следующими

$$\begin{array}{ccc|ccc} \multicolumn{3}{c}{\textit{Свободные}} & \multicolumn{3}{c}{\textit{Базисные}} \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_6 & y_1 & y_2 & y_3 \\ \multicolumn{3}{c}{\textit{Базисные}} & \multicolumn{3}{c}{\textit{Свободные}} \end{array}$$

Решим, например, двойственную ЗЛП, построенную для определения проигрыша предприятия B. Каноническая форма задачи имеет вид:

$$\max : \widetilde{f} = y_1 + y_2 + y_3$$

$$\left.\begin{aligned} 50y_1 + 15y_2 + 20y_3 + y_4 = 1 \\ 25y_1 + 40y_2 + 30y_3 + y_5 = 1 \\ 10y_1 + 30y_2 + 60y_3 + y_6 = 1 \end{aligned}\right\}$$

$$y_j \geq 0, \quad (j = 1..6)$$

Решив ее симплекс –методом, приходим к таблице 13.4, в которой содержится оптимальный план $y^* = \left(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*, y_6^*\right)$ =(0,0133; 0,0094; 0,0098;0;0;0). При этом $\widetilde{f}(y^*) = 0{,}0325$.

С учетом основной теоремы двойственности и соответствия между переменными оптимальный план исходной задачи запишется в виде $x^* = \left(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*\right)$=(0,0102;0,0180;0,0043;0;0;0) и $f(x^*) = 0{,}0325$.

По формулам $\nu = \dfrac{1}{f_{\max}}, \quad \dfrac{p_i}{\nu} = x_i, \quad \dfrac{q_j}{\nu} = y_j \ (i = 1..m, \quad j = 1..n)$ получим цену игры $\nu = \dfrac{1}{0{,}0325} \approx 30{,}77$ и вероятности p_i^*, q_j^* для оптимальных смешанных стратегий соответственно предприятий A и B:

$$p_1^* = 0{,}0102 \cdot 30{,}77 = 0{,}314 \quad q_1^* = 0{,}0133 \cdot 30{,}77 = 0{,}409$$

$$p_2^* = 0{,}0180 \cdot 30{,}77 = 0{,}554 \quad q_2^* = 0{,}0094 \cdot 30{,}77 = 0{,}289$$

$$p_3^* = 0{,}0043 \cdot 30{,}77 = 0{,}132 \quad q_1^* = 0{,}0098 \cdot 30{,}77 = 0{,}302$$

Таблица 13.4 –Последняя симплекс-таблица

БП	СЧ	-y1	-y2	-y3	-y4	-y5	-y6
y1	0,0133	1	0	0	0,0234	-0,0047	-0,0055
y2	0,0094	0	1	0	-0,0188	0,0438	-0,0156
y3	0,0098	0	0	1	0,0056	-0,0211	0,0254
$\widetilde{f}$	0,0325	0	0	0	0,0102	0,0180	0,0043

Таким образом, оптимальными смешанными стратегиями сельскохозяйственных предприятий *A* и *B* являются стратегии p^*=(0,314; 0,554; 0,132) и =(0,409;0,289;0,302) соответственно при гарантии получении предприятием А независимо от стратегий предприятия В прибыли не менее 30,77 тыс. ден. ед. Убыток предприятия В при этом составит не более 30,77 тыс. ден. ед.

Итак, из общей суммы средств α тыс. ден. ед., выделенных предприятием А на строительство трех объектов, на долю первого объекта должно выделяться 31,4%, второго – 55,4% и третьего -13,2% этой суммы. Аналогично распределяются b тыс. ден. ед. предприятием *B*. Так на долю первого объекта должно выделяться 40,9%, второго – 28,9% и третьего -30,2% этой суммы. Такое распределение денежных средств предприятиями *A* и *B* по трем строящимся объектам позволит им получить максимальную прибыль 30,77 тыс. ден. ед.

ТЕМА 12. ИГРЫ С ПРИРОДОЙ. КРИТЕРИИ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ

Лекция 13.2. Игры с природой. Принцип недостаточного основания Лапласа. Критерий Байеса. Максиминный критерий Вальда. Критерий минимального риска Сэвиджа. Критерий пессимизма-оптимизма Гурвица

Игры с природой

Управление производственными процессами осуществляется путем реализации последовательности принимаемых решений. Для этого необходима информация о состоянии объекта управления в условиях его работы. В случае отсутствия достаточно полной информации возникает неопределенность в принятии решения. Причины этого могут быть различными:

- невозможность получения информации к моменту принятия решения;
- слишком высокие затраты на получение информации;
- невозможность устранения неопределенности по причинам объективного характера и т. д.

Естественно, по мере совершенствования средств сбора информации, передачи и обработки ее неопределенность ситуации в момент принятия управленческих решений будет уменьшаться.

Существование неустранимой неопределенности связано со случайным характером многих явлений. Например, случайный характер спроса на продукцию делает невозможным точное прогнозирование объема ее выпуска. Принятие решения в этом случае связано с риском. Или, например, прием партии товара для контроля на соответствие стандарту также связан с риском. Правда, неопределенность при контроле может быть устранена в случае контроля всего товара, выпускаемого для реализации. Однако это может оказаться слишком дорогостоящим мероприятием.

С целью уменьшения неблагоприятных последствий в каждом конкретном случае следует учитывать степень риска и имеющуюся информацию. И здесь *лицо, принимающее решение* (ЛПР), вступает в игровые отношения с некоторым абстрактным лицом, которое условно молено назвать *"природой"*. Иными словами, ЛПР должно уметь находить управленческое решение, когда природа не выбирает сознательно свои оптимальные стратегии. Вместе с тем мы иногда располагаем некоторыми вероятностными характеристиками состояния природы. Такого рода ситуации принято называть *играми с природой*.

Любую хозяйственную деятельность человека можно рассматривать как игру с природой. В широком смысле под "природой" будем понимать совокупность неопределенных факторов, влияющих на эффективность принимаемых решений.

Задачей ЛПР является принятие наилучшего управленческого решения в каждой конкретной ситуации. Качество принимаемого решения зависит от информированности ЛПР о ситуации, в которой принимается решение. В случае неопределенности неквалифицированный экономист отказывается принимать решение или принимает его без достаточного обоснования. Хороший экономист руководствуется правилом: "информация — это деньги". Умение использовать даже неполную информацию для обоснования принимаемых решений — это задача экономиста.

Безразличие природы к результату игры (выигрышу) и возможность получения экономистом (статистиком) или ЛПР дополнительной информации о ее состоянии отличают игру с природой от обычной матричной игры, в которой принимают участие два сознательных игрока.

Игры с природой представляют собой основную модель теории принятия решений в условиях частичной неопределенности.

Множество состояний природы обозначим через П, отдельное состояние — $П_j$, $j = 1..n$. Множество решений (стратегий) статистика обозначим через *А,* отдельное решение — A_i, $i = 1..m$.

Если на множествах состояний природы *П* и решений статистика *А* потребуется определить распределение вероятностей, то необходимо производить эксперимент, целью которого будет нахождение распределения некоторой случайной переменной, зависящей от состояния природы.

Для *i*-го решения A_i статистика *А* и j-го состояния природы $П_j$ имеем некоторое число, обозначающее функцию потерь $L(A_i, П_j)$, которая, как правило, является случайной переменной.

Во взаимоотношениях с природой статистик может использовать любые из стратегий $A_1, A_2, ..., A_m$ в зависимости от состояний $П_j$ природы. Имея ряд стратегий $A_1, A_2, ..., A_m$, статистик должен руководствоваться некоторым правилом поведения, с помощью которого он определяет, какую стратегию A_i ему выбрать. Иными словами, статистик отыскивает оптимальное поведение, которое и будет его оптимальной стратегией. При этом он может пользоваться как чистыми, так и смешанными стратегиями.

Чтобы выразить в количественной форме упомянутое выше некоторое правило поведения статистика, которым он должен руководствоваться, предположим, что есть возможность численно оценить величиной a_{ij} эффективность каждой комбинации $(A_i, П_j)$, иначе говоря, качество решения A_i. Тем самым будет определена так называемая платежная матрица игры с природой

$$\begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \multicolumn{4}{c}{\dotfill} \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix}$$

на основе которой в дальнейшем и будут сформулированы "правила поведения" — критерии выбора оптимальной стратегии статистика.

Элемент a_{ij} - назовем выигрышем статистика, если он использует стратегию A_i при состоянии природы $П_j$.

Решение игры с природой несколько отличается от решения обычной матричной игры, где оба игрока ведут игру сознательно.

Отличие состоит прежде всего в упрощении игры. Выявление дублирующих и доминируемых стратегий производится только для стратегий статистика. Стратегии природы нельзя опускать, поскольку она не имеет "умысла" навредить статистику, более того, она может реализовать состояния, заведомо выгодные статистику. Иногда при решении игры с природой используется *матрица рисков.* Элементы r_{ij} матрицы рисков равны разности между максимально возможным выигрышем и тем выигрышем, который статистик получит в тех же условиях $П_j$, применяя стратегию A_i т. е. $r_{ij} = \beta_j - a_{ij}$ где $\beta_j = \max\limits_i a_{ij}$

Принцип недостаточного основания Лапласа. Критерий Байеса. Максиминный критерий Вальда. Критерий минимального риска Сэвиджа. Критерий пессимизма-оптимизма Гурвица

Оптимальную стратегию статистика можно определить, используя ряд критериев. Так, при известном распределении вероятностей различных состояний $П_j$ природы пользуются ***критерием Байеса***. Показателем в этом критерии служит либо величина среднего выигрыша, либо величина среднего риска.

Платежную матрицу $\lfloor a_{ij} \rfloor_{m\times n}$ представим в виде табл. 13.5

Таблица 13.5 – Платежная матрица

Стратегия статистика A_i	Состояние природы $П_j$				Средний выигрыш
					$\bar{a}_1$
A_1	a_{11}	a_{12}	…	a_{1n}	$\bar{a}_2$
A_2	a_{21}	a_{22}	…	a_{2n}	…
…	…	…	…	…	$\bar{a}_m$
A_m	a_{m1}	a_{m2}	…	a_{mn}	
q_j	q_1	q_2	…	q_n	

По критерию Байеса за оптимальную принимается та чистая стратегия A_i, при которой максимизируется средний выигрыш $\bar{a}_i$ статистика, т. е. обеспечивается $\bar{a} = \max\limits_i \bar{a}_i$ где $\bar{a}_i = \sum\limits_{j=1}^{n} a_{ij} q_j$, $i = 1..m$.

Матрицу рисков представим в виде табл.13.6. За оптимальную стратегию статистика принимается чистая стратегия A_i при которой минимизируется средний риск, т.е. обеспечивается $\bar{r} = \min\limits_i \bar{r}_i$ где

$$\bar{r}_i = \sum_{j=1}^{n} r_{ij} q_j, \; i = 1..m.$$

Таблица 13.6 – Матрица рисков

Стратегия статистика A_i	Состояние природы $П_j$				Средний риск
					$\bar{r}_1$
A_1	r_{11}	r_{12}	…	r_{1n}	$\bar{r}_2$
A_2	r_{21}	r_{22}	…	r_{2n}	…
…	…	…	…	…	$\bar{r}_m$
A_m	r_{m1}	r_{m2}	…	r_{mn}	
q_j	q_1	q_2	…	q_n	

В случае, когда вероятности состояний природы правдоподобны, для их оценки используют ***принцип недостаточного основания Лапласа,*** согласно которому все состояния природы полагаются равновероятными, т. е. Оптимальной считается стратегия, обеспечивающая максимум среднего выигрыша.

Если вероятности состояний природы неизвестны, то для решения игр с природой — выбора оптимальной стратегии статистика — можно использовать несколько критериев.

Максиминный критерий Вальда совпадает с критерием выбора максиминной стратегии, позволяющей получать нижнюю чистую цену α в парной игре с нулевой суммой. По критерию Вальда за оптимальную принимается чистая стратегия, которая в наихудших условиях гарантирует максимальный выигрыш, т. е. $\alpha = \max_i \min_j a_{ij}$.

Критерий минимального риска Сэвиджа рекомендует выбирать в качестве оптимальной стратегии ту, при которой величина максимального риска минимизируется в наихудших условиях, т. е. обеспечивается $\min_i \max_j r_{ij}$

Критерии Вальда и Сэвиджа ориентируют статистика на самые неблагоприятные состояния природы, т.е. эти критерии выражают пессимистическую оценку ситуации.

Критерий Гурвица является критерием пессимизма-оптимизма. За оптимальную принимается та стратегия, для которой выполняется соотношение $\max_i \left(\lambda \min_j a_{ij} + (1-\lambda) \max_j a_{ij} \right)$

где . $0 \le \lambda \le 1$При $\lambda = 0$имеем критерий крайнего оптимизма, а при $\lambda = 1$— критерий пессимизма Вальда. Если $0 < \lambda < 1$то имеем нечто среднее. При желании подстраховаться в данной ситуации λ принимают близким к единице. В общем случае число λ выбирают исходя из опыта или субъективных соображений.

Решение игры с природой по рассмотренным критериям позволяет более обоснованно принимать ту стратегию, которая гарантирует статистику больший выигрыш по сравнению с выигрышем, принимаемым статистиком интуитивно или исходя из опыта.

Пример 13.4. В соответствии со спросом на продукцию *q-й* номенклатуры в городе планируется построить предприятие по производству этой продукции. Неопределенность спроса в период t приводит к тому, что необходимо рассчитать объем выпускаемой продукции V_q, который должен быть не меньше уровня спроса S_q, чтобы не потерять потенциально возможный доход от реализации продукции, а также не должен превышать уровень спроса, так как предприятие будет нести убытки, связанные в основном с уценкой. Предположим, что в течение года (по кварталам) спрос на продукцию *q-й* номенклатуры выражается величинами 10, 20, 30, 40 тыс. шт. В таком случае и планирующий орган предприятия может принять одно из следующих решений: построить предприятие, которое могло бы удовлетворить спрос потребителей в 10, 20, 30, 40 тыс. шт. *q-й* продукции. Работа подобных предприятий показывает, что предприятие терпит издержки от нереализованной единицы *q-й* продукции 5 ден. ед., а доход от реализации единицы продукции составляет 15 ден. ед. Функцию платежей можно записать в виде кусочно-линейной функции потерь:

$$L\left(S_q, V_q\right)=\begin{cases} K'_q\left(V_q - S_q\right), & V_q \geq S_q \\ K''_q\left(S_q - V_q\right), & V_q < S_q \end{cases}$$

Требуется: 1) придать описанной ситуации игровую схему, установить характер игры и выявить ее участников; 2) вычислить элементы платежной матрицы и составить ее; 3) дать обоснованные рекомендации планирующему органу на строительство предприятия, которое могло бы обеспечить спрос потребителей на *q-ю* продукцию.

При изучении работы аналогичных предприятий планирующий орган располагает некоторой дополнительной информацией, снижающей неопределенность ситуации: известны вероятности спроса на данную продукцию по кварталам года: 0,3; 0,2; 0,4; 0,1; спрос на продукцию в каждом квартале равновероятен; о вероятностях спроса на указанную продукцию по кварталам ничего определенного сказать нельзя.

Решение. В качестве статистика выступает планирующий орган, который может принять одно из следующих решений: построить предприятие, способное удовлетворить спрос потребителей в 10 тыс. ед. продукции (стратегия A_1); построить предприятие мощностью в 20тыс. ед. продукции (стратегия A_2) построить предприятие мощностью в 30 тыс. ед. продукции (стратегия A_3); построить предприятие мощностью в 40 тыс. ед. продукции (стратегия A_4).

Второй играющей стороной — природой — будем считать совокупность объективных внешних условий, в которых формируется спрос потребителей. Спрос по кварталам года различен: в первом квартале — 10 тыс. ед. — будет означать состояние $П_1$,спрос на продукцию во втором квартале в объеме 20 тыс. ед. — состояние $П_2$, спрос на продукцию в третьем квартале в объеме 30 тыс. ед. — состояние $П_3$; спрос на продукцию в четвертом квартале в объеме 40 тыс. ед. — состояние $П_4$. Итак, описанная ситуация представляет собой игру с природой.

Рассчитаем элементы платежной матрицы (табл. 13.6).

Таблица 13.6 – Расчет платежной матрицы

Мощность предприятия, тыс. ед.	Стратегия статистка A_i	Спрос потребителей по кварталам года, тыс. ед.			
		10	20	30	40
		Состояние спроса $П_j$			
		$П_1$(10)	$П_2$(20)	$П_3$(30)	$П_4$(40)
10	A_1	150	150	150	150
20	A_2	100	300	300	300
30	A_3	50	250	450	450
40	A_4	0	200	400	600
	q_j	0,3	0,2	0,4	0,1

Так, в ситуации (A_1, $П_1$) элемент a_{11} вычисляется следующим образом. Плановый орган принимает решение построить предприятие мощностью в 10 тыс. ед., что и соответствует состоянию спроса в 10 тыс. ед. Доход от производства 10 тыс. ед. продукции a_{11} = 10* 15 = 150 тыс. ден. ед.

Элемент a_{12} в ситуации (A_1 , $П_2$) рассчитываем так. Предприятие строится в расчете на выпуск 10 тыс. ед. продукции, а спрос на нее составляет 20 тыс. ед. Если бы предприятие могло обеспечить этот спрос, то доход составил бы 20*15 = 300 тыс. ден. ед. Однако спрос удовлетворяется лишь на 10 тыс. ед. продукции, следовательно, предприятие недополучит доход 10*15 = 150 тыс. ден. ед. Элемент a_{12} =20*15 — 10*15 = 150.

Аналогично определяются и другие элементы табл. 3, например элемент a_{31} для ситуации (A_3, $П_1$). Предприятие строится в расчете на выпуск 30 тыс. ед. продукции, а спрос на нее составляет 10 тыс. ед., тогда доход предприятия от реализации 10 тыс. ед. составит 10*15 = 150 тыс. ден. ед., а от нереализованной продукции (20 тыс. ед.) предприятие терпит издержки 20*5 = 100 тыс. ден. ед., доход предприятия в ситуации (A_3, $П_1$). составит a_{32}= 10*15 — 20*5 = 50 тыс. ден. ед.

Аналогично рассчитываем и остальные элементы платежной матрицы (табл. 3).

Вычисляем средние выигрыши:

$$\overline{a}_1 = 150 * 0,3 + 150 * 0,2 + 150 * 0,4 + 150 * 0,1 = 150$$

Аналогично $\overline{a}_2 = 240, \overline{a}_3 = 290, \quad \overline{a}_4 = 260$

<u>Оптимальной</u> стратегией по Байесу является A_3 поскольку ей соответствует максимальная средняя прибыль $\overline{a}_3 = \max(150, 240, 290, 260)$ (тыс. ден. ед.)

По критерию Лапласа, когда средние выигрыши равны: $q_1 = q_2 = q_3 = q_4 = \dfrac{1}{4}$

$$\overline{\alpha}_1 = \frac{1}{4}(150 + 150 + 150 + 150) = 150, \quad \overline{\alpha}_2 = 250, \quad \overline{\alpha}_3 = 300, \quad \overline{\alpha}_4 = 300$$

Оптимальными стратегиями по Лапласу являются *Аз* и A_4, так как им соответствует максимальная прибыль, равная 300 тыс. ден. ед.

По критерию Вальда оптимальной является стратегия A_1 , для которой прибыль достигает наибольшей величины, равной 150 тыс. ден. ед. В самом деле,

$$\alpha = \max_i \min_j a_{ij} = \max(150, 100, 50, 0) = 150 \text{ тыс. ден. ед.}$$

Построим матрицу рисков (табл. 13.7).

Таблица 13.7 – Матрица рисков

Мощность предприятия, тыс. ед.	Стратегия статистка A_i	Спрос потребителей по кварталам года, тыс. ед.				$\max_j r_{ij}$
		10	20	30	40	
		Состояние спроса $П_j$				
		$П_1(10)$	$П_2(20)$	$П_3(30)$	$П_4(40)$	
10	A_1	0	150	300	450	450
20	A_2	50	0	150	300	300
30	A_3	100	50	0	150	150
40	A_4	150	100	50	0	150

По критерию Сэвиджа оптимальными являются стратегии *Аз* и A_4, для которых в наихудших условиях величина r риска принимает наименьшее значение, равное 150 тыс. ден. ед. В самом деле, $r = \min_i \max_j r_{ij} = \min(450,300,150,150) = 150$

Таким образом, в результате решения игры с природой по различным критериям чаще других рекомендовалась стратегия *Аз*-

Следовательно, нужно строить предприятие мощностью в 30 тыс. ед. продукции. Прибыль при этом, если вероятности спроса известны, составит 290 тыс. ден. ед., при равновероятных условиях спроса — 300 тыс. ден. ед.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Дайте следующие определения: игра, исход игры, стратегия, партия, ход.
2. По каким признакам можно классифицировать игры?
3. Матричные игры.
4. Что такое платежная матрица?
5. Что называется нижней чистой ценой игры (максимином)?
6. Что называется верхней чистой ценой игры (минимаксом)?
7. Когда игра имеет седловую точку?
8. Что такое смешанные стратегии для игрока A и игрока B?
9. Как выявить в матрице игры доминирование одних стратегий над другими?
10. Применение матричной игры к задаче линейного программирования.
11. Какого рода ситуации принято называть играми с природой?
12. Принцип недостаточного основания Лапласа.
13. Критерий Байеса.
14. Максиминный критерий Вальда.
15. Критерий минимального риска Сэвиджа.
16. Критерий пессимизма-оптимизма Гурвица.

РАЗДЕЛ 6. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

ТЕМА 13. ЗАДАЧИ МАССОВОГО ОБСЛУЖИВАНИЯ

Лекция 14. Общая характеристика систем массового обслуживания. Классификация систем массового обслуживания и их основные характеристики. Задачи анализа одноканальных систем массового обслуживания. Задачи анализа многоканальных систем массового обслуживания. Задача синтеза (оптимизации) одноканальной замкнутой системы массового обслуживания с ожиданием. Задача синтеза (оптимизации) многоканальной замкнутой системы массового обслуживания с ожиданием.

Общая характеристика систем массового обслуживания. Классификация систем массового обслуживания и их основные характеристики.

При исследовании операций часто приходится сталкиваться с работой своеобразных систем, называемых системами массового обслуживания (СМО). Примерами таких систем могут служить: телефонные станции, ремонтные мастерские, билетные кассы, справочные бюро, магазины, парикмахерские и т. п.

Каждая СМО состоит из какого-то числа обслуживающих единиц (или «приборов»), которые мы будем называть к а н а л а м и обслуживания. Каналами могут быть: линии связи, рабочие точки, кассиры, продавцы, лифты, автомашины и др. СМО могут быть одноканальными и многоканальными.

Всякая СМО предназначена для обслуживания какого-то п о т о к а з а я в о к (или «требований»), поступающих в какие-то случайные моменты времени. Обслуживание заявки продолжается какое-то, вообще говоря, случайное время $T_{об}$, после чего канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времен обслуживания приводит к тому, что в какие-то периоды времени на входе СМО скапливается излишне большое число заявок (одни либо становятся в очередь, либо покидают СМО необслуженными); в другие же периоды СМО будет работать с недогрузкой или вообще простаивать.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем; состояние СМО меняется скачком в моменты появления каких-то событий (или прихода новой заявки, или окончания обслуживания, или момента, когда заявка, которой надоело ждать, покидает очередь).

Предмет теории массового обслуживания — построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками — показателями эффективности СМО. описывающими, с той или другой точки зрения, ее способность справляться с потоком заявок. В качестве таких показателей (в зависимости от обстановки и целей исследования) могут применяться разные величины, например: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналом; среднее число заявок в очереди и среднее время ожидания обслуживания: вероятность того, что число заявок в очереди превысит какое-то значение, и т. д. Среди заданных условий работы СМО мы намерение не выделяем элементов решения: ими могут быть, например, число каналов, их производительность, режим работы СМО и т. д.

Математический анализ работы СМО очень облегчается, если процесс этой работы — марковский. Для этого достаточно, чтобы все потоки событий, переводящие систему из состояния в состояние (потоки заявок, потоки «обслуживании»), были простейшими. Если это свойство нарушается, то математическое описание процесса становится гораздо сложнее и довести его до явных, аналитических формул удается лишь в редких случаях.

Системы массового обслуживания делятся на типы (или классы) но ряду признаков. Первое деление: СМО с о т к а з а м и 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 уходит, а становится в очередь и ожидает возможности быть обслуженной. На практике чаще встречаются (и имеют большее значение) СМО с очередью; недаром теория массового обслуживания имеет второе название: «теория очередей».

СМО с очередью подразделяются на разные виды, в зависимости от того, как организована очередь — ограничена она или не ограничена. Ограничения могут касаться как длины очереди, так и времени ожидания (так называемые «СМО с нетерпеливыми заявками»). При анализе СМО должна учитываться также и «дисциплина обслуживания»—заявки могут обслуживаться либо в порядке поступления (раньше пришла, раньше обслуживается), либо в случайном порядке. Нередко встречается так называемое обслуживание с приоритетом — некоторые заявки обслуживаются вне очереди. Приоритет может быть как абсолютным, когда заявка с более высоким приоритетом и вытесняет из обслуживания заявку с низшим (например, пришедший в парикмахерскую клиент высокого ранга прогоняет с кресла обыкновенного клиента), так и относительным — когда начатое обслуживание доводится до конца, а заявка с более высоким приоритетом имеет лишь право на лучшее место в очереди.

Существуют СМО с так называемым многофазовым обслуживанием, состоящим из нескольких последовательных или «фаз» (например, покупатель, пришедший в магазин, должен сначала выбрать товар, затем оплатить его в кассе, затем получить на контроле).

Кроме этих признаком, СМО делятся на два класса: «открытые» и «замкнутые». В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято). В замкнутой СМО — зависят. Например, если один рабочий обслуживает группу станков, время от времени требующих наладки, то интенсивность потока «требований» со стороны станков зависит от того, сколько их уже неисправно и ждет наладки. Это — пример замкнутой СМО. Классификация СМО далеко не ограничивается приведенными разновидностями.

Оптимизация работы СМО может производиться под разными углами зрения: с точки зрения организаторов (или владельцев) СМО или с точки зрения обслуживаемых клиентов. С первой точки зрения желательно «выжать все, что возможно» из СМО и добиться того, чтобы ее каналы были предельно загружены. С точки зрения клиентов желательно всемерное уменьшение очередей, которые зачастую становятся настоящим «бичом быта», приводя к бессмысленной трате сил и времени и, в конечном итоге, к понижению производительности труда. При решении задач оптимизации в теории массового обслуживания существенно необходим «системный подход», полное и комплексное рассмотрение всех последствий каждого решения. Например, с точки зрения клиентов СМО желательно увеличение числа каналов обслуживания; но ведь работу каждого канала надо оплачивать, что удорожает обслуживание. Построение математической модели позволяет решить оптимизационную задачу о разумном числе каналов с учетом всех «за» и «против».

Задачи анализа одноканальных систем массового обслуживания

Задача анализа детерминированной системы. Постановка задачи.

Пусть исследуется производственный процесс, в котором поступление требований происходит через равные промежутки времени $\Delta t_n = const$ (т. е. интенсивность потока поступлений требований $\lambda = 1/\Delta t_n = const$) и обслуживание производится через равные промежутки времени $\Delta t_{об} = const$ (т.е. интенсивность обслуживания $\mu = 1/\Delta t_{об} = const$). Имеется один канал обслуживания. Предполагается, что $\Delta_{об}/\Delta t_n = \lambda/\mu < 1$ (в противном случае очередь будет бесконечно возрастать) и что к началу обслуживания в системе имеется уже *n* требований. Определить через какое время очередь исчезнет.

Задача анализа разомкнутой системы с ожиданием (потоки требований пуассоновские). Постановка задачи.

Пусть задана некоторая система массового обслуживания, для которой справедливы следующие гипотезы:

- вероятность поступления требований не зависит от принятого начала отсчета времени, а зависит только от продолжительности периода наблюдений (стационарность потока);
- не поступают в систему и не покидают ее одновременно два или более требований (поток ординарный);
- поступление одного требования не зависит от поступления другого (отсутствие последствия).

Известны также интенсивность λ поступления потока требований (среднее число поступлений требований в в единицу времени $\lambda = 1/\Delta t_n$ и интенсивность μ обслуживания требований (среднее число обслуживаний в единицу времени $\mu = 1/\Delta t_{об}$. Требуется определить основные характеристики системы:

- вероятность простоя канала обслуживания P_0;
- вероятность того, что в системе находится *n* требований, P_n;
- среднее число требований, находящихся в системе, $N_{сист}$(в очереди и обслуживании);
- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее время ожидания требования в системе $T_{сист}$.

Задача анализа замкнутой системы с ожиданием (потоки требований пуассоновские). Постановка задачи.

Пусть исследуется некоторая замкнутая система массового обслуживания с ограниченным количеством требований в системе, т.е. обслуженные требования вновь возвращаются в систему обслуживания (например, экскаватор или автосамосвал). Интенсивность поступления одного требования в систему известна и равна λ. Интенсивность обслуживания требований известна и равна μ. Число требований, нуждающихся в обслуживании равно *m*. Требуется определить основные характеристики системы:

- вероятность того, что в системе находится *n* требований, P_n;
- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее число требований, находящихся в системе, $N_{сист}$;
- среднее время ожидания требования в системе $T_{сист}$;
- среднее время ожидания требования в очереди $T_{оч}$.

Задачи анализа многоканальных систем массового обслуживания

Задача анализа разомкнутой системы с ожиданием (потоки требований пуассоновские). Постановка задачи.

Пусть известны интенсивность λ поступления потока требований в систему и интенсивность μобслуживания этих требований. Число каналов обслуживания N_k. Требуется определить:

- вероятность того, что в системе имеется *n* требований, $P_n(t)$;
- вероятность простоя каналов обслуживания $P_0(t)$;

- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее время ожидания требования в очереди $T_{оч}$.
- среднее число свободных каналов обслуживания $N_{св}$.

Задача анализа разомкнутой системы с отказом требований пуассоновские). Постановка задачи.

Пусть имеется некоторая разомкнутая система массового обслуживания. Пусть известны интенсивность поступления потока требований в систему и интенсивность μ облуживания каждым каналом. Если требование застало все *N* каналов занятыми, то оно получает отказ и покидает систему.

Требуется определить:

- вероятность P_0 того, что все каналы свободны
- вероятность P_n того, что занято *n* каналов обслуживания;
- среднее число занятых каналов обслуживания.

Задача анализа замкнутой системы с ожиданием (потоки требований пуассоновские). Постановка задачи.

Пусть имеется некоторая замкнутая система массового обслуживания, в которой обслуженные требования вновь возвращаются в систему. Интенсивность поступления одного требования в систему известна и равна λ, интенсивность обслуживания каждого канала известна и равна μ. Число каналов обслуживания *N*, число требований, нуждающихся в обслуживании *m*. Будем считать, что $N \le m$. Требуется определить:

- вероятность того, что в системе находятся *n* требований, $P_n(t)$;
- вероятность простоя каналов обслуживания $P_0(t)$;
- среднее число требований, ожидающих начала обслуживания, $N_{оч}$;
- среднее время ожидания требования в очереди $T_{оч}$.

Задача синтеза (оптимизации) одноканальной замкнутой системы массового обслуживания с ожиданием

Постановка задачи и выбор критерия оптимизации. Пусть исследуется одноканальная замкнутая система массового обслуживания, для которой известны характеристики канала обслуживания и характеристики требований, поступающих на обслуживание. Требуется определить оптимальную структуру системы, т.е. оптимальное число требований $m_{опт}$, необходимых для обслуживания канала, чтобы эффективность системы была максимальна. В качестве критерия оптимизации примем удельные приведенные затраты, характеризующие затраты всей системы на одно обслуживание.

Аналитически выражение критерия оптимизации для определения оптимальной структуры одноканальной замкнутой системы массового обслуживания с ожиданием запишется так:

$$Y = \frac{P_0 C_{пк} + (1-P_0)C_{рк} + mC_{пт}}{\mu(1-P_0)} + \frac{E_н(S_к + mS_m)}{T_г \mu(1-P_0)} \text{ (руб./ед. продукции)},$$

где P_0 – вероятность простоя канала обслуживания; m – число требований, нуждающихся в обслуживании; $C_{пк}$ – средние затраты при простое канала обслуживания в течение часа из-за несвоевременного поступления требований на обслуживание, руб.; $C_{рк}$ – средние затраты при работе канала обслуживания в течение часа, руб; μ – интенсивность обслуживания канала, 1/ч; $\mu = 1/t_{об}$; $t_{об}$ – среднее время обслуживания требования каналом, ч; $C_{пт}$ – средние затраты содержания требования в течение часа, руб; $S_к, S_m$ – капитальные вложения соответственно на канал обслуживания и требование, руб; $T_г$ – годовой режим работы – число часов работы системы в году; $E_н$ – нормативный коэффициент эффективности.

Задача синтеза (оптимизации) многоканальной замкнутой системы массового обслуживания с ожиданием

<u>**Постановка задачи и выбор критерия оптимизации.**</u> Пусть исследуется некоторая многоканальная замкнутая система массового обслуживания. Известны характеристики каналов обслуживания и характеристики требований, поступающих на обслуживание. Требуется определить такую структуру многоканальной системы, чтобы эффективность системы была максимальна. В качестве критерия оптимизации принимаем целевую функцию – удельные приведенные затраты, т.е. затраты, приходящиеся на одно обслуживание. Используем обозначения и допущения те же, что и для одноканальной замкнутой системы.

Аналитическое выражение критерия оптимизации для определения оптимальной структуры многоканальной замкнутой системы массового обслуживания с ожиданием запишется так:

$$Y = \frac{P_0 C_{пк} N + (1 - P_0) C_{рк} N + N C_{пт}}{\mu(1 - P_0)} + \frac{E_н (S_к N + m S_m)}{T_г \mu(1 - P_0)},$$

где N – число каналов обслуживания в системе.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Приведите примеры систем массового обслуживания (СМО).
2. Что может быть каналами обслуживания.
3. Для чего предназначена всякая СМО?
4. Что изучает предмет теория массового обслуживания?
5. Дайте классификацию систем массового обслуживания.
6. Задача анализа детерминированной системы. Постановка задачи.
7. Задача анализа разомкнутой системы с отказом требований пуассоновские). Постановка задачи.
8. Задача анализа разомкнутой многоканальной системы с ожиданием (потоки требований пуассоновские). Постановка задачи.
9. Задача анализа разомкнутой многоканальной системы с отказом требований пуассоновские). Постановка задачи.
10. Задача анализа замкнутой многоканальной системы с ожиданием (потоки требований пуассоновские). Постановка задачи.
11. Задача синтеза (оптимизации) одноканальной замкнутой системы массового обслуживания с ожиданием.
12. Задача синтеза (оптимизации) многоканальной замкнутой системы массового обслуживания с ожиданием

РАЗДЕЛ 7. ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ТЕХНОЛОГИИ МАШИНОСТРОЕНИЯ

ТЕМА 14. СТАТИСТИЧЕСКИЕ МЕТОДЫ В ТЕХНОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Лекция 15. Исследование влияния технологических факторов на точность обработки и шероховатость поверхности. Обработка экспериментальных данных по способу наименьших квадратов.

Исследование влияния технологических факторов на точность обработки и шероховатость поверхности

При исследовании технологических процессов часто приходится сталкиваться с величинами переменного характера, которые и результате опытов могут принимать любые численные значения в определенных интервалах и, следовательно, по своему характеру относятся к категории случайных величин. Например, суммарная погрешность обработки, отдельные виды погрешностей обработки, средняя высота неровностей на обработанной поверхности, величина разбивки отверстий при развертывании, величина усадки отверстия после дорнования, сила резания при обработке металлов и т. п. представляют собой случайные величины. Под влиянием большого числа факторов они могут при испытаниях принимать любые численные значения (в определенных интервалах) и наблюденные значения этих величин будут иметь рассеивание.

Большей частью случайные величины, встречающиеся в технике, в том числе и в технологии машиностроения, подчиняются закону нормального распределения. Экспериментальным путем всегда можно проверить, подчиняется ли распределение интересующей нас случайной величины закону нормального распределения. Повторяя опыт не менее 50 раз, можно рассматривать его как случайную большую выборку из некоторой генеральной совокупности. Проверив выборку на случайность и нормальность, можно убедиться в справедливости предположения о подчинении изучаемого признака нормальному распределению или по эмпирической кривой распределения установить близость его к какому-либо другому закону распределения случайных величин, рассматриваемых в теории вероятностей и математической статистике.

Если распределение окажется нормальным или близким к нему, то в дальнейших исследованиях может быть широко использован выборочный метод, подробно разработанный для выборок из нормальных совокупностей. С помощью выборочного метода можно с достаточной надежностью и требуемой точностью решать такие важные и часто встречающиеся в исследованиях задачи, как установление влияния различных факторов технологического процесса на исследуемый признак и установление связей между изучаемыми качественными или количественными признаками какого-либо массового процесса.

Обработка экспериментальных данных по способу наименьших квадратов

Исследования могут преследовать различные цели. Одной из них может быть установление функциональных или корреляционных связей между переменными величинами и выражение этих связей в виде эмпирических формул. При этом отыскание функциональных связей, например между двумя переменными величинами, носит условный характер.

В природе нет связей только между двумя величинами. Обычно каждая величина зависит от ряда других величин, которые часто заранее не известны. Поэтому задача отыскания функциональной связи между двумя переменными величинами возможна только в случае, если влияние других аргументов на изучаемую величину либо пренебрежимо мало, либо они сохраняют (хотя бы приблизительно) постоянные значения во всех наблюдениях. Последнее предположение означает, что остальные аргументы входят в функциональную зависимость в качестве постоянных параметров. Если при таких предположениях функциональная связь между переменными величинами обнаруживается, то возникает задача выбора формы связи и вида эмпирической формулы. Если функциональной связи не обнаружено, а установлено наличие корреляционной между переменными величинами, то и здесь возникает задача выбора формы связи и вида эмпирического уравнения корреляционной связи.

Для установления формы связи (прямолинейной или криволинейной) прибегают к графическому методу. По экспериментальным данным в системе декартовых или логарифмических координат строят эмпирическую кривую. По ее виду подбирают наиболее близкую теоретическую кривую, уравнение которой известно. Это уравнение и принимается в качестве эмпирической формулы, определяющей функциональную или корреляционную связь между изучаемыми величинами. Выбрав, таким образом, наиболее подходящий вид формулы, определяют ее параметры. Для этого пользуются способом наименьших квадратов, сущность которого заключается в следующем.

Пусть имеются некоторые переменные, связанные между собой уравнением вида:

$$ax + by + yz = N \tag{15.1}$$

Чтобы определить неизвестные значения коэффициентов, необходимо было бы найти опытным путем значения их величины при трех различных комбинациях переменных, в результате получились бы три уравнения:

$$\begin{aligned} &\alpha x_1 + \beta y_1 + \gamma z_1 = N_1; \\ &\alpha x_2 + \beta y_2 + \gamma z_2 = N_2; \\ &\alpha x_3 + \beta y_3 + \gamma z_3 = N_3. \end{aligned} \tag{15.2}$$

имеющих три неизвестных α, β, γ, значения которых можно определить, решив систему трех уравнений. Если переменным *x, y, z* придать некоторые четвертые значения: x_4, y_4, z_4 и определить опытным путем значение N_4 то окажется, что равенство (15.1) при значениях коэффициентов, вычисленных по уравнениям (15.2),- удовлетворено не будет:

$$\alpha x_4 + \beta y_4 + \gamma z_4 \neq N. \tag{15.3}$$

Такой результат объясняется следующим. Если бы коэффициенты были определены из уравнений (15.2) на основании математически точного обмера переменных *x, y, z* и *N*, равенство (15.1) было бы удовлетворено и при других значениях переменных *x, y, z*. Однако всякие измерения являются лишь относительно точными, так как экспериментатор и прибор, которым он пользуется, дают ряд неизбежных ошибок. Кроме того, значения измеряемых величин могут колебаться в каждом опыте под влиянием случайных причин.

Ошибки измерений и колебаемость измеряемых величин приводят к результатам, выраженным неравенством (15.3)/

Очевидно, значения неизвестных α, β, γ надо находить в результате не трех (математически достаточных) опытов, а значительно большего количества их, чтобы получить наиболее надежные значения этих коэффициентов.

В этом случае получим систему *n* уравнений:

$$\left.\begin{aligned} &ax_1 + \beta y_1 + \gamma z_1 = N_1; \\ &ax_2 + \beta y_2 + \gamma z_2 = N_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &ax_n + \beta y_n + \gamma z_n = N_n \end{aligned}\right\} \qquad (15.4)$$

Полученная система уравнений носит название «избыточной» вследствие того, что число уравнений больше, чем число входящих в них неизвестных. Сами же уравнения называются «условными», так как они не вполне совместимы, т. е. значения неизвестных, определенные, например, из первых трех уравнений, не будут равны значениям этих же неизвестных, определенных из других трех уравнений этой системы.

Таким образом, возникает вопрос, каковы должны быть значения коэффициентов α, β, γ, чтобы избыточная система уравнений удовлетворялась наилучшим образом, т. е. ошибки $E_t = N - N_t$, где N — расчетные, а N_t — опытные данные или

$$\left.\begin{aligned} &E_1 = ax_1 - \beta y_1 - \gamma z_1 - N_1; \\ &E_2 = ax_2 - \beta y_2 - \gamma z_2 - N_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &E_n = ax_n - \beta y_n - \gamma z_n - N_n \end{aligned}\right\} \qquad (15.5)$$

должны быть наименьшими.

Для решения поставленной задачи пользуются принципом, предложенным Лежандром и математически обоснованным в дальнейшем Лапласом и Гауссом. Согласно этому принципу из всех возможных величин α, β, γ наиболее удовлетворительными будут те, при которых сумма квадратов ошибок будет наименьшая:

$$E_1^2 + E_2^2 + \ldots + E_n^2 = \min.$$

Возведение в квадрат, например, первого уравнения системы (15.5) дает

$$E_1^2 = \alpha^2 x_1^2 + \alpha x_1 \beta y_1 + \beta^2 y_1^2 + 2\alpha\gamma x_1 z_1 + 2\beta y_1 \gamma z_1 - 2\alpha x_1 N_1 - 2\beta y_1 N_1 + \gamma_2 z_2 - 2\gamma z_1 N_1 + N_1^2.$$

Отсюда получаем следующее уравнение:

$$\sum_1^n E_i^2 = \alpha^2 \sum_1^n x_i^2 + \beta^2 \sum_1^n y_i^2 + \gamma^2 \sum_1^n z_i^2 + 2\alpha\beta \sum_1^n x_i y_i + 2\alpha\gamma \sum_1^n x_i z_i + 2\beta\gamma \sum_1^n y_i z_i - 2\alpha \sum_1^n x_i N_i - 2\beta \sum_1^n y_i N_i - 2\gamma \sum_1^n z_i N_i - 2\sum_1^n N_i^2$$

(15.6)

Для соблюдения условия $\sum\limits_1^n E_i^2 = \min$ необходимо, чтобы сумма частных производных по α, β, γ, уравнения (15.6) обращалась в нуль, а следовательно, и каждая из этих частных производных должна обратиться в нуль:

$$\frac{\partial \sum\limits_1^n E_i^2}{\partial \alpha} = 0, \quad \frac{\partial \sum\limits_1^n E_i^2}{\partial \beta} = 0, \quad \frac{\partial \sum\limits_1^n E_i^2}{\partial \gamma} = 0 .$$

Тогда значения α, β, γ, которые обращают каждую из этих производных в нуль, будут представлять собой наиболее подходящие значения коэффициентов.

Дифференцируя уравнение (15.6) последовательно получим:

$$\left.\begin{aligned} &\alpha \sum x^2 + \beta \sum xy + \gamma \sum xz - \sum xN = 0 \\ &\beta \sum y^2 + \alpha \sum xy + \gamma \sum yz - \sum yN = 0 \\ &\gamma \sum z^2 + \alpha \sum xz + \beta \sum yz - \sum zN = 0 \end{aligned}\right\} \qquad (15.7)$$

Эти уравнения носят название «нормальных». Решая их относительно α, β, γ получим наилучшие значения коэффициентов, удовлетворяющих принципу наименьших квадратов.

Сравнивая уравнения (15.5) и (15.7), нетрудно заключить, что для получения нормального уравнения, например для α, необходимо все члены из условных уравнений (15.5) помножить на коэффициент при α и затем почленно сложить. Так же получаются нормальные уравнения для других коэффициентов. В связи с этим отпадает необходимость прибегать каждый раз к возведению в квадрат и дифференцированию уравнений, так как все можно свести к указанным арифметическим действиям, которые выполняются механически.

Если уравнение имеет вид линейной зависимости

$\bar{y} = a + bx,$

то для определения коэффициентов a и b по способу наименьших квадратов необходимо по аналогии с предыдущим составить по экспериментальным данным сумму:

$$\frac{\partial \sum_1^n E_i^2}{\partial \alpha} = 2\left[-\sum_1^n y_i + b\sum_1^n x_i + \sum_1^n a\right] = 0; \quad \sum_1^n E_i^2 = \sum_1^n (y_i - \bar{y})^2 = \sum_1^n (y_i - a - bx_i)^2 \qquad (15.8)$$

Возведя правую часть уравнения (15.8) в квадрат и взяв частные производные по a и b и приравняв их нулю, получим (15.9)

$$\left.\begin{aligned} &na + b\sum_1^n x_i = \sum_1^n y_i; \\ &a\sum_1^n x_i + b\sum_1^n x_i^2 + \sum_1^n x_i y_i. \end{aligned}\right\} \qquad (15.9)$$

Система уравнений (15.9) очень легко решается, предварительно заменить x_i на $x_i' = (x_i - \bar{X})$, где $\bar{X} = \frac{1}{n}\sum_1^n x_i$. Тогда исходное уравнение примет вид $\bar{y} = a' + b(x_i - \bar{X})$. Так как $\sum_1^n x_i' = 0$, то уравнения примут вид $n\alpha' = \sum_1^n y_i$;

$b\sum_1^n x_i'^2 = \sum_1^n x_i' y_i$, откуда $a' = \frac{\sum_1^n y_i}{n}$;

$$b = \frac{\sum_1^n x_i' y_i}{\sum_1^n x'^2} = \frac{\sum_1^n (x_i - \bar{X})}{\sum_1^n (x_i - \bar{X})^2}.$$

Для перехода от a' к а уравнения $\bar{y} = a + bx$ служит равенство $a = a' - b\bar{X}$.

ТЕМА 15. СТАТИСТИЧЕСКИЙ АНАЛИЗ ТОЧНОСТИ МЕХАНИЧЕСКОЙ ОБРАБОТКИ И СТАТИСТИЧЕСКОЕ РЕГУЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Лекция 16. Статистический анализ посредством больших выборок. Статистический анализ посредством малых выборок

Статистический анализ посредством больших выборок

Статистический анализ следует производить после того, как станок проработает некоторое время, необходимое для стабилизации температуры системы СПИД. Это время колеблется в пределах одного-двух часов. В результате этого погрешности обработки, вызываемые температурными деформациями элементов системы СПИД, превратятся из функциональных в постоянные и процесс обработки при нормальном его ходе будет характеризоваться точностной диаграммой I или IV типа.

Статистический анализ посредством большой выборки заключается в следующем.

Со станка берется большая (текущая) выборка, состоящая из деталей, изготовленных подряд одна за другой при неизменной настройке и других неизменных условиях. Объем выборки устанавливается в зависимости от желаемой точности и надежности определения меры рассеивания σ суммарной погрешности обработки. Для практических целей можно принять точность вычисления оценки σ по выборочному *s,* равную $\varepsilon = \pm 0{,}2s$ с вероятностью $\alpha = 0{,}95$. Тогда объем выборки достаточно сделать равным $n = 50$. Однако с увеличением n точность ε возрастает и поэтому часто принимают $n \geq 100$

Все детали выборки должны быть измерены шкальным измерительным инструментом с ценой деления измерительной шкалы,
равной $\left(\frac{1}{6} \div \frac{1}{10}\right) 2\delta$, где 2δ —допуск на измеряемый размер. На основании результатов измерений деталей выборки составляется таблица распределения размеров выборки. При составлении таблицы все наблюденные размеры разбиваются на интервалы, число которых выбирается так, чтобы ширина интервала была больше не менее чем в 2 раза цены деления шкалы измерительного инструмента. Это

делается для того, чтобы компенсировать погрешности измерения. Затем производится вычисление статистических характеристик выборки $\bar{x}$ и s, которые и принимаются в качестве оценок параметров $\overline{X_0}$ и σ_0 распределения генеральной совокупности, из которой взята выборка. После этого производится проверка гипотезы нормальности распределения. Выборку необходимо проверить также на случайность и убедиться в стабильности центра рассеивания погрешностей в процессе отбора пробы.

При положительных результатах проверки гипотез нормальности и случайности распределения выборки процесс может быть отнесен к IV типу точности, для которого суммарная погрешность обработки определяется по формуле

$$\Delta = \Delta_n + \Delta_c ;$$

где Δ_n — постоянные погрешности; Δ_c — случайные погрешности. Для определения суммарной величины случайных погрешностей во всей настроечной партии необходимо в качестве оценки σ_0 принять $\sigma = sz_2$, где s — среднее квадратическое отклонение выборки, а z_2 — коэффициент, определяемый в зависимости от объема выборки по таблице. Тогда

$$\Delta_c = 6z_2 s = 6\sigma \tag{16.1}$$

Фактическая величина постоянных погрешностей Δ_n или резерв допуска, приходящийся на долю постоянных погрешностей, определяется по следующим формулам:

а) для наружных поверхностей

$$\Delta_n = \overline{X} - 3\sigma - HO ; \tag{16.2}$$

б) для внутренних поверхностей

$$\Delta_n = BO - \overline{X} - 3\sigma ; \tag{16.3}$$

где BO и HO — верхнее, и нижнее предельные отклонения измеряемого размера с учетом их знаков;

$\overline{X}$ — среднее значение отклонений размеров от их номинала.

Для оценки точности процесса необходимо сравнить полученную суммарную погрешность Δ с допуском 2σ на размер детали. Точность процесса считается достаточной или избыточной, если удовлетворяется неравенство

$$\Delta \le 2\sigma .$$

Однако на практике возможен брак даже и при избыточной точности процесса, если настройка станка была выполнена с погрешностью, величина которой превышала допустимое значение.

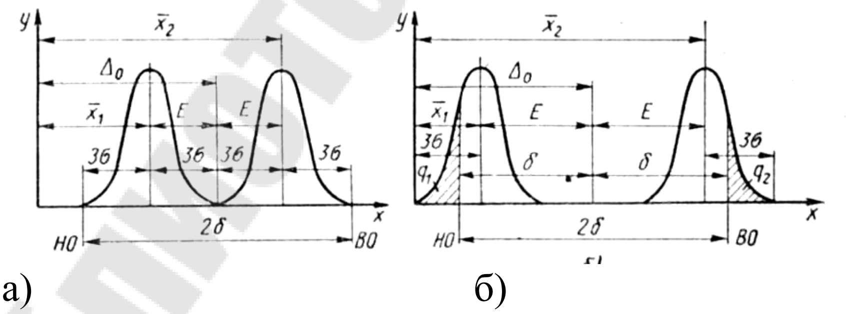
Обозначим через Δ_0 координату середины поля допуска относительно номинального значения размера, величина которой определяется по формуле

$$\Delta_0 = \frac{B0 + H0}{2},$$

где *B0* и *H0* — верхнее и нижнее предельные отклонения размера по чертежу с учетом знаков. Среднее значение действительных отклонений измеряемого размера от его номинала обозначим через $\overline{X}$. Величину смещения $\overline{X}$ от Δ_0 обозначим через *E*. Тогда

$$\mathrm{E} = \overline{X} - \Delta_0$$

На рис. 16.1 показаны два крайних положения кривой нормального распределения в поле допуска, когда смещение $\overline{X}$ от координаты середины поля допуска Δ_0 находится в пределах допустимых значений, и два других крайних положения кривой нормального распределения, когда $\overline{X}$ смещено относительно Δ_0 на величину, превышающую допустимое значение. В результате этого возникает брак, т. е. часть, деталей q_1 или q_2 будет иметь отклонения размеров, выходящие за пределы допуска.



а) б)

Рисунок 16.1 – Допустимые *(а)* и недопустимые (б) смещения центра рассеивания погрешности

Из рис. 16.1, *а* видно, что допускаемая погрешность настройки $\Delta_{н.д.}$ режущего инструмента на размер равна

$$\Delta_{н.д.} = \delta - 3\sigma \qquad (16.4)$$

Фактическая величина погрешности настройки $\Delta_{н.ф.}$ определится по формуле

$$\Delta_{н.ф} = \text{Е}_{ф} = \left|\overline{X} - \Delta_0\right|. \qquad (16.5)$$

Для работы без брака должно быть соблюдено неравенство

$$\Delta_{н.ф} \le \Delta_{н.д} \qquad (16.6)$$

Если это неравенство соблюдено не будет и $\Delta_{н.ф} \ge \Delta_{н.д}$ то при обработке настроечной партии неизбежен брак даже при избыточной точности процесса. Вероятный процент этого брака (рис. 16.1, *б)* можно определить по формуле

$$q = \left[0{,}5 - \Phi\left(\frac{\delta - \text{Е}_{ф}}{\sigma}\right)\right]100 \qquad (16.7)$$

Для сравнительной оценки точности аналогичных операций можно пользоваться коэффициентом точности K_T:

$$K_T = \frac{6\sigma}{2\delta} = \frac{3\sigma}{\delta} \qquad (16.8)$$

При $K_T \le 1$ точность процесса достаточная, а при $K_T > 1$ недостаточная. Для оценки точности настройки станка пользуются коэффициентом точности настройки *е:*

$$e = \frac{|\text{Е}|}{2\delta} \qquad (16.9)$$

При этом допустимое значение e:

$$e_д = \frac{2\delta - 6\sigma}{2\delta} = 1 - K_T \qquad (16.10)$$

Фактическое значение $e_ф$ определится по формуле

$$e_ф = \frac{\left|\overline{X} - \Delta_0\right|}{2\delta} \qquad (16.11)$$

Условия работы без брака выразятся неравенствами:

$K_T \leq 1$

$$e_{ф} \leq e_{д} \qquad (16.12)$$

Для оценки устойчивости процесса по большой выборке достаточно подтверждений гипотез нормальности и случайности выборки. Если эти гипотезы подтверждаются, то процесс можно считать устойчивым во времени.

Статистический анализ посредством малых выборок

Метод малых выборок имеет ряд преимуществ перед методом больших выборок. Основными преимуществами его являются, во-первых, уменьшение объема вычислительных работ, во-вторых, возможность следить за динамикой изменения точности процесса во времени, чего нельзя сделать с помощью метода больших выборок. Метод больших выборок может дать представление лишь о точности и устойчивости процесса в период взятии выборки, которые могут сохраниться и в дальнейшем, если после взятия выборки условия протекания процесса не изменяются. В действительности такой неизменности производственных условий заранее предвидеть нельзя. Например, при работе на прутковом автомате в течение смены производится несколько раз замена материала (смена прутка), смена инструмента в связи с износом, под-настройка станка и т. д., которые могут вносить значительные коррективы в полученные ранее параметры распределения. Метод малых выборок, если последние берут в течение всей смены регулярно через определенные промежутки времени, позволяет получить полную картину состояния процесса в течение исследуемого периода, выяснить степень его устойчивости, а также выявить причины недостаточной устойчивости процесса во времени, если она есть.

Статистический анализ малыми выборками производится следующим образом. Выборки объемом *n* — $5 \div 10$ шт. берутся через определенные фиксированные промежутки времени (например, через 15—30 *мин)*. Период времени для отбора проб устанавливается опытным путем и зависит от производительности станка, объема выборки и степени устойчивости технологического процесса.

Каждая выборка должна быть проверена на случайность по методу последовательных разностей, для каждой выборки нужно вычислить $\overline{X}$ и s^2. Далее необходимо для каждых двух смежных выборок проверить гипотезу однородности дисперсий выборок при помощи критерия *T*. Если гипотеза однородности дисперсий выборок подтверждается, то это свидетельствует о стабильности рассеивания или о том, что сравниваемые выборки взяты из одной и той же генеральной совокупности. При подтверждении гипотезы однородности дисперсий двух выборок следует проверить гипотезу однородности двух выборочных средних по критерию *t*-Стюдента.

Подтверждение гипотезы равенства двух смежных выборочных средних означает, что центр настройки станка не изменился в момент взятия данной выборки и остался таким, каким был при взятии предыдущей выборки, т. е. процесс находится в стабильном состоянии. Когда гипотеза равенства двух средних выборок не подтверждается, это свидетельствует о смещении центра настройки станка во время взятия данной выборки. Так как выборки берутся через определенные промежутки времени, то при обнаружении смещения центра настройки или изменения зоны рассеивания можно определить период времени, через который наступило нарушение стабильности процесса.

Обнаружив факт нарушения стабильности процесса, можно установить и область, в которой следует искать причину этого явления. Неоднородность выборочных дисперсий, свидетельствующая о нестабильности рассеивания, указывает на то, что причину этого следует искать в станке или установочно-зажимном приспособлении, или в механических свойствах обрабатываемого материала. Например, из-за недопустимого биения шпинделя станка, недопустимого биения зажимной цанги или зажимного патрона, резкого увеличения твердости материала в каком-либо прутке может значительно увеличиться и зона рассеивания размеров.

Неоднородность выборочных средних говорит о смещении центра настройки, а значит, причиной этого может служить либо износ режущего инструмента, либо ослабление крепления инструмента, либо другая причина, но связанная с настройкой инструмента на размер.

Таким образом, беря в течение смены через определенные интервалы времени малые выборки из текущей продукции станка, вычисляя средние и дисперсии выборок путем сравнения и оценки их расхождения при помощи критериев *T* и *t,* можно установить моменты разладок процесса и даже источники этих разладок. Устраняя причины разладок, можно привести его в состояние, когда рассеивание размеров в каждый момент времени будет носить более или менее стабильный характер. Что же касается центра рассеивания размеров, то можно добиться такого положения, что смещение его во времени будет зависеть главным образом от интенсивности размерного износа режущего инструмента и, следовательно, носить вполне закономерный характер. Другими словами, в результате статистического анализа процесса его привести в устойчивое состояние.

После того как процесс будет приведен в устойчивое состояние, следует установить, к какому типу точности он относится. Для этой цели необходимо параллельно с вычислением средних арифметических выборок строить кривую изменения средних для каждого номера выборки. По внешнему виду полученной кривой можно установить и тип точности изучаемого процесса. Если же процесс относится к IV типу, то для определения суммарной погрешности обработки необходимо определить $\overline{X}_0$ и σ_0 генеральной совокупности, из которой брались малые выборы, по следующим формулам:

$$\overline{X_0} \approx \frac{1}{m}\sum_{i=1}^{m}\overline{X_i}$$

где m — число выборок, и

$$\sigma_0 \approx z_2\overline{S}$$

где $\overline{S} = \sqrt{\frac{\sum s_i^2}{m}}$;

z_2 определяется по таблице; при этом в качестве n нужно принимать mn (n — объем выборки, m — число выборок).

При вычислении $\overline{X_0}$ и $\overline{S}$ необходимо резко выделяющиеся значения $\overline{X_i}$ и s_i^2, носящие совершенно случайный характер, исключать из суммы.

Для упрощения расчетов при вычислениях s_i^2 и $\overline{S}$ можно пользоваться размахами выборок $R = x_{max} - x_{mln}$. В математической статистике доказывается, что между средним значением размахов

$\overline{R} = \frac{1}{m}\sum_{i=1}^{m} R_i$ малых выборок ($n \leq$ 10) и $\overline{S}$ этих выборок существует следующая зависимость $\overline{S} = \frac{\overline{R}}{d_n}$

где d_n — коэффициент, зависящий от объема выборки n.

Доверительные границы для σ_0 генеральной совокупности, и которой извлекались выборки, определяются неравенством (16.13)

$$\frac{\overline{R}}{d_n} \cdot \frac{1}{1 + t\frac{\gamma_n}{\sqrt{m}}} \leq \sigma_0 \leq \frac{\overline{R}}{d_n} \cdot \frac{1}{1 - t\frac{\gamma_n}{\sqrt{m}}}$$

или

$$\alpha\frac{\overline{R}}{d_n} < \sigma_0 < \beta\frac{\overline{R}}{d_n}. \tag{16.14}$$

Для расчетов необходимо принять

$$\sigma_0 = \beta\frac{\overline{R}}{d_n}, \tag{16.15}$$

где

$$\beta = \frac{1}{1 - t\frac{\gamma_n}{\sqrt{m}}} \tag{16.16}$$

γ_n - коэффициент, зависящий от объема выборки n, значения приведены в табл. 16.1; m –число выборок; t – аргумент функции $2\Phi(t)=1-q$ (q –доверительный уровень значимости, обычно q=0,05).

Таблица 16.1 – Значения коэффициентов d_n, γ_n

n	4	5	6	7	8	9	10
d_n	2,059	2,326	2,534	2,704	2,847	2,970	3,037
γ_n	0,427	0,371	0,335	0,308	0,288	0,272	0,259

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Обработка экспериментальных данных по способу наименьших квадратов.
2. Когда следует статистический анализ?
3. Статистический анализ посредством больших выборок.
4. Статистический анализ посредством малых выборок.

Список использованных источников

1. Вентцель Е.С. Исследование операций: задачи, принципы, методология. -Москва: Наука, 1980. – 208 с.
2. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. -Москва: Радио и связь, 1983. – 416 с.
1. Вентцель Е.С., Овчаров Л..А Теория вероятностей и ее инженерные приложения. -Москва: , 1988. – 480 с.
2. Волков И. К., Загоруйко Е. А. Исследование операций: учеб. для вузов. – Москва: Изд-во МГТУ им. Н.Э. Баумана, 2000. – 436 с.
3. Климович Ф.Ф., Присевок А.Ф. Математическое моделирование технологических задач в машиностроении. Учебно-методическое пособие по лабораторным работам для студентов машиностроительных специальностей высших учебных заведений. – Мн.:БГПА, 2000. – 88с.
4. Кудрявцев Е.М. Исследование операций в задачах, алгоритмах программах. – Москва: Радио и связь, 1984. – 184 с.
5. Кузнецов А.В., Сакович В.А., Холод Н.И. Высшая математика. Математическое программирование: учеб. пособие для эконом. спец. вузов. – Минск: Выш. шк., 1984. – 221 с.
6. Кузнецов А.В., Холод Н.И. Математическое программирование: учеб. – 2-ое изд., перераб. и доп. – Минск: Выш. шк., 2001. – 351 с.
7. Кузнецов А.В., Холод Н., Костевич Л.С. Руководство к решению задач по математическому программированию: учеб. пособие.– 2-ое изд., перераб. и доп. – Минск: Выш. шк., 2001. – 448 с.
3. Кузнецов О.П., Адельсон-Вельский Г.М. Дискретная математика для инженера. -Москва: Энергия, 1980. – 344 с.
8. Кузнецов Ю. И., Кузубов В. И, ВолощенкоА.В. Математическое программирование: учеб. пособие. – 2-е изд., перераб. и доп. – Москва: Высш. школа, 1980. – 300 с.
9. Мурашко В.С. Практическое пособие к выполнению лабораторных работ по курсу «Математическое моделирование технологических задач в машиностроении» для студентов спец. Т03.01.01 – «Технология машиностроения». – Гомель: ГГТУ им. П.О. Сухого, 1999. – 60с. (М/у №2416)
10. Мурашко В.С. Оптимизация режимов резания. Практическое пособие по курсу «Математическое моделирование технологических задач в машиностроении» для студентов спец. Т.03.01.000 «Технология машиностроения» очной и заочной формы обучения». – Гомель: ГГТУ им. П.О. Сухого, 2003. – 36с. (М/у №2866).

11. Палай И.А. Линейное программирование: учеб. пособие. – Москва: ЭКСМО, 2008. – 256 с.
12. Просветов Г. И. Дискретная математика: задачи и решения: учеб. пособие. – Москва: БИНОМ. Лаборатория знаний, 2008. – 222 с.
13. Пляскин И.И. Оптимизация технических решений в машиностроении . -Москва: Машиностроение, 1982. – 176 с.
14. Сборник задач и упражнений по высшей математике: Математическое программирование: учеб. пособие. / Под общ. ред. Кузнецова А.В., Рутковского Р.А. – 2-ое изд., перераб. и доп. – Минск: Выш. шк., 2002. – 447 с.
15. Солонин И. С. Математическая статистика в технологии машиностроения. – Москва: «Машиностроение», 1972. – 216с.
16. Тарасик В.П. Математическое моделирование технических систем: учебник для вузов.- Минск: Дизайн ПРО, 1997. -640 с.
4. Таха Х. Введение в исследование операций. – Москва: Мир, 1985. – Том 1, 479с.
17. Таха Х. Введение в исследование операций. – Москва: Мир, 1985. – Том 2, 496 с.
18. Тимковский В.Г. Дискретная математика в мире станков и деталей – Москва: Наука, 1992 – 145 с.
19. Тимковский В.Г. Дискретная математика в мире станков и деталей – М.: Наука, 1992 – 145 с.
20. Щербаков С.А. Моделирование решений технологических задач. Учебное пособие по курсу «Основы математическое моделирования» для студентов специальности 12.01- «Технология машиностроения». – Гомель: ГГТУ им. П.О. Сухого, 1995. – 66с. (М/у №1948)