

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ В ОТРАСЛИ**  
**(Конспект Лекций)**

**Оглавление**

ВВЕДЕНИЕ .....	2
ЛЕКЦИЯ 1 ПОНЯТИЕ О МОДЕЛИРОВАНИИ. МОДЕЛЬ.....	4
ЛЕКЦИЯ 2 КЛАССИФИКАЦИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ. ТРЕБОВАНИЯ, ПРЕДЪЯВЛЯЕМЫЕ К МАТЕМАТИЧЕСКИМ МОДЕЛЯМ .....	12
ЛЕКЦИЯ 3. ОСНОВЫ ТЕОРИИ МНОЖЕСТВ И ТЕОРИИ ГРАФОВ.....	15
ЛЕКЦИЯ 4. ОБЩАЯ ПОСТАНОВКА И ВИДЫ ЗАДАЧ ПРИНЯТИЯ РЕШЕНИЙ.....	18
ЛЕКЦИЯ 5. ГРАФО – АНАЛИТИЧЕСКИЙ МЕТОД РЕШЕНИЯ ЗАДАЧ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ .....	21
ЛЕКЦИЯ 6. МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ.....	25
ЛЕКЦИЯ 7. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ (ПОИСК ЭКСТРЕМУМА ФУНКЦИИ ОДНОЙ ПЕРЕМЕННОЙ).....	31
ЛЕКЦИЯ 8. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ (ПОИСК ЭКСТРЕМУМА ФУНКЦИИ N – ПЕРЕМЕННЫХ).....	35
ЛЕКЦИЯ 9. МЕТОДЫ РЕШЕНИЯ МНОГОКРИТЕРИАЛЬНЫХ ЗАДАЧ ОПТИМИЗАЦИИ.....	38
ЛЕКЦИЯ 10 МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СИЛОВОГО ВЗАИМОДЕЙСТВИЯ В ЗОНЕ РЕЗАНИЯ ПРИ ИЗГОТОВЛЕНИИ ДЕТАЛЕЙ НА СТАНКАХ.....	43
ЛЕКЦИЯ 11. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ УПРУГИХ ДЕФОРМАЦИЙ В ТЕХНОЛОГИЧЕСКОЙ СИСТЕМЕ .....	47
ЛЕКЦИЯ 12. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТОЧНОСТИ ОБРАБОТКИ ДЕТАЛЕЙ НА СТАНКАХ .....	51
ЛЕКЦИЯ 13. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ УПРАВЛЕНИЯ ПРОИЗВОДИТЕЛЬНОСТЬЮ, СЕБЕСТОИМОСТЬЮ И ТОЧНОСТЬЮ ОБРАБОТКИ ДЕТАЛЕЙ НА МЕТАЛЛОРЕЖУЩИХ СТАНКАХ .....	56
ЛЕКЦИЯ 14 ОБЪЕМНОЕ ПЛАНИРОВАНИЕ РАБОТЫ ТЕХНОЛОГИЧЕСКИХ СТАНОЧНЫХ СИСТЕМ .....	61
ЛЕКЦИЯ 15. ОСНОВЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ .....	66
ЛЕКЦИЯ 16. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ПРОСТЕЙШИХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ.....	72
ЛЕКЦИЯ 17 ОСНОВЫ ТЕОРИИ ПРОИЗВОДИТЕЛЬНОСТИ И НАДЕЖНОСТИ АВТОМАТИЧЕСКИХ И АВТОМАТИЗИРОВАННЫХ СТАНОЧНЫХ СИСТЕМ .....	77
ЛЕКЦИЯ 18. ПРОИЗВОДИТЕЛЬНОСТЬ И НАДЕЖНОСТЬ АВТОМАТИЧЕСКИХ И АВТОМАТИЗИРОВАННЫХ СТАНОЧНЫХ СИСТЕМ .....	81
ЛЕКЦИЯ 19. ОПЕРАТИВНО – КАЛЕНДАРНОЕ ПЛАНИРОВАНИЕ В ТЕХНОЛОГИЧЕСКИХ СИСТЕМАХ НА ОСНОВЕ ТЕОРИИ РАСПИСАНИЙ.....	86
ЛЕКЦИЯ 20. ЛОГИЧЕСКИЕ МОДЕЛИ ПРОЦЕССОВ И ОБЪЕКТОВ .....	91
ЛЕКЦИЯ 21. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И ОБЪЕКТОВ .....	112
ЛЕКЦИЯ 22. ОДНОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ.....	121
ЛЕКЦИЯ 23. МНОГОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ.....	124
ЛЕКЦИЯ 24. МЕТОДИКА ФАКТОРНОГО ЭКСПЕРИМЕНТА.....	127

## Введение

**Наименование курса** – «Математическое моделирование в отрасли».

**Цель курса** – приобретение навыков разработки и использования математических моделей для описания, исследования и оптимизации процессов в машиностроении.

**Задачи курса** 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.

### **Литература, рекомендуемая для изучения курса:**

1. <http://ispu.ru/library/lessons/Koposov/4.html>
2. Системы автоматизированного проектирования: Учеб. пособие для вузов: В 9 кн./ И.П.Норенков. Кн.1. Принципы построения и структура.- М.: Высшая школа, 1986.- 127 с.
3. Горбатов В.А. Основы дискретной математики: Учеб. пособие для студентов вузов – М.: Высшая школа, 1986.- 311 с.
4. Корячко В.П. и др. Теоретические основы САПР: Учебник для вузов/В.П.Корячко,В.М.Курейчик,И.П.Норенков.-М.: Энергоатомиздат, 1987.- 400 с.
5. Ногин В.Д., Протоdjяконов И.О., Евлампиев И.И. Основы теории оптимизации: Учеб. пособие для студентов вузов/ Под ред. И.О.Протоdjяконова, - М.: Высшая школа,1986. – 384 с.
6. Дегтярев Ю.И. Исследование операций: Учеб. пособие для вузов по специальности АСУ.- М.: Высшая школа, 1986. – 320 с.
7. Вентцель Е.С. Исследование операций: задачи, принципы, методология. – 2-е изд. – М.: Наука, 1988. – 208 с.
8. Банди Б. Методы оптимизации. Вводный курс: Пер. с англ. – М.: Радио и связь, 1988. – 128 с.
9. Бобров В.Ф. Основы теории резания металлов. – М.: Машиностроение, 1975. – 344 с.
10. Грановский Г.И., Грановский В.Г. Резание металлов: Учебник для машиностроительных и приборостроительных спец. вузов. – М.: Высшая школа, 1985. – 304 с.
11. Кован В.М. Основы технологии машиностроения. – М.: Машгиз, 1959.-496 с.
12. Колесов И.М. Основы технологии машиностроения: Учебник для машиностроительных вузов. – М.: Машиностроение, 1997. – 592 с.
13. Марголит Р.Б. Эксплуатация и наладка станков с программным управлением и промышленных роботов: Учеб. пособие для машиностроительных техникумов. – М.: Машиностроение, 1991.-592 с.
14. Проников А.С. Расчет и конструирование металлорежущих станков. – М.: Высшая школа, 1968. – 431 с.
15. Маталин А.А. Технология механической обработки. – Л.: Машиностроение (Ленинградское отделение), 1977. – 464 с.
16. Проектирование технологии: Учебник для студентов машиностроительных специальностей вузов/ И.М. Баранчукова, А.А.Гусев, Ю.Б.Крамаренко и др.; Под общ. ред. Ю.М.Соломенцева. – М.: Машиностроение, 1990. – 416 с.
17. Васильев Г.Н. Автоматизация проектирования металлорежущих станков. – М.: Машиностроение, 1987. – 280 с.
18. Танаев В.С., Сотсков Ю.Н., Струевич В.А. Теория расписаний. Многостадийные системы. – М.: Наука, 1989. – 328 с.
19. Чудаков А.Д., Фалевич Б.Я. Автоматизированное оперативно-календарное планирование в гибких комплексах механообработки. – М.: Машиностроение, 1986. – 224 с.

- 20.Ивченко Г.И., Каштанов В.А., Коваленко И.Н. Теория массового обслуживания: Учеб. пособие для вузов. – М.: Высшая школа, 1982. – 256с.
- 21.Дащенко А.И., Белоусов А.П. Проектирование автоматической линий: Учеб. пособие для машиностроительных специальностей вузов. – М.: Высшая школа, 1983. – 328 с.
- 22.Васильев В.Н. Организация, управление и экономика гибкого интегрированного производства в машиностроении. – М.: Машиностроение, 1986. – 312 с.

### **Список понятий, знание которых необходимо на момент начала изучения курса**

Для успешного освоения курса необходимо знание следующих основных понятий и тем из приведенных ниже дисциплин:

1. Математика:
  - основы теории множеств и теории графов;
  - основы теории оптимизации;
  - теория вероятностей.
2. Экономика и планирование:
  - методы расчета себестоимости изделий.
3. Технология машиностроения:
  - основы изготовления различных деталей.
4. Технологические машины и оборудование в нефтегазовом комплексе:
  - конструкции машин и оборудования НГП.
8. Проектирование в машиностроительном производстве:
  - виды и конструкции машин в НГП;
  - понятия надежности и производительности машин НГП.

## **Лекция 1 Понятие о моделировании. Модель.**

**Моделирование** – это метод исследования сложных объектов, явлений или процессов на их моделях или на реальных установках с применением методов теории подобия при постановке и обработке результатов эксперимента.

**Модель** (от лат. Modulus –мера, образец, норма) - физическая система (устройство, схема, установка, система машин) или математическое описание компонентов и функций, отображающие существенные свойства какого-либо объекта, процесса или явления.

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

С древнейших времен при изучении сложных процессов, явлений при конструировании новых устройств и сооружений человек применяет модели. Хорошо построенная модель, как правило, доступнее для исследования, нежели реальный объект. Кроме того, некоторые объекты вообще не могут быть исследованы непосредственным образом: недопустимы, например, эксперименты с экономикой страны в познавательных целях, неосуществимы эксперименты с планетами Солнечной системы и т.п.

Важное назначение модели состоит в том, что с ее помощью выявляются наиболее существенные факторы, формирующие те или иные свойства объекта, поскольку сама модель отражает лишь некоторые основные характеристики исходного объекта.

Модель позволяет также научиться правильно управлять объектом, проверяя различные варианты управления на модели этого объекта. Экспериментировать в этих целях с реальным объектом в лучшем случае бывает неудобно, а зачастую просто вредно, а порой невозможно в силу ряда причин.

Таким образом, модель нужна для того, чтобы понять, как устроен конкретный объект, какова его структура, основные свойства, законы развития и взаимодействия с окружающим миром, а также чтобы научиться управлять объектом или процессом.

Хорошо построенная модель, как правило, обладает удивительным свойством: ее изучение дает некоторые новые знания об объекте оригинале.

**Процесс построения модели называется моделированием.**

Существует несколько приемов моделирования, которые можно изобразить в виде схемы.

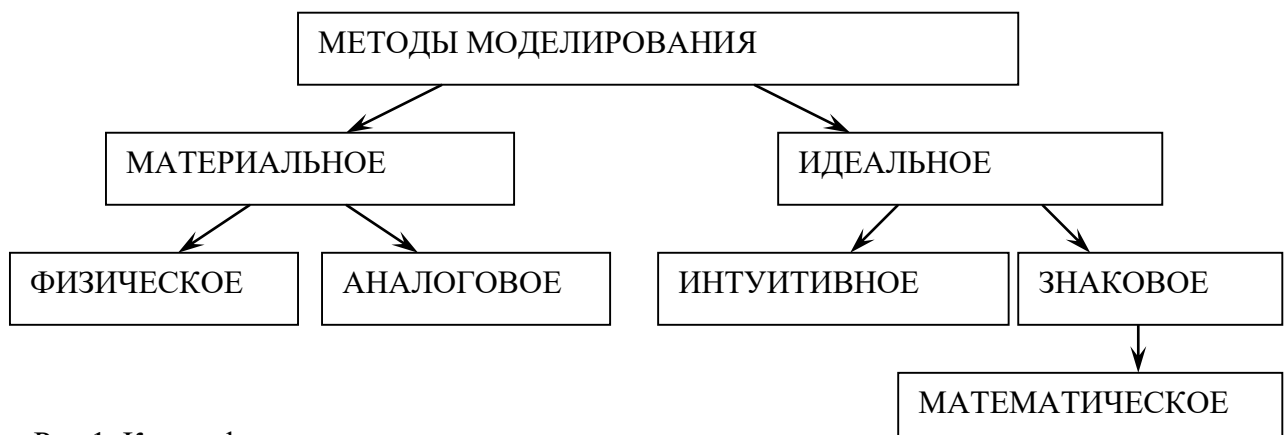


Рис.1. Классификация методов моделирования.

**К материальным** относятся такие способы моделирования, при которых исследование ведется на основе модели, воспроизводящей основные геометрические, физические, динамические и функциональные характеристики изучаемого объекта. Основными разновидностями материального моделирования являются физические и аналоговые моделирования.

**Физическим** принято называть моделирование, при котором реальному объекту противопоставляется его увеличенная или уменьшенная копия, допускающая исследование (как правило, в

лабораторных условиях) с помощью последующего перенесения свойств изучаемых процессов и явлений с модели на объект на основе теории подобия. Вот несколько примеров физических моделей: в астрономии – планетарий, в архитектуре – макеты зданий, в самолетостроении – модели летательных аппаратов.

**Аналоговое** моделирование основано на аналогии процессов и явлений, имеющих различную физическую природу, но одинаково описываемых формально (одними и теми же математическими уравнениями). Наиболее простой пример – изучение механических колебаний с помощью электрической схемы, описываемой теми же дифференциальными уравнениями.

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

**Под интуитивным** понимают моделирование, основанное на интуитивном представлении об объекте исследования не поддающимся формализации, либо не нуждающемся в ней. Примером интуитивной модели окружающего мира можно считать жизненный опыт каждого человека.

**Знаковым** называется моделирование, использующее в качестве моделей знаковые преобразования какого-либо вида: схемы, графики, чертежи, формулы, наборы символов и т.д., а также включающее совокупность законов по которым можно оперировать с выбранными знаковыми преобразованиями и их элементами.

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

**Процесс математического моделирования можно подразделить на 4 этапа.**

**Первый этап** – формулирование законов, связывающих основные объекты модели. Этот этап завершается записью в математических терминах сформулированных качественных представлений о связях между объектами модели.

**Второй этап** – исследование математических задач, к которым приводят математические модели. Основным вопросом здесь является решение прямой задачи, т. е. получение в результате анализа модели выходных данных для дальнейшего их сопоставления с результатами наблюдения изучаемых явлений. Здесь главную роль приобретают математический аппарат и вычислительная техника.

**Третий этап** – выяснение того, удовлетворяет ли принятая модель критерию практики, согласуются ли результаты наблюдений с теоретическими следствиями модели в пределах точности наблюдений.

**Четвертый этап** – последующий анализ модели в связи с накоплением данных об изучаемых явлениях. Возникает возможность модернизации модели, т.е. построения новой, более совершенной математической модели.

## **Технология подготовки и решения задач с помощью компьютера**

(По материалам сайта [http://www.aec.neva.ru/book/informatika/theory/chapter8/1\\_8\\_2.html](http://www.aec.neva.ru/book/informatika/theory/chapter8/1_8_2.html))

Итак, что называют математической моделью?

**Математическая модель — это система математических соотношений — формул, уравнений, неравенств и т.д., отражающих существенные свойства объекта или явления.**

Чтобы описать явление, необходимо выявить самые существенные его свойства, закономерности, внутренние связи, роль отдельных характеристик явления. Выделив наиболее важные факторы, можно пренебречь менее существенными.

Наиболее эффективно математическую модель можно реализовать на компьютере в виде алгоритмической модели — так называемого "вычислительного эксперимента".

Конечно, результаты вычислительного эксперимента могут оказаться и не соответствующими действительности, если в модели не будут учтены какие-то важные стороны действительности.

Итак, создавая математическую модель для решения задачи, нужно:

- **выделить предположения, на которых будет основываться математическая модель;**
- **определить, что считать исходными данными и результатами;**
- **записать математические соотношения, связывающие результаты с исходными данными.**

При построении математических моделей далеко не всегда удается найти формулы, явно выражающие искомые величины через данные. В таких случаях используются математические методы, позволяющие дать ответы той или иной степени точности.

Существует не только математическое моделирование какого-либо явления, но и визуально-натурное моделирование, которое обеспечивается за счет отображения этих явлений средствами машинной графики, т.е. перед исследователем демонстрируется своеобразный "компьютерный мультфильм", снимаемый в реальном масштабе времени. Наглядность здесь очень высока.

### **Математическое моделирование процессов в машиностроении**

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.

В чистом виде (по – отдельности) математические модели данных объектов применяются редко, как правило, они комбинированные. Например, в математических моделях ТС используются математические модели ТП, в которых, в свою очередь, применяются математические модели ФП.

Математическая модель объекта моделирования – это система математических элементов (чисел, переменных, уравнений, неравенств, множеств, матриц, графов и т.д.) и отношений между ними, адекватно отражающая некоторые свойства объекта, существенные с точки зрения инженера, для решения той или иной задачи.

Математические модели разрабатываются для:

1. Описания ФП, ТП, ТС.(физических, технологических процессов, технологических систем)
2. Исследования ФП, ТП, ТС.
3. Проектирования ТП, ТС.
4. Оптимизации в ходе проектирования ТП, ТС и организации работы ТС.
5. Построения систем автоматизированного проектирования.

### **Методы оптимизации.**

(По материалам сайта [http://dl/sumdu.edu/ua/](http://dl.sumdu.edu/ua/))

Процесс оптимизации лежит в основе всей инженерной деятельности, поскольку функции специалиста состоят в том, чтобы, с одной стороны, проектировать новые, более эффективные, менее дорогие технические системы, а, с другой стороны, разрабатывать методы повышения качества функционирования существующих систем.

В практической деятельности часто из многих возможных вариантов решений задачи необходимо выбрать оптимальный. Например, из нескольких вариантов перевозки сырья потребителям необходимо выбрать наиболее дешевый, но такой, который учитывает ограничения на допустимые термины поставок; из возможных планов раскроя материала выбрать такой, который позволит выполнить план при наименьшем количестве отходов и т.д.

Во множестве случаев задача поиска оптимального решения может быть формализована и решена точно или приблизительно известными методами.

### **Постановка задачи оптимизации.**

Оптимизация в широком смысле слова находит применение в науке, технике и в любой другой области человеческой деятельности. Оптимизация - целенаправленная деятельность, заключающаяся в получении наилучших результатов при соответствующих условиях.

Поиски оптимальных решений привели к созданию специальных математических методов и уже в 18 веке были заложены математические основы оптимизации (вариационное исчисление, численные методы и др.).

Однако до второй половины 20 века методы оптимизации во многих областях науки и техники применялись очень редко, поскольку практическое использование математических методов оптимизации требовало огромной вычислительной работы, которую без ЭВМ реализовать было крайне трудно, а в ряде случаев - невозможно. Особенно большие трудности возникали при решении задач оптимизации процессов в химической технологии из-за большого числа параметров и их сложной взаимосвязи между собой. При наличии ЭВМ задача заметно упрощается.

Постановка задачи оптимизации предполагает существование конкурирующих свойств процесса, например,

- количество продукции - "расход сырья" - количество продукции "качество продукции"

Выбор компромиссного варианта для указанных свойств и представляет собой процедуру решения оптимизационной задачи.

При постановке задачи оптимизации необходимо:

1. Наличие объекта оптимизации и цели оптимизации. При этом формулировка каждой задачи оптимизации должна требовать экстремального значения лишь одной величины, т.е. одновременно системе не должно приписываться два и более критериев оптимизации, т.к. практически всегда экстремум одного критерия не соответствует экстремуму другого.

Типичный пример неправильной постановки задачи оптимизации: "Получить максимальную производительность при минимальной себестоимости". Ошибка заключается в том, что ставится задача поиска оптимума 2-х величин, противоречащих друг другу по своей сути. Правильная постановка задачи могла быть следующей: а) получить максимальную производительность при заданной себестоимости; б) получить минимальную себестоимость при заданной производительности; В первом случае критерий оптимизации - производительность, а во втором - себестоимость.

2. Наличие ресурсов оптимизации, под которыми понимают возможность выбора значений некоторых параметров оптимизируемого объекта. Объект должен обладать определенными степенями свободы - управляющими воздействиями.

3. Возможность количественной оценки оптимизируемой величины, поскольку только в этом случае можно сравнивать эффекты от выбора тех или иных управляющих воздействий.

4. Учет ограничений. **ФОРМУЛИРОВКА МАТЕМАТИЧЕСКОЙ ЗАДАЧИ ОПТИМИЗАЦИИ**

В достаточно общем виде математическую задачу оптимизации можно сформулировать следующим образом: ***Минимизировать (максимизировать) целевую функцию с учетом ограничений на управляемые переменные.***

Под *минимизацией (максимизацией)* функции  $n$  переменных  $f(x) = f(x_1, \dots, x_n)$  на заданном множестве  $U$   $n$ -мерного векторного пространства  $E_n$  понимается определение хотя бы одной из точек минимума (максимума) этой функции на множестве  $U$ , а также, если это необходимо, и минимального (максимального) на  $U$  значения  $f(x)$ .

При записи математических задач оптимизации в общем виде обычно используется следующая символика:

$$f(x) \rightarrow \min (\max), x \text{ принадлежит } U,$$

где  $f(x)$  - целевая функция, а  $U$  - допустимое множество, заданное ограничениями на управляемые переменные.

### **ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ**

Задачи одномерной минимизации представляют собой простейшую математическую модель оптимизации, в которой целевая функция зависит от одной переменной, а допустимым множеством является отрезок вещественной оси:

$$f(x) \rightarrow \min, x \text{ принадлежит } [a, b].$$

Максимизация целевой функции эквивалентна минимизации ( $f(x) \rightarrow \max$ ) эквивалентна минимизации противоположной величины ( $-f(x) \rightarrow \min$ ), поэтому, не умаляя общности можно рассматривать только задачи минимизации.

К математическим задачам одномерной минимизации приводят прикладные задачи оптимизации с одной управляемой переменной. Кроме того, необходимость в минимизации функций одной



переменной возникает при реализации некоторых методов решения более сложных задач оптимизации.

Для решения задачи минимизации функции  $f(x)$  на отрезке  $[a, b]$  на практике, как правило, применяют приближенные методы. Они позволяют найти решения этой задачи с необходимой точностью в результате определения конечного числа значений функции  $f(x)$  и ее производных в некоторых точках отрезка  $[a, b]$ . Методы, использующие только значения функции и не требующие вычисления ее производных, называются прямыми методами минимизации.

Большим достоинством прямых методов является то, что от целевой функции не требуется дифференцируемости и, более того, она может быть не задана в аналитическом виде. Единственное, на чем основаны алгоритмы прямых методов минимизации, это возможность определения значений  $f(x)$  в заданных точках.

Рассмотрим наиболее распространенные на практике прямые методы поиска точки минимума. Самым слабым требованием на функцию  $f(x)$ , позволяющим использовать эти методы, является ее унимодальность. Поэтому далее будем считать функцию  $f(x)$  унимодальной на отрезке  $[a, b]$ .

*Метод перебора*

*Метод поразрядного поиска*

*Метод деления пополам*

*Метод золотого сечения*

## **МЕТОДЫ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ**

Задача безусловной оптимизации состоит в нахождении минимума или максимума функции в отсутствие каких-либо ограничений. Несмотря на то что большинство практических задач оптимизации содержит ограничения, изучение методов безусловной оптимизации важно с нескольких точек зрения. Многие алгоритмы решения задачи с ограничениями предполагают сведение ее к последовательности задач безусловной оптимизации.

Другой класс методов основан на поиске подходящего направления и последующей минимизации вдоль этого направления. Обоснование методов безусловной оптимизации может быть естественным образом распространено на обоснование процедур решения задач с ограничениями.

Алгоритмы безусловной минимизации функций многих переменных можно сравнивать и исследовать как с теоретической, так и с экспериментальной точек зрения.

Первый подход может быть реализован полностью только для весьма ограниченного класса задач, например, для сильно выпуклых квадратичных функций. При этом возможен широкий спектр результатов от получения бесконечной минимизирующей последовательности в методе циклического покоординатного спуска до сходимости не более чем за  $n$  итераций в методе сопряженных направлений.

Мощным инструментом теоретического исследования алгоритмов являются теоремы о сходимости методов. Однако, как правило, формулировки таких теорем абстрактны, при их доказательстве используется аппарат современного функционального анализа. Кроме того, зачастую непросто установить связь полученных математических результатов с практикой вычислений. Дело в том, что условия теорем трудно проверяемы в конкретных задачах, сам факт сходимости мало что дает, а оценки скорости сходимости неточны и неэффективны. При реализации алгоритмов также возникает много дополнительных обстоятельств, строгий учет которых невозможен (ошибки округления, приближенное решение различных вспомогательных задач и т.д.) и которые могут сильно повлиять на ход процесса.

Поэтому на практике часто сравнение алгоритмов проводят с помощью вычислительных экспериментов при решении так называемых специальных тестовых задач. Эти задачи могут быть как с малым, так и с большим числом переменных, иметь различный вид нелинейности.

Они могут быть составлены специально и возникать из практических приложений, например задача минимизации суммы квадратов, решение систем нелинейных уравнений и т.п.

Вид, состав, сложность математической модели зависит от того, какой объект она описывает и для каких целей разработана.

**Пример 1.** Пусть требуется описать и исследовать процесс резания (точения) с точки зрения действующих сил, которые представим известными в теории резания уравнениями:



$$P_X = C_{PX} t^{X_{PX}} s^{Y_{PX}} K_{ОЩ};$$

$$P_Y = C_{PY} t^{X_{PY}} s^{Y_{PY}} K_{ОЩ};$$

$$P_Z = C_{PZ} t^{X_{PZ}} s^{Y_{PZ}} K_{ОЩ}.$$

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  $X_p \cong 1, Y_p \cong 0,75$ );  $K_{ОЩ} = K_M K_{ИХ...}$  – 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.2)

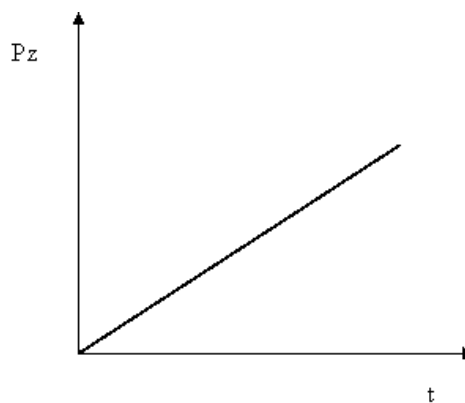


Рис.2. Зависимость силы резания  $P_z$  от глубины резания  $t$

Использовать математическую модель процесса резания в таком виде, например, для проведения оптимизации режимов резания нельзя, она слишком проста. Для оптимизации математическая модель должна иметь следующие элементы:

1. Критерий или критерии оптимизации.
2. Целевую функцию.
3. Систему ограничений.
4. Систему уравнений, описывающих объект.
5. Входные, выходные и внутренние параметры.
6. Управляемый (варьируемый) или управляемые (варьируемые) параметры, которые выделяются из числа внутренних параметров.

**Пример 2.** Однопараметрическая однокритериальная оптимизация режимов резания. Пусть требуется определить оптимальную подачу  $s_{опт}$ , при этом глубина резания постоянна ( $t = const$ ).

1. Критерий оптимизации - сила резания  $P_Z$ .
2. Целевая функция – зависимость между критерием (критериями) оптимизации и подлежащими оптимизации параметрами с указанием направления (вида) экстремума. В нашем примере она имеет следующий вид:

$$P_Z = C_{PZ} t^{X_{PZ}} s^{Y_{PZ}} K_{ОЩ} \rightarrow \max.$$

3. Система ограничений включает в себя лишь одно из них:

$$0,07 \leq s \leq 4,16, \text{ мм/об.}$$

4. Система уравнений. Математическая модель в нашем примере включает в себя одно уравнение

$$P_Z = C_{PZ} t^{X_{PZ}} s^{Y_{PZ}} K_{ОЩ}.$$

5. Входные параметры –  $C_{PZ}, X_{PZ}, Y_{PZ}, K_{ОБЦ}$ ,

- выходной параметр –  $P_Z$ ,
- внутренние параметры –  $t, s$ .

6. Управляемый (варьируемый) параметр –  $s$ .

Задача оптимизации решается в два этапа (см. рис.3):

Этап 1. Определение области допустимости решений (работаем с ограничениями) – ОДР.

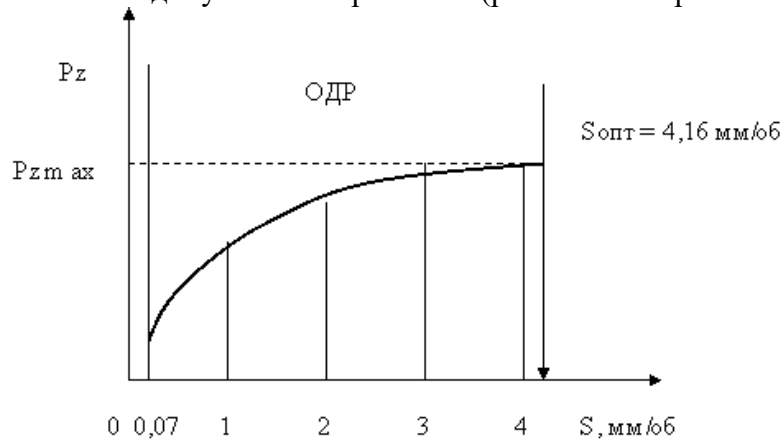


Рис. 3 Графическая иллюстрация решения задачи оптимизации

Этап 2. Определение оптимального параметра –  $s_{ОПТ}$ .

В общем виде математическая модель объекта записывается:

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

где  $Y$  – вектор выходных параметров,  $Y = (y_1, y_2, \dots, y_m)$ ;

$X$  – вектор внутренних параметров,  $X = (x_1, x_2, \dots, x_n)$ ;

$Q$  – вектор внешних (входных) параметров,  $Q = (q_1, q_2, \dots, q_p)$ .

Для приведенной выше математической модели процесса резания:

$$Y = (P_X, P_Y, P_Z);$$

$$X = (t, s);$$

$$Q = (C_{PX}, C_{PY}, C_{PZ}, X_{PX}, X_{PY}, X_{PZ}, Y_{PX}, Y_{PY}, Y_{PZ}, K_{ОБЦ}).$$

Структурная схема объекта моделирования представлена на рис.4 .

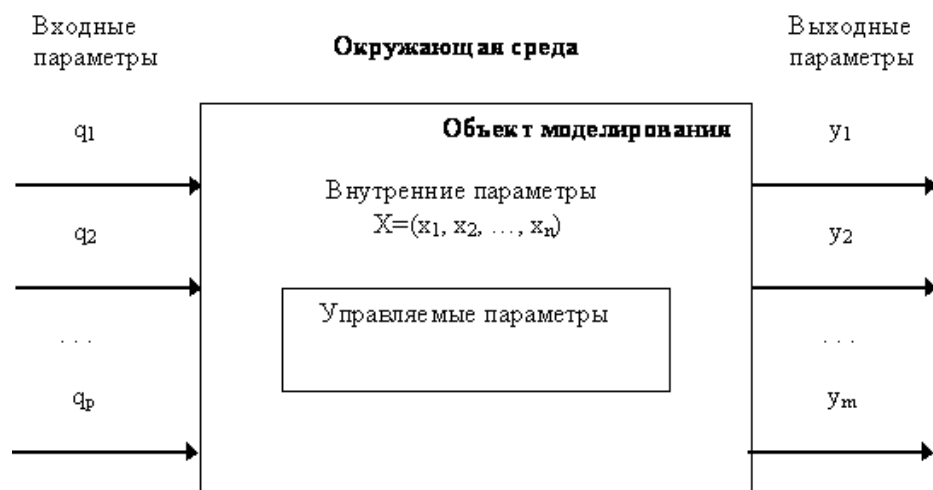


Рис.4. Структурная схема объекта моделирования

Следует иметь в виду, что любой объект не может существовать обособленно, без взаимосвязи с другими объектами (окружающей средой). Для процесса резания элементами окружающей среды являются обрабатываемый и инструментальный материалы, смазочно-охлаждающая жидкость, образующаяся в результате обработки стружка и т.д. Для технологического процесса в ходе его реализации- это технологическое оборудование, режущий инструмент, элементы приспособления и т.д.

Для технологической системы (механического участка из универсальных станков или станков с ЧПУ, автоматической линии и др.) элементами окружающей среды являются другие технологические системы, цеховой транспорт и т.д.

### **Литература**

1. Васильев Ф.П. Численные методы решения экстремальных задач. - М.: Наука, 1980.
2. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации. - М.: Наука, 1986.
3. Поляк Б.Т. Введение в оптимизацию. - М.: Наука, 1983.
4. Сеа Ж. Оптимизация. Теория и алгоритмы. - М.: Мир, 1973.
5. Зангвилл У. Нелинейное программирование. Единый подход. - М.: Сов. радио, 1973.
6. Банди Б. Методы оптимизации (вводный курс). - М.: Радио и связь, 1988.
7. Компьютерное методическое пособие по методам параметрической оптимизации. МГТУ им. Баумана, 1997г.

### **Вопросы для самоконтроля.**

1. Дайте определение понятия “модель”.
2. Что такое моделирование?
3. Классификация способов моделирования.
4. Определение математического моделирования.
5. Каковы основные этапы моделирования?

### **Тест-контроль знаний**

1. Модель – это:
  - \* Физическая система (устройство, схема, установка, система машин) или математическое описание компонентов и функций, отображающие существенные свойства какого-либо объекта, процесса или явления.
  - \* Физическая система (устройство, схема, установка, система машин)
  - \* Математическое описание компонентов и функций, отображающие существенные свойства какого-либо объекта, процесса или явления.
2. Знаковое моделирование – это:
  - \* Частный случай интуитивного моделирования.
  - \* Моделирование, использующее в качестве моделей знаковые преобразования какого-либо вида: схемы, графики, чертежи, формулы, наборы символов и т.д.
3. Описание и исследование основных законов механики И. Ньютона выполнено при помощи -
  - \* Знакового моделирования.
  - \* Интуитивного моделирования.
  - \* Математического моделирования.
4. Жизненный опыт каждого человека – это пример:
  - \* Аналогового моделирования.
  - \* Интуитивного моделирования.
  - \* Знакового моделирования.
5. С помощью вычислительного эксперимента можно на компьютере реализовать -
  - \* Модель, полученную при помощи физического моделирования.
  - \* Модель, полученную при помощи аналогового моделирования.
  - \* Модель, полученную при помощи математического моделирования.

## Лекция 2 Классификация математических моделей. Требования, предъявляемые к математическим моделям

Классификация в любой области знаний чрезвычайно важна. Она позволяет обобщить накопленный опыт, упорядочить понятия предметной области. Не является исключением в этом смысле и математическое моделирование. В табл.2.1 показаны виды математических моделей по различным признакам классификации.

Таблица 2.1 Классификация математических моделей

Признаки классификации	Виды математических моделей
1. Принадлежность к иерархическому уровню	1. Модели микроуровня 2. Модели макроуровня 3. Модели метауровня
2. Характер отображаемых свойств объекта	1. Структурные 2. Функциональные
3. Способ представления свойств объекта	1. Аналитические 2. Алгоритмические 3. Имитационные
4. Способ получения модели	1. Теоретические 2. Эмпирические
5. Особенности поведения объекта	1. Детерминированные 2. 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, расщепки 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 в примерах, приведенных в лекции 1. Ввиду того, что нами будет рассматриваться далее целый ряд аналитических моделей, поговорим о них более подробно.

Аналитическое моделирование основано на косвенном описании моделируемого объекта с помощью набора математических формул. Язык аналитического описания содержит следующие основные группы семантических элементов: критерий (критерии), неизвестные, данные, математические операции, ограничения. Наиболее существенная характеристика аналитических моделей заключается в том, что модель не является структурно подобной объекту моделирования. Под структурным подобием здесь понимается однозначное соответствие элементов и связей модели элементам и связям моделируемого объекта. К аналитическим относятся модели, построенные на основе аппарата математического программирования, корреляционного, регрессионного анализа. Аналитическая модель всегда представляет собой конструкцию, которую можно проанализировать и ре-

шить 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 удастся 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.

Имитационное моделирование основано на прямом описании моделируемого объекта. Существенной характеристикой таких моделей является структурное подобие объекта и модели. Это значит, что каждому существенному с точки зрения решаемой задачи элементу объекта ставится в соответствие элемент модели. При построении имитационной модели описываются законы функционирования каждого элемента объекта и связи между ними.

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:

Универсальности. Точности. Адекватности. 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 лекции 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  $Y = (y_1, y_2, \dots, y_m)$ ,  $y_{имм}$  -  $i$ -ый параметр, рассчитанный с помощью модели, а  $y_{ид}$  - истинное значение того же параметра. Тогда относительная погрешность математической модели по  $i$ -му параметру будет равна:

$$E_i = \frac{|y_{имм} - y_{ид}|}{y_{ид}}$$

По этой формуле рассчитываются погрешности для каждого выходного параметра, в результате получается вектор погрешностей  $E = (E_1, E_2, \dots, E_m)$ . В целом для математической модели погрешность оценивается следующим образом:

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

Например, оценим погрешность рассмотренной в лекции 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.

Т. к. 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

$$OA = \{X, Q / E_{мм} \leq \delta\},$$

где  $\delta$  – 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 соответствующего алгоритма и программы, реализующей математическую модель на ЭВМ.
4. **Наглядность**, т.е. удобное визуальное восприятие модели.







Рис 3.1. Граф для примера 2

Две встречные дуги в графе могут быть заменены линией без стрелок на концах. Эта линия называется ребром графа (рис. 3.2).

Матрицы смежности обоих графов, представленных на рис 3.2., имеют вид:

$$B = \begin{array}{cc|cc} & & m_1 & m_2 & & \\ & & 0 & 1 & & m_1 \\ & & 1 & 0 & & m_2 \\ & & & & & \end{array}$$

**Пример 3.** Пусть требуется изобразить граф, представленный следующей матрицей смежности:

$$B = \begin{array}{cccc|cccc} & & m_1 & m_2 & m_3 & m_4 & & \\ & & 0 & 1 & 1 & 0 & & m_1 \\ & & 0 & 0 & 0 & 1 & & m_2 \\ & & 0 & 1 & 0 & 1 & & m_3 \\ & & 0 & 1 & 0 & 0 & & m_4 \\ & & & & & & & \end{array}$$

Граф, соответствующий этой матрице, изображен на рис. 3.3.

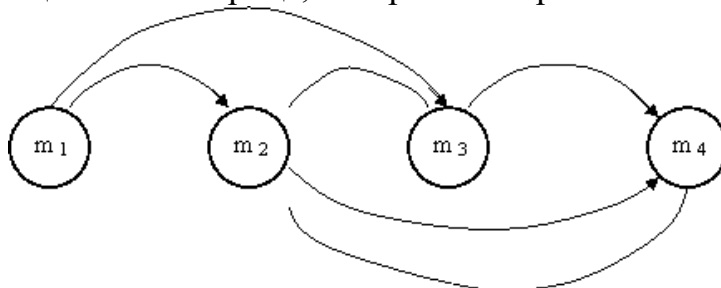


Рис. 3.3. Граф для примера 3

**Пример 4.** Имеется эскиз детали, представленный на рис.3.4.

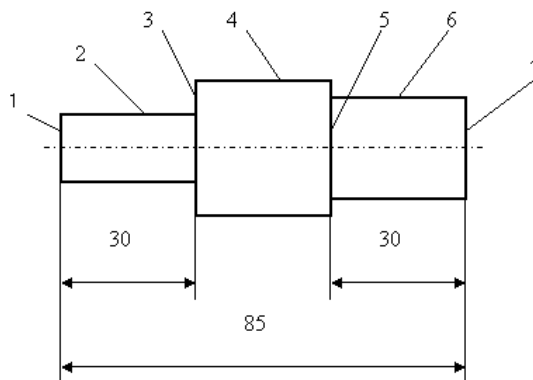


Рис. 3.4. Эскиз детали для примера 4

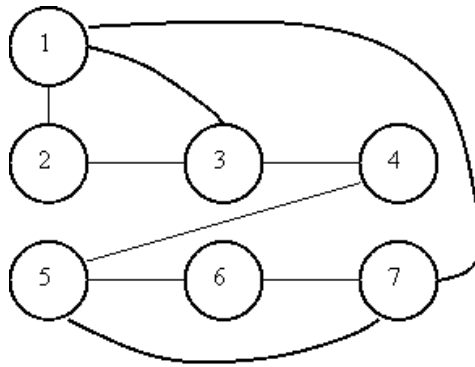


Рис.3.5. Граф для примера 4

Требуется изобразить в виде графа и матрицы смежности конструктивные и размерные связи этой детали.

Граф для этого примера изображен на рис. 3.5.

На рис. 3.5 **конструктивные связи детали** изображены (**тонкими линиями**) в виде ребер графа, т.к. здесь безразлично, связана, например, конструктивно поверхность 2 с поверхностью 3 или наоборот (они просто связаны между собой). Размеры также представлены между поверхностями, в этом примере нет приоритета, какая из поверхностей является базой при обработке детали.

Поэтому и **размерные связи** изображены (**толстыми линиями**) в виде ребер, а не дуг графа. Если же важно будет указать, что одна поверхность является базовой, а положение другой задано относительно ее, то в графе связь между этими поверхностями следует изобразить в виде дуги, исходящей из вершины, отождествляющей первую поверхность и входящей в вершину, отождествляющую вторую поверхность.

Матрица смежности данного графа имеет следующий вид:

$$\begin{array}{c}
 \begin{array}{cccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
 \begin{array}{c}
 \left| \left| \begin{array}{cccccccc}
 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
 1 & 0 & 0 & 0 & 1 & 1 & 0
 \end{array} \right| \right| & \begin{array}{c}
 1 \\
 2 \\
 3 \\
 4 \\
 5 \\
 6 \\
 7
 \end{array}
 \end{array}
 \end{array}
 \end{array}$$

## Лекция 4. Общая постановка и виды задач принятия решений.

### Математическая постановка и разрешимость задач оптимизации

При проектировании любых технических объектов, технологических процессов и систем всегда решаются задачи выбора и принятия решений.

**Задачей принятия решения** называют кортеж (совокупность)

$$\Omega = \langle X, O \rangle,$$

где  $X$  – множество вариантов решения задачи;

$O$  – принцип оптимальности, дающий представление о качестве вариантов, в простейшем случае – это правило их предпочтения друг перед другом.

**Решением задачи принятия решений** называется множество  $X_{\text{опт}}$ , которое является подмножеством множества  $X$  ( $X_{\text{опт}} \subset X$ ), полученное на основе принципа оптимальности.

Задачи принятия решений классифицируются по наличию информации об  $X$  и  $O$  и бывают трех видов:

1.  $X$  и  $O$  – неизвестны. Это общая задача принятия решений. Данные для получения  $X_{\text{опт}}$  определяют в данной задаче в процессе ее решения.
2.  $X$  – неизвестно,  $O$  – известно (эта задача поиска вариантов).
3.  $X$  и  $O$  – известны (это задача оптимизации).

В общем случае задача принятия решения решается в два этапа:

**1 этап:** Задача формализуется, т.е. строится ее математическая модель, в которой конкретные физические, технические, технологические, экономические условия и требования к объекту воплощаются в виде задачи оптимизации с определенной целевой функцией и допустимым множеством вариантов.

**2 этап:** Решение задачи оптимизации с использованием известных методов.

### Основы теории оптимизации

«Теория оптимизации», с одной стороны, является самостоятельной наукой, а, с другой стороны, составной частью науки под названием «исследование операций».

**Операция** (в данной науке) – это совокупность взаимосогласованных действий, направленных на достижение вполне определенной цели.

Обязательно должно быть сформулирована цель. Если есть разные пути достижения этой цели, то необходимо найти наилучший из них.

### Математическая постановка задачи оптимизации

Постановка задачи оптимизации включает в себя множество допустимых решений  $X = \{x_1, x_2, \dots, x_n\}$  и числовую функцию  $f$ , определенную на этом множестве, которая называется целевой функцией.

Нельзя отождествлять критерий (критерии) оптимальности и целевую функцию.

**Целевая функция** – это аналитическая зависимость между критерием (критериями) оптимальности и подлежащими оптимизации параметрами с указанием направления экстремума. Отличие понятий «критерий» и «целевая функция» состоит в следующем:

1. Целевая функция может включать в себя более одного критерия.
2. Для целевой функции всегда и обязательно указывается вид экстремума:  
 $f(x) \rightarrow \max$  ( $f(x) \rightarrow \min$ )

Различают два вида задач оптимизации: Задачу минимизации. 2. Задачу максимизации.

Чтобы решить задачу **минимизации** функции  $f$  на множестве  $X$ , необходимо найти такой вектор  $x_0 \in X$  (а также соответствующее значение целевой функции  $f(x_0)$ ), чтобы неравенство:  $f(x_0) \leq f(x)$  выполнялось для всех  $x \in X$ . При этом  $x_0$  называют оптимальным решением (точнее здесь – минимальным решением), а  $f(x_0)$  – оптимумом (минимумом).

Чтобы решить задачу **максимизации** функции  $f$  на множестве  $X$ , необходимо найти такой вектор  $x_0 \in X$  ( а также соответствующее значение целевой функции  $f(x_0)$  ), чтобы неравенство:  $f(x_0) \geq f(x)$  выполнялось для всех  $x \in X$ . При этом  $x_0$  называют оптимальным (максимальным) решением, а  $f(x_0)$  – оптимумом ( максимумом ).

В общем виде находится именно вектор  $x_0$ , т.к., например, при решении двухпараметрической задачи, он будет включать в себя два параметра, трехпараметрической – три параметра и т.д.

### Локальный и глобальный минимум (максимум) целевой функции

При решении задач оптимизации следует иметь в виду, какой вид имеет целевая функция.

Например,  $a \leq x \leq b, f(x) \rightarrow \min$ . Целевая функция имеет вид, как показано на рис. 4.1.

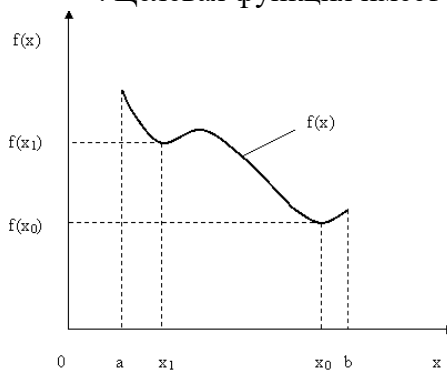


Рис.4.1. Иллюстрация понятий «локальный минимум» и «глобальный минимум» целевой функции. Здесь т.  $x_0$  - глобальный минимум, а т.  $x_1$  - локальный минимум целевой функции.

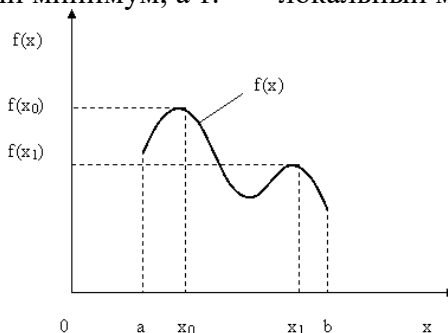


Рис.4.2. Иллюстрация понятий «локальный максимум» и «глобальный максимум» целевой функции

Пусть теперь  $a \leq x \leq b, f(x) \rightarrow \max$ , а целевая функция имеет вид, как показано на рис.4.2.

Здесь  $x_0$  - точка глобального максимума, а  $x_1$  - точка локального максимума целевой функции.

### Разрешимость задач оптимизации

Приведенная выше задача оптимизации имеет решение не при любых целевых функциях и допустимых множествах. Существуют задачи, в которых невозможно найти оптимальное решение и экстремум целевой функции. Например, не существует точек минимума функции одной переменной  $f$  на множестве  $X(f(x) \rightarrow \min)$  в случаях, приведенных ниже на рис. 4.3, 4.4 и 4.5.

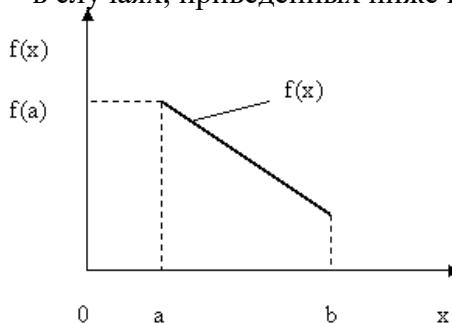


Рис.4.3. Иллюстрация случая, когда множество допустимых решений не замкнуто. Здесь граница «а» множества допустимых решений в интервал входит, а граница «b» нет.

$X = [a, b)$  - множество  $X$  не замкнуто, следовательно,  $f(b)$  – не существует.

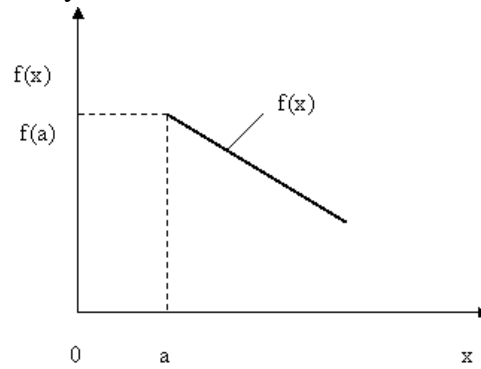


Рис.4.4. Иллюстрация неограниченности множества допустимых решений

В случае, представленном на рис.4.4, определена лишь одна левая граница множества допустимых решений.  $X = [a, \infty)$ , т.е. множество допустимых решений неограниченно.

На рис.4.5. показан еще один случай, когда задача оптимизации не имеет однозначного решения.

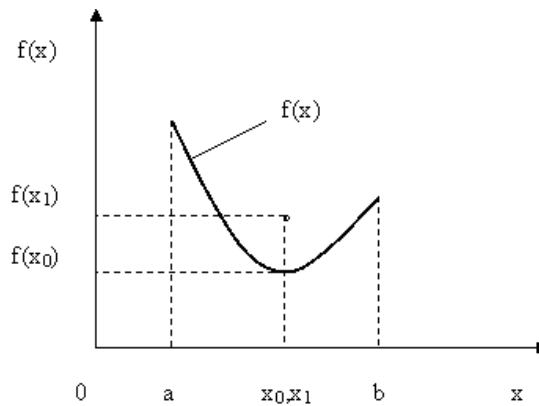


Рис.4.5. Иллюстрация случая, когда функция  $f(x)$  не является непрерывной. Здесь функция  $f(x)$  не является непрерывной, т.к. в т.  $x_0, x_1$  существуют два значения функции –  $f(x_0)$  и  $f(x_1)$ .

Следовательно, задача оптимизации разрешима, если выполняются следующие три условия:

1. Множество допустимых решений  $X$  замкнуто, т.е. если предельные точки принадлежат этому множеству.
2. Множество  $X$  ограничено.
3. Целевая функция  $f(x)$  непрерывна.

Это нестрогая формулировка **теоремы Вейерштрасса**.

## Лекция 5. Графо – аналитический метод решения задач математического программирования

Задачами **математического программирования** называют **однокритериальные задачи оптимизации**. Методы их решения оперируют с детерминированными математическими моделями. Напомним, что детерминированная математическая модель отражает поведение объекта с позиций полной определенности в настоящем и будущем.

Эти модели в исследовании операций занимают одно из главных мест. Это обусловлено тем, что в них отражены разнообразные проблемы распределения ограниченных ресурсов в экономике, военном деле, создании новой техники и т.д. Пути решения этих проблем так или иначе связаны с планированием целенаправленной деятельности, т.е. с разработкой определенных установок на будущее.

Термин «программирование» (от английского «programming») - составление плана или программы действий) здесь следует понимать в смысле «поиск наилучших планов» (в отличие от того толкования, которое принято специалистами по программному обеспечению ЭВМ, – «составление программы для ЭВМ»).

**Задача математического программирования** формулируется следующим образом: найти значения переменных  $x_1, x_2, \dots, x_n$ , доставляющие максимум (минимум) заданной целевой функции  $y = f(x_1, x_2, \dots, x_n)$  при условиях:

$$g_j(x_1, x_2, \dots, x_n) \leq (\geq, =) b_j, (j = \overline{1, m}).$$

Различают два вида задач математического программирования:

1. Задачи линейного программирования.
2. Задачи нелинейного программирования.

В первых задачах функция  $\mathcal{Y}$  и ограничения  $g_j$  линейны относительно переменных  $x$ . Во вторых задачах целевая функция  $\mathcal{Y}$  и (или) условия  $g_j$  имеют разного рода нелинейности.

### Графоаналитический метод решения задач оптимизации

Этим методом вручную решаются простые задачи оптимизации. Математические модели в этих задачах не должны быть сложными, т.к. в противном случае требуется много времени для их решения. Для начала рассмотрим однопараметрическую однокритериальную задачу оптимизации.

**Постановка задачи:** Дан один критерий  $\mathcal{Y}$ . Объект (процесс) описан уравнением (уравнениями), включающими один искомый параметр  $y = f(x)$ . Имеется система ограничений:

$$\begin{aligned} 1) x &\geq a_1; \\ 2) a_2 &\leq x \leq b_1; \text{ и т.д.} \end{aligned}$$

Необходимо найти оптимальное значение параметра  $x - x_{\text{опт}}$ , обращающее целевую функцию  $f(x)$  в максимум или минимум.

Задача решается в два этапа:

1. Построение области допустимых решений (ОДР).
2. Нахождение в пределах ОДР оптимального решения.

При построении ОДР на первом этапе рассматривается система ограничений. Все ограничения должны быть выполнены. Выполнение первого ограничения в приведенной выше постановке задачи оптимизации означает, что искомое значение параметра  $x$  должно находиться правее  $a_1$ , причем,  $a_1$  в разрешенный интервал входит (рис.5.1). Выполнение второго ограничения означает, что искомое значение параметра  $x$  должно находиться в интервале (на отрезке)  $[a_2, b_1]$ , следует иметь в виду, что границы интервала в ОДР входят.

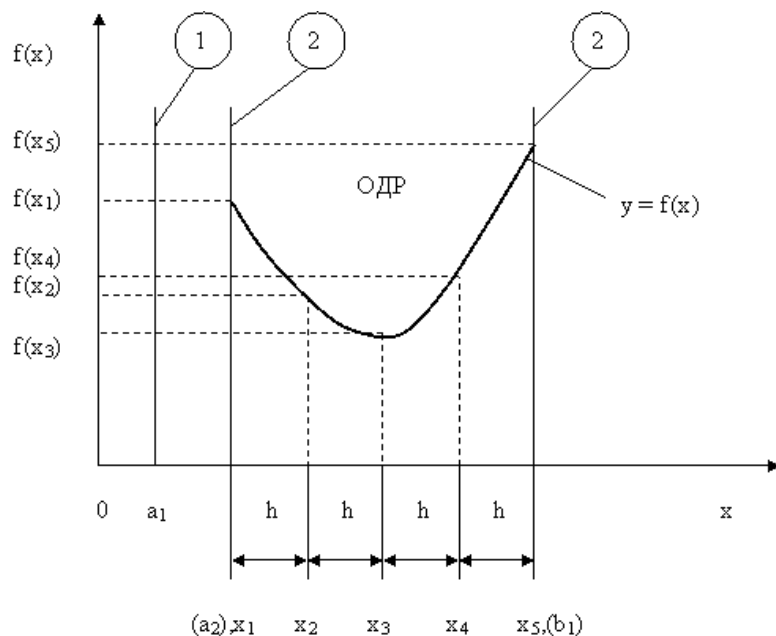


Рис.5.1. Графическая иллюстрация решения однопараметрической однокритериальной задачи оптимизации

Когда однопараметрическая однокритериальная задача оптимизации решается с применением графоаналитического метода вручную, то на втором этапе применяют метод перебора. Суть его заключается в следующем. В пределах ОДР через определенный интервал  $h$  выбирается ряд значений параметра  $x$ . В рассматриваемом нами случае ОДР разбита на четыре отрезка, и выбрано пять значений параметра  $x$ . Для этих значений параметра  $x$  рассчитываются соответствующие значения целевой функции. Среди них находят минимальное (максимальное) значение. Значение параметра  $x_i$ , обращающее целевую функцию в минимум (максимум), является оптимальным. Если в рассматриваемом нами случае  $f(x)$  стремится к минимуму, то  $x_{\text{опт}} = x_3$ , если к максимуму, то  $x_{\text{опт}} = x_5$ .

При решении практических задач оптимизации всегда следует иметь в виду, какова целевая функция. Это значительно упрощает работу как при решении задач оптимизации вручную с применением графоаналитического метода, так и при решении таких задач с использованием компьютерных программ. Причем, это относится и к случаю использования готовых программ, и, что особенно важно, к разработке собственных программ.

Рассмотрим, например, следующий частный случай, когда целевая функция линейная (рис.5.2.).

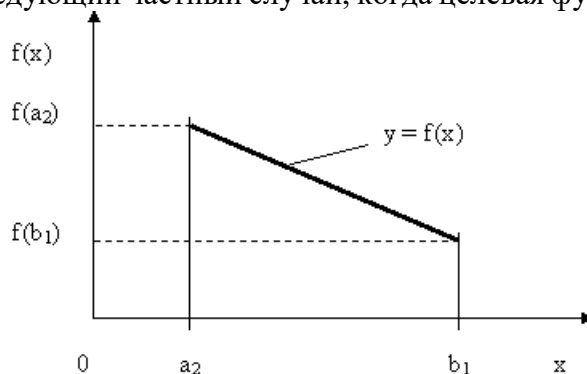


Рис.5.2. Графическая иллюстрация решения однопараметрической однокритериальной задачи оптимизации для случая линейной целевой функции

В данном случае на втором этапе вычисляют значения целевой функции только на границах ОДР. Эти значения сравнивают и выбирают наименьшее или наибольшее. Для примера, приведенного на рис. 5.2, если  $f(x) \rightarrow \min$ , то  $x_{\text{опт}} = b_1$ , если  $f(x) \rightarrow \max$ , то  $x_{\text{опт}} = a_2$ .



Рассмотрим теперь графоаналитический метод решения многопараметрической однокритериальной задачи оптимизации. Вручную этим методом реально можно решить максимум двухпараметрическую задачу оптимизации.

**Постановка задачи:**

Дан один критерий  $\mathcal{Y}$ . Объект(процесс) описан уравнением (уравнениями), включающими в себя ряд параметров  $y = f(x_1, x_2, \dots, x_n)$ . Имеется система ограничений:

$$\begin{aligned} &1) a_1 \leq x_1 \leq b_1; \\ &2) a_2 \leq x_2 \leq b_2; \\ &\dots \\ &n) a_n \leq x_n \leq b_n; \\ &\dots \\ &n+1) x_1 \cdot x_2 \leq c_1; \\ &n+2) x_2 \cdot x_3 \leq c_2; \\ &\dots \end{aligned}$$

Требуется определить оптимальное значение ряда параметров  $x_{1\text{опт}}, x_{2\text{опт}}, \dots, x_{m\text{опт}} (m \leq n)$ , обращающих целевую функцию  $f(x_1, x_2, \dots, x_n)$  в максимум или минимум.

**Пример.** Дан критерий  $y = x_2 / x_1$ . Требуется найти  $x_{1\text{опт}}$ , и  $x_{2\text{опт}}$ , обращающие в максимум целевую функцию  $y = x_2 / x_1 \rightarrow \max$ . Ограничения:

$$\begin{aligned} &1 \leq x_1 \leq 8; \\ &2 \leq x_2 \leq 12; \\ &x_1 \cdot x_2 \geq 10. \end{aligned}$$

Задача решается опять в два этапа:

1. Построение ОДР.
2. Нахождение в пределах ОДР оптимального решения.

Построение ОДР в данной задаче в отличие от задачи однопараметрической заключается в том, что работать нужно в двух направлениях. В итоге в плоскости  $x_1 \text{О} x_2$  ОДР будет представлять собой многогранник (рис. 5.3).

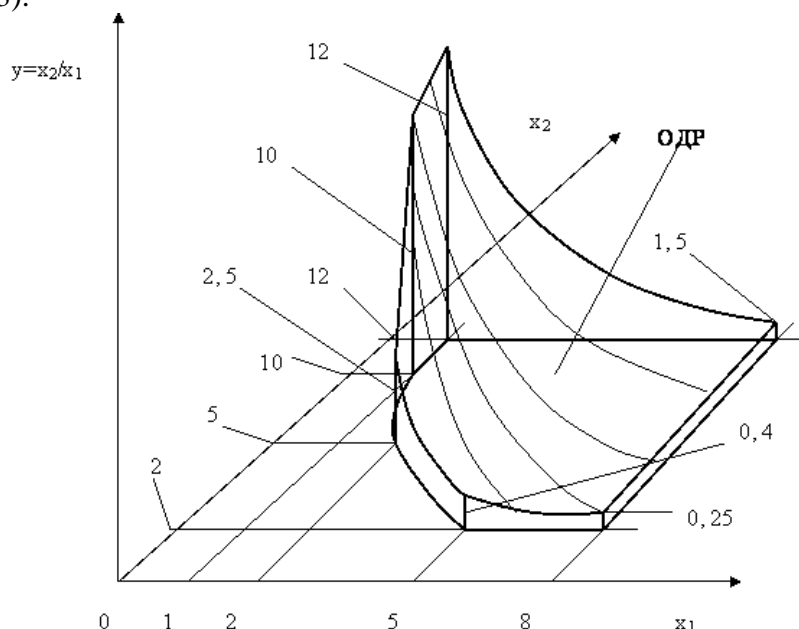


Рис.5.3. Графическая иллюстрация решения двухпараметрической однокритериальной задачи оптимизации

Для построения нелинейного ограничения  $x_1 \cdot x_2 \geq 10$  сначала необходимо приравнять левую и правую части неравенства и построить соответствующую кривую.

$$x_1 \cdot x_2 = 10; x_2 = 10 / x_1;$$

$x_1$	<b>1</b>	<b>2</b>	<b>5</b>
$x_2$	<b>10</b>	<b>5</b>	<b>2</b>

После этого нужно определить направление допустимости искомых параметров  $x_1$  и  $x_2$ . Чтобы не ошибиться, для этого можно применить следующий прием. Выберем произвольную точку на плоскости  $x_1$  и  $x_2$  с любой стороны кривой. Например, выберем точку с координатами  $x_1 = 5, x_2 = 5$ , т.е. «справа - вверху» от кривой. Вычислим значение левой части неравенства:  $5 \cdot 5 = 25; 25 > 10$ , следовательно, неравенство выполняется. Это значит, что выбранная точка находится в допустимой области искомых параметров. Т.е. допустимая область искомых параметров находится «справа - вверху» от кривой.

На втором этапе необходимо вычислить значения целевой функции в пределах ОДР. В данном примере искомая точка, определяющая оптимальные значения искомых параметров, находится на границе ОДР:  $x_{1\text{опт}} = 1, x_{2\text{опт}} = 12$ . Если  $x_2 / x_1 \rightarrow \min$ , то  $x_{1\text{опт}} = 8, x_{2\text{опт}} = 2$ .

## Лекция 6. Методы решения задач линейного программирования

**Постановка задачи.** Требуется найти оптимальные значения  $n$  переменных  $x_1, x_2, \dots, x_n$ , доставляющие экстремум **линейной** целевой функции

$$y = c_1x_1 + c_2x_2 + \dots + c_nx_n = \sum_{j=1}^n c_jx_j.$$

Имеется конечное число **линейных** ограничений вида:

$$a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pn}x_n \geq b_p, p = \overline{1, k_1};$$

$$a_{l1}x_1 + a_{l2}x_2 + \dots + a_{ln}x_n \leq b_l, l = \overline{1, k_2};$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m, m = \overline{1, k_3}.$$

### Графоаналитический метод решения задач линейного программирования

**Пример.** Найти оптимальные значения  $x_{1opt}, x_{2opt}$ .

**Целевая функция**

$$y = x_1 + 2x_2 \rightarrow \max.$$

**Ограничения:**

$$2x_1 + x_2 \leq 10;$$

$$x_1 + 3x_2 \leq 18;$$

$$x_1 \geq 0;$$

$$x_2 \geq 0.$$

**1 этап.** Построение области допустимых решений (ОДР).

Для формирования ОДР необходимо в системе координат  $x_1 \text{ vs } x_2$  построить линии, соответствующие ограничениям, приравняв их левые и правые части, и определить направления расположения допустимых значений искомых переменных в соответствии со знаками неравенств (рис.6.1).

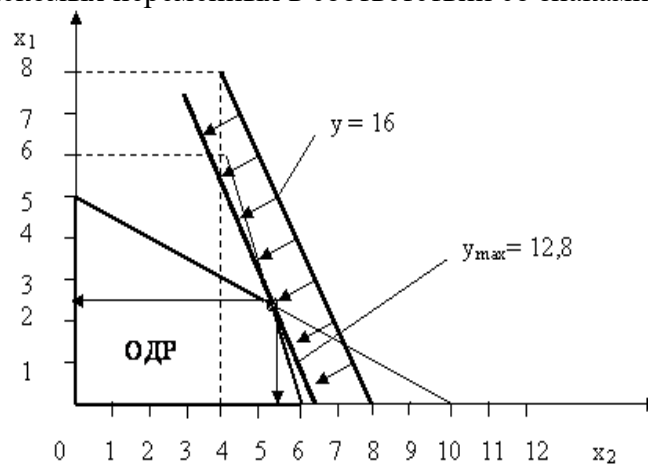


Рис.6.1. Графическая иллюстрация решения задачи линейного программирования

Вычисления для построения первых двух ограничений:

$$2x_1 + x_2 \leq 10 \rightarrow 2x_1 + x_2 = 10 \Rightarrow x_2 = 10 - 2x_1;$$

$x_1$	0	5
$x_2$	10	0

$$x_1 + 3x_2 \leq 18 \rightarrow x_1 + 3x_2 = 18 \Rightarrow x_2 = (18 - x_1) / 3;$$

$x_1$	0	6
$x_2$	6	4

Направления допустимости значений переменных  $x_1$  и  $x_2$  в соответствии с первыми двумя ограничениями – «вниз - влево». В соответствии с третьим ограничением значения переменной  $x_1$

должны находиться выше оси  $Ox_2$ , а в соответствии с четвертым ограничением значения переменной  $x_2$  должны находиться правее оси  $Ox_1$ . Следует иметь в виду, что границы ОДР в область допустимых решений входят. Это объясняется тем, что в ограничениях применяются знаки неравенств «меньше – равно» и «больше – равно».

**2 этап.** Окончательное определение оптимальных значений переменных  $x_1$  и  $x_2$ .

Для этого необходимо сначала построить произвольную прямую для целевой функции, приравняв ее выражение к любому числу, так, чтобы эта прямая располагалась в пределах выбранного масштаба рисунка. Приравняем выражение целевой функции, например, к числу «16» и построим соответствующую прямую линию.

$$y = x_1 + 2x_2 \rightarrow \max; x_1 + 2x_2 = 16 \Rightarrow x_2 = (16 - x_1) / 2;$$

$x_1$	0	8
$x_2$	8	4

После этого необходимо построить прямую линию, параллельную данной прямой, так, чтобы она **коснулась** границы ОДР. Координаты точки **касания** этой прямой с границей ОДР будут оптимальными значениями переменных  $x_{1opt}$  и  $x_{2opt}$ .

Для точного определения координат точки касания линии целевой функции границы ОДР (точных значений  $x_{1opt}$  и  $x_{2opt}$ ) необходимо решить систему, включающую в себя два уравнения: " $2x_1 + x_2 = 10$ " и " $x_1 + 3x_2 = 18$ ". Решим эту систему уравнений:  $2x_1 + x_2 = 10 \Rightarrow x_2 = 10 - 2x_1$ . Подставим это значение  $x_2$  во второе уравнение:

$$x_1 + 3(10 - 2x_1) = 18 \Rightarrow x_1 + 30 - 6x_1 - 18 = 0 \Rightarrow -5x_1 + 12 = 0 \Rightarrow x_1 = 12 / 5 = 2,4; x_2 = 10 - 2 \cdot 2,4 = 5,2.$$

Т.е.  $x_{1opt} = 2,4; x_{2opt} = 5,2$ . При этом максимальное значение целевой функции  $y_{max} = x_1 + 2x_2 = 2,4 + 2 \cdot 5,2 = 12,8$ .

### Симплекс метод решения задач линейного программирования

ОДР в двухпараметрической задаче линейного программирования представляет собой плоский многогранник (см. предыдущий пример), а в общем виде это выпуклый многогранник.

**Теорема:** экстремум целевой функции в задаче линейного программирования, если он существует, всегда является абсолютным и достигается хотя бы в одной крайней точке многогранника, определяющего ОДР.

Примечание. Крайние точки – это точки пересечения границ ОДР.

Идея симплекс метода заключается в направленном переборе крайних точек ОДР. Этот метод является классическим в линейном программировании.

Пример. **Требуется минимизировать** целевую функцию

$$y = x_1 + x_2 + x_3 \rightarrow \min$$

при следующих ограничениях:

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1; \\ -x_1 + 2x_2 - x_3 &= 2; \\ x_1 &\geq 0; \\ x_2 &\geq 0; \\ x_3 &\geq 0. \end{aligned}$$

$$x_{1opt}, x_{2opt}, x_{3opt} - ?$$

Поставим в соответствие этой задаче следующую систему линейных уравнений:

$$y - x_1 - x_2 - x_3 = 0; \tag{6.1}$$

$$x_1 - x_2 + 2x_3 = 1; \tag{6.2}$$

$$-x_1 + 2x_2 - x_3 = 2. \tag{6.3}$$

Решим эту систему методом последовательного исключения неизвестных (методом Гаусса). В его основе лежат следующие элементарные преобразования:

1. Умножение какого - либо уравнения на число, отличное от нуля;

2. Сложение двух уравнений и последующая замена одного из них получившейся суммой;

3. Перемена местами любых двух уравнений.

С помощью элементарных преобразований получим следующую систему:

$$y + 3x_3 = 7; \quad (6.4)$$

$$x_1 + 3x_3 = 4; \quad (6.5)$$

$$x_2 + x_3 = 3. \quad (6.6)$$

Уравнение (6.6) – это сумма уравнений (6.2) и (6.3).

Уравнение (6.5) – это сумма уравнений (6.2) и (6.6).

Уравнение (6.4) – это сумма уравнения (6.1) и уравнений (6.5) и (6.6).

Пусть  $x_3 = 0$ , при этом значения переменных будут следующими:  $y = 7, x_1 = 4, x_2 = 3, x_3 = 0$ .

Эти значения являются решением системы уравнений (6.4) – (6.6) и, значит, системы уравнений (6.1) – (6.3). Причем,  $x_1, x_2, x_3 \geq 0$ , т.е. найденное решение допустимо.

Т.к.  $y = 7 - 3x_3$  (см.6.4), то для уменьшения значения целевой функции необходимо увеличить  $x_3$ , но увеличивать его бесконечно нельзя, иначе величины  $x_1$  и  $x_2$  станут отрицательными ( $x_1 = 4 - 3x_3, x_2 = 3 - x_3$ , эти выражения получены соответственно из уравнений 6.5 и 6.6).

Согласно первому уравнению  $x_3$  может быть равным  $4/3$ , согласно второму уравнению 3. Для того, чтобы все условия задачи выполнялись, необходимо взять наименьшее из этих значений -  $x_3 = \min \{4/3, 3\} = 4/3$  (если принять  $x_3 = 3$ , то  $x_2 = 0$ , что приемлемо, а  $x_1 = 4 - 3x_3 = 4 - 3 \cdot 3 = -5$ , что не приемлемо для нашей задачи, т.е. необходимо принять  $x_3 = 4/3$ ). Тогда из уравнения (6.5) следует, что  $x_1 = 0$ .

С помощью элементарных преобразований приведем систему уравнений (6.4) – (6.6) к следующей системе:

$$y - x_1 = 3; \quad (6.7)$$

$$x_3 + 1/3x_1 = 4/3; \quad (6.8)$$

$$x_2 - 1/3x_1 = 5/3. \quad (6.9)$$

Система (6.7) – (6.9) из системы (6.4) – (6.6) получена следующим образом:

$$x_1 + 3x_3 = 4(6.5) \Rightarrow 3x_3 = 4 - x_1;$$

$$y + 3x_3 = 7(6.4) \Rightarrow y = 7 - 3x_3 \Rightarrow y = 7 - (4 - x_1) = 7 - 4 + x_1 = 3 + x_1 \Rightarrow y - x_1 = 3.$$

$$x_1 + 3x_3 = 4(6.5),$$

разделим это уравнение на 3, получим  $x_3 + 1/3x_1 = 4/3(6.8) \Rightarrow x_3 = 4/3 - 1/3x_1$  - подставим в (6.6)

$$x_2 + x_3 = 3(6.6) \Rightarrow x_2 + 4/3 - 1/3x_1 = 3 \Rightarrow x_2 - 1/3x_1 = 3 - 4/3 \Rightarrow x_2 - 1/3x_1 = 5/3(6.9).$$

Запишем решение системы уравнений (6.7) – (6.9):

$$y = 3, x_1 = 0, x_2 = 5/3, x_3 = 4/3.$$

Здесь  $y = 3 + x_1$ , уменьшать  $x_1$  для дальнейшего уменьшения целевой функции нельзя, т.к. оно станет отрицательным, увеличивать его нет смысла, т.к. при этом увеличивается целевая функция. Следовательно, полученное значение  $y = 3$  является **минимальным**, и ему соответствуют оптимальные значения искомых переменных:  $x_{1\text{опт}} = 0, x_{2\text{опт}} = 5/3, x_{3\text{опт}} = 4/3$ . Эти оптимальные значения переменных справедливы и для исходной задачи.

### Симплекс – метод. Этапы поиска решений

Для использования симплекс - метода, задачу линейного программирования необходимо привести к стандартному виду: найти  $x_1, x_2, \dots, x_n$ , доставляющие минимум целевой функции

$$y = \sum_{j=1}^n c_j x_j$$

при условиях:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1;$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2;$$

$$\dots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m;$$

$$\forall x_j \geq 0 (j = \overline{1, n}).$$

Здесь значок  $\forall$  означает «все».

С формальной точки зрения нет необходимости строго различать задачи поиска минимума или максимума целевой функции. Одна задача сводится к другой изменением знака  $\mathcal{Y}$  на противоположный. Все  $b_i$  должны быть положительными, в противном случае соответствующее уравнение умножается на (-1).

Рассмотрим этапы поиска решений симплекс – методом на следующем примере.

**Пример.** Модель раскроя листовых материалов. Выбор наилучшего варианта.

Имеется некоторый материал в виде стандартных листов, которые необходимо раскроить для получения не менее 80шт. деталей типа 1 и не менее 40шт. деталей типа 2. Известны четыре способа раскроя листа, каждый из которых дает результат, представленный в таблице 6.1.

Таблица 6.1 Исходные данные для решения задачи

Способы раскроя листа	1	2	3	4
Результат	3 детали типа 1; 1 деталь типа 2	2 детали типа 1; 6 деталей типа 2	1 деталь типа 1; 9 деталей типа 2	0 деталей типа 1; 13 деталей типа 2

Требуется так провести операцию изготовления деталей, чтобы общий расход листов оказался минимальным.

Пусть  $x_j$  – количество листов, раскраиваемых  $j$  – м способом ( $j = \overline{1, 4}$ ). Тогда целевая функция, будет иметь следующий вид:

$$\mathcal{Y} = x_1 + x_2 + x_3 + x_4 \rightarrow \min .$$

Здесь  $\mathcal{Y}$  – это общий суммарный расход листов.

Формулировка задачи. Найти  $x_1, x_2, x_3, x_4$ , доставляющие минимум приведенной выше целевой функции при следующих ограничениях:

$$3x_1 + 2x_2 + x_3 \geq 80; \quad (6.10)$$

$$x_1 + 6x_2 + 9x_3 + 13x_4 \geq 40; \quad (6.11)$$

$$\forall x_j \geq 0 (j = \overline{1, 4}). \quad (6.12)$$

Ограничение (6.10) обеспечивает заданное число деталей типа 1, ограничение (6.11) – заданное число деталей типа 2. Ограничение (6.12) необходимо, т.к. количество листов не может быть отрицательным.

Для того, чтобы для решения данной задачи применить симплекс метод, необходимо систему (6.10) – (6.12) привести к **стандартному виду**:

$$3x_1 + 2x_2 + x_3 - x_5 = 80; \quad (6.13)$$

$$x_1 + 6x_2 + 9x_3 + 13x_4 - x_6 = 40; \quad (6.14)$$

$$\forall x_j \geq 0 (j = \overline{1, 6}). \quad (6.15)$$

Очевидно следующее решение системы (6.13) – (6.15):  $x_1 = x_2 = x_3 = x_4 = 0, x_5 = -80, x_6 = -40$ .

Это решение называется **базисным**. Оно неприемлемо, т.к.  $x_5$  и  $x_6$  отрицательны. Следовательно, систему (6.13) – (6.15) необходимо преобразовать и решать ее в два этапа.

**1 этап.** Вспомогательный.

Он выполняется для нахождения **приемлемого (допустимого) и оптимального для данного этапа базисного решения**, т.е. по существу координат одной крайней точки области допустимых решений с целью направленного перебора на втором этапе остальных крайних точек для выбора оптимального решения для задачи в целом.

Преобразуем систему ( 6.13) – (6.15) в систему ( 6.16) – (6.18):

$$3x_1 + 2x_2 + x_3 - x_5 + x_7 = 80; \quad (6.16)$$

$$x_1 + 6x_2 + 9x_3 + 13x_4 - x_6 + x_8 = 40; \quad (6.17)$$

$$\forall x_j \geq 0 (j = \overline{1,8}). \quad (6.18)$$

Ставим вспомогательную задачу. Необходимо найти  $x_1, x_2, \dots, x_8$ , доставляющие минимум следующей целевой функции:  $\varpi = x_7 + x_8 \rightarrow \min$ . Здесь легко находятся исходное допустимое базисное решение:

$$x_1, x_2, \dots, x_6 = 0, x_7 = 80, x_8 = 40$$

После этого начинается реализация симплекс – алгоритма для решения вспомогательной задачи. Работа при этом осуществляется с системой (6.16) – (6.18).

1. Выражаются базисные переменные  $x_7$  и  $x_8$  через свободные переменные  $x_1, x_2, \dots, x_6$ .

$$x_7 = 80 - 3x_1 - 2x_2 - x_3 + x_5; x_8 = 40 - x_1 - 6x_2 - 9x_3 - 13x_4 + x_6.$$

2. Определяется целевая функция:

$$\varpi = x_7 + x_8 = 120 - 4x_1 - 8x_2 - 10x_3 - 13x_4 + x_5 + x_6.$$

3. Анализируется выражение для целевой функции, в нем присутствуют отрицательные коэффициенты, значит текущее базисное решение не оптимально.

**Теорема.** Допустимое базисное решение является оптимальным, если коэффициенты  $c_j$  при свободных переменных  $x_j$  в выражении для целевой функции неотрицательны.

Для текущего базисного решения  $\varpi = 80 + 40 = 120$ , т.к.  $x_1, x_2, \dots, x_6 = 0$ .

4. В выражении для целевой функции находится переменная, имеющая наибольший по модулю отрицательный коэффициент, она вводится в базис для получения нового базисного решения.

Для этого анализируются зависимости текущих базисных переменных (здесь  $x_7$  и  $x_8$ ), от вводимой в базис переменной (здесь  $x_4$ , т.к. она имеет коэффициент «-13») при условии, что свободные переменные равны нулю ( $x_1, x_2, \dots, x_6 = 0$ ):  $x_7 = 80, x_8 = 40 - 13x_4$ . Очевидно, что  $x_4$  можно увеличить самое большее до  $40/13$ , при этом

$$x_8 = 40 - 13 \cdot (40/13) = 0,$$

т.е. получим новое базисное решение:

$$x_7 = 80, x_4 = 40/13, x_1 = x_2 = x_3 = x_5 = x_6 = x_8 = 0$$

Снова выполняем симплекс алгоритм:

1. Опять выражаем базисные переменные через свободные:

$$x_7 = 80 - 3x_1 - 2x_2 - x_3 + x_5; x_8 = 0.$$

2.  $\varpi = x_7 + x_8 = 80 - 3x_1 - 2x_2 - x_3 + x_5$ .

3. Минимум  $\varpi$  не достигнут (текущее базисное решение не оптимально), т.к. выражение для целевой функции содержит переменные с отрицательными коэффициентами.

$\varpi = 80$ , т.к.  $x_1 = x_2 = x_3 = x_5 = 0$ , т.е. значение целевой функции уменьшилось по сравнению с предыдущим.

4. Вводим в базис  $x_1$ , т.к. эта переменная в выражении для целевой функции имеет коэффициент «-3».  $x_7 = 80 - 3x_1$ , т.к.  $x_2 = x_3 = x_5 = 0$ .  $x_4 = (40 - x_1)/13$ , т.к.  $x_1 + 13x_4 = 40$  (см. уравнение (6.17), принимая во внимание, что  $x_2 = x_3 = x_6 = x_8 = 0$ ). Здесь  $x_1 = \min \{80/3, 40\}$ , т.е.  $x_1$  можно увеличить до  $80/3$  (до 40 нельзя, т.к. при этом  $x_7 = 80 - 3 \cdot 40 = -40 < 0$ ). Следовательно,



$x_7 = 80 - 3x_1 = 80 - (80/3) \cdot 3 = 0$ ,  $x_1 = 80/3$ ,  $x_4 = (40 - 80/3)/13 = 40/39$ , т.е. получили новое базисное решение:  $x_1 = 80/3, x_4 = 40/39, x_2 = x_3 = x_5 = x_6 = x_7 = x_8 = 0$ .

Это базисное решение на данном этапе является оптимальным (выражение для целевой функции  $\varpi = x_7 + x_8$  не содержит отрицательных коэффициентов). При этом

$$\begin{aligned}\varpi &= x_7 + x_8 = 120 - 4x_1 - 8x_2 - 10x_3 - 13x_4 + x_5 + x_6; \\ \varpi &= 120 - 4 \cdot (80/3) - 8 \cdot 0 - 10 \cdot 0 - 13 \cdot (40/39) + 0 + 0 = 0.\end{aligned}$$

Целевую функцию уменьшить нельзя, т.к.  $\forall x_j \geq 0$ .

Переменные  $x_7 = x_8 = 0$ , т.е. они перешли в разряд свободных переменных и могут быть отброшены. Окончательное базисное решение  $x_1 = 80/3, x_4 = 40/39, x_2 = x_3 = x_5 = x_6 = 0$  является оптимальным для вспомогательного этапа 1 и допустимым для системы (6.13) – (6.15).

**2 этап.** Работаем с системой (6.13)–(6.15). Имея допустимое базисное решение  $x_1 = 80/3, x_4 = 40/39, x_2 = x_3 = x_5 = x_6 = 0$ , полученное в результате минимизации  $\varpi$  на этапе 1, оценим его для исходной задачи с целевой функцией  $y = x_1 + x_2 + x_3 + x_4$ . Тем самым начинается повторное использование симплекс - алгоритма, но уже для исходной задачи.

1. С помощью системы (6.13) – (6.15) получим:

$$\begin{aligned}x_1 &= \frac{80 - 2x_2 - x_3 + x_5}{3}; \\ x_4 &= \frac{40 - x_1 - 6x_2 - 9x_3 + x_6}{13}.\end{aligned}$$

2.  $y = x_1 + x_2 + x_3 + x_4 = (40 \cdot 27 - 3x_2 + 12x_3 + 3x_6)/39$  (приводится без вывода).

3. Данное базисное решение не оптимально, т.к. переменная  $x_2$  имеет отрицательный коэффициент.

4. Вводим в базис переменную  $x_2$ .

$$x_1 = \frac{80 - 2x_2}{3}; \quad x_4 = \frac{40 - 16x_2}{39} \text{ (выражение для } x_4 \text{ приводится без вывода).}$$

$x_2 = \min \{40, 40/16\}$ ,  $x_2$  можно увеличить самое большее до  $40/16 = 5/2$ , при этом

$$\begin{aligned}x_4 &= \frac{40 - 16 \cdot (40/16)}{39} = 0, \\ x_1 &= \frac{80 - 2 \cdot (5/2)}{3} = 25,\end{aligned}$$

т.е. получили новое базисное решение:  $x_1 = 25, x_2 = 5/2, x_3 = x_4 = x_5 = x_6 = 0$ .

Снова выполняем симплекс – алгоритм:

1,2. Выражаем  $y$  через свободные переменные  $x_3, x_4, x_5, x_6$ .

$$y = x_1 + x_2 + x_3 + x_4 = \frac{440 + 5x_3 + 3x_4 + 5x_5 + 5x_6}{16} \text{ (приводится без вывода).}$$

3. В выражении для целевой функции все коэффициенты при переменных положительны. Т.е. текущее базисное решение является оптимальным, значит  $x_{1\text{опт}} = 25, x_{2\text{опт}} = 2,5, x_{3\text{опт}} = 0, x_{4\text{опт}} = 0$ . Т.е. наилучший вариант технологии состоит в применении только первого и второго способов раскроя листов, причем первым способом необходимо раскроить 25 листов, а вторым 2,5 листа, тогда наименьший общий суммарный расход листов составит 27,5 шт. Количество получаемых деталей при этом будет: деталь типа 1 –

$$3x_1 + 2x_2 + x_3 = 3 \cdot 25 + 2 \cdot 2,5 + 0 = 80 \text{шт.},$$

деталь типа 2 –

$$x_1 + 6x_2 + 9x_3 + 13x_4 = 25 + 6 \cdot 2,5 + 9 \cdot 0 + 13 \cdot 0 = 40 \text{шт.}$$

## Лекция 7. Численные методы решения задач нелинейного программирования (поиск экстремума функции одной переменной)

Напомним, что задача математического программирования формулируется следующим образом: найти значения переменных  $x_1, x_2, \dots, x_n$ , доставляющие максимум или минимум целевой функции  $y = f(x_1, x_2, \dots, x_n)$  при условиях  $g_j(x_1, x_2, \dots, x_n) = (\leq, \geq) b_j, (j = \overline{1, m})$ .

Особенность задач нелинейного программирования заключается в том, что функции  $\mathcal{U}$  и (или)  $g_j$  не линейны. Разработано множество численных методов решения задач нелинейного программирования.

### Классификация численных методов решения задач нелинейного программирования

1. Численные методы поиска экстремума функции одной переменной.

1.1. Классический метод.

1.2. Метод равномерного перебора.

1.3. Метод золотого сечения.

1.4. Метод Фибоначчи и т.д.

2. Численные методы поиска экстремума функции  $n$  – переменных.

2.1. Численные методы в задачах без ограничений.

2.1.1. Метод покоординатного спуска.

2.1.2. Метод Хука – Дживса.

2.1.3. Градиентный метод.

2.1.4. Метод Ньютона.

2.1.5. Метод сопряженных направлений и т.д.

2.2. Численные методы в задачах с ограничениями.

2.2.1. Метод покоординатного спуска.

2.2.2. Метод условного градиента.

2.2.3. Метод барьерных функций.

2.2.4. Метод штрафных функций.

2.2.5. Метод линеаризации и т.д.

Универсального метода, с помощью которого можно было бы решить любую задачу оптимизации, не существует. Поэтому для решения конкретной задачи применяют один или несколько своих численных методов.

### Методы поиска экстремума функции одной переменной

Эти методы применяются в однопараметрических задачах оптимизации. В них ищется один оптимальный параметр. Целевая функция – это функция одной переменной.

**Постановка задачи.** Найти значение переменной  $x$ , доставляющее минимум или максимум целевой функции  $y = f(x)$ , при условиях  $g_j(x) = (\leq, \geq) b_j, (j = \overline{1, m})$ .

### Классический метод минимизации (максимизации) функции одной переменной

Пусть  $a \leq x \leq b$ , функция  $f(x)$  непрерывна на этом отрезке и имеет на нем непрерывную производную. Вычисляют значение производной  $f'(x)$  и определяют **критические** точки, т.е. такие внутренние точки отрезка  $[a, b]$ , в которых производная обращается в нуль или не существует. В окрестности каждой такой критической точки исследуют знак производной и отбирают те из них, при переходе через которые производная меняет знак с минуса на плюс (это точки локального минимума) или с плюса на минус (это точки локального максимума). Затем вычисляют значения целевой функции в этих точках и на границах отрезка  $[a, b]$ . Эти значения сравнивают между собой и определяют точку, в которой достигается минимум (максимум) целевой функции. Эта точка является точкой глобального минимума (максимума) функции  $f(x)$  на отрезке  $[a, b]$ .

При решении реальных задач оптимизации данный метод применяется редко, т.к. зачастую производную целевой функции определить сложно или невозможно.

### Метод равномерного перебора

Пусть дана функция  $y = f(x) \rightarrow \min$  (см. рис 7.1).

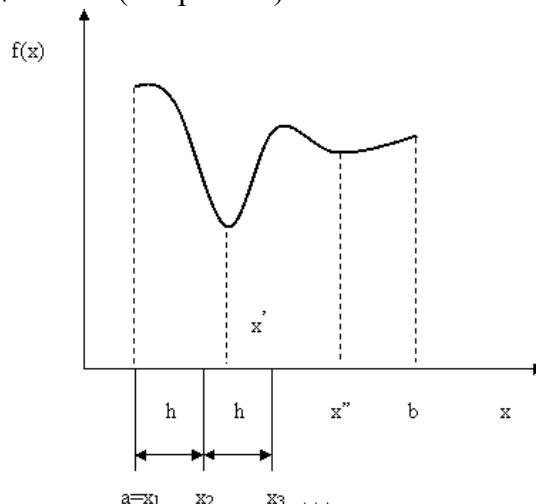


Рис.7.1. Графическая иллюстрация метода равномерного перебора

В соответствии с данным методом алгоритм поиска  $x_{opt}$  заключается в следующем. Фиксируют величину шага  $h > 0$ . Вычисляют значения целевой функции в точках  $x_1 = a$  и  $x_2 = x_1 + h - f(x_1)$  и  $f(x_2)$ . Полученные значения сравнивают. Запоминают меньшее из этих двух значений. Далее выбирается точка  $x_3 = x_2 + h$  и в ней вычисляются значения целевой функции  $f(x_3)$ . Сравнивается оставшееся на предыдущем шаге значение и значение  $f(x_3)$ . Наименьшее из них опять запоминают. Так поступают до тех пор, пока очередное значение  $x$  не превысит  $b$ . Последнее оставшееся значение является приближенным значением глобального минимума.

Трудности при использовании данного метода. Если целевая функция имеет узкую впадину, подобную приведенной на рисунке, то можно ее проскочить, и вместо точки глобального минимума определить точку локального минимума. Т.е. вместо  $x'$  можно найти  $x''$ . Эта проблема частично снимается, если выбрать очень маленький шаг, но при этом потребуется много времени (в том числе и машинного) для решения задачи.

### Метод золотого сечения

Рассматриваемая в данном методе функция должна быть **униmodalной**. Функция  $f(x)$  является **униmodalной** на отрезке  $[a, b]$ , если она на этом отрезке имеет единственную точку глобального минимума и слева от этой точки является строго убывающей, а справа строго возрастающей. Другими словами, функция  $f(x)$  униmodalна, если выполняются следующие соотношения (рис.7.2):

Суть метода золотого сечения заключается в том, чтобы определить точку глобального минимума на отрезке  $[a, b]$  за минимальное количество шагов, т.е. за минимальное количество вычислений целевой функции.

В соответствии с данным методом в каждый текущий момент времени рассматривается всегда две точки, например, в начальный момент точки  $x_1$  и  $x_2$  так, чтобы  $a < x_1 < x_2 < b$ . При этом возможен один из двух случаев (рис.7.3):

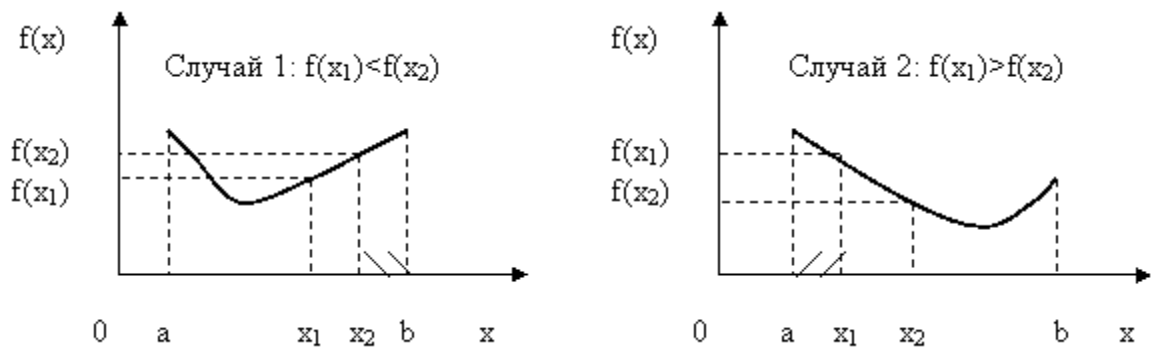


Рис.7.3. Иллюстрация обоснования исключения отрезков

Согласно свойству унимодальной функции в первом случае искомая точка  $x_{\min}$  не может находиться на отрезке  $[x_2, b]$ , во втором случае на отрезке  $[a, x_1]$  (показаны штриховкой). Значит, область поиска сужается, и следующую точку  $x_3$  необходимо брать на одном из укороченных отрезков:  $[a, x_2]$  - случай 1 или  $[x_1, b]$  - случай 2.

Теперь следует определиться, где на исходном отрезке  $[a, b]$  необходимо выбирать точки  $x_1$  и  $x_2$ . Первоначально ничего не известно о положении точки  $x_{\min}$  (графиков нет, и они не строятся, здесь мы их приводим для наглядной иллюстрации сути метода, при реальной оптимизации есть только выражение для целевой функции). Поэтому любой из приведенных выше случаев возможен с одинаковой вероятностью. Это означает, что лишним может оказаться любой из отрезков:  $[x_2, b]$  или  $[a, x_1]$ . Отсюда ясно, что точки  $x_1$  и  $x_2$  следует выбирать **симметрично** относительно середины отрезка  $[a, b]$ .

Далее для того, чтобы максимально сузить область поиска, эти точки должны быть поближе к середине исходного отрезка. Однако слишком близко к середине отрезка их тоже брать не следует, т.к. мы хотим построить алгоритм, для реализации которого необходимо **общее минимальное** количество вычислений целевой функции. Рассмотрим рис. 7.4.

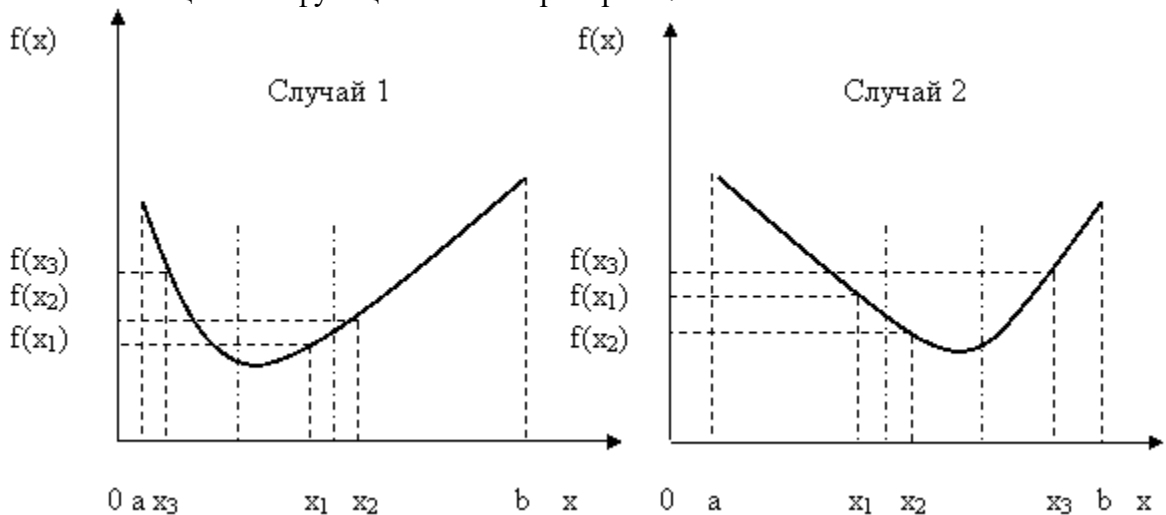


Рис.7.4. Иллюстрация обоснования расположения точек на отрезке

Выбирая на первом шаге сравниваемые точки слишком близко к середине отрезка  $[a, b]$ , мы исключим из рассмотрения большой отрезок  $[x_2, b]$  для случая 1 или  $[a, x_1]$  для случая 2. Но на втором шаге величина исключаемого отрезка **значительно уменьшится** (будет исключен отрезок  $[a, x_3]$  для случая 1 или отрезок  $[x_3, b]$  для случая 2).

Таким образом, с одной стороны, точки следует брать рядом с серединой отрезка, а, с другой стороны, слишком близко друг от друга их брать нельзя. Т.е. необходимо найти некую **«золотую середину»**. Для этого рассмотрим для простоты вместо отрезка  $[a, b]$  отрезок  $[0, 1]$  единичной длины – рис.7.5.

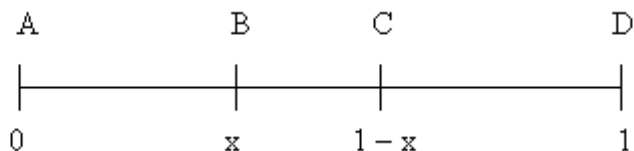


Рис.7.5. Обоснование «золотой середины» расположения точек на отрезке

На этом рисунке  $|AB| = |CD| = x$ ,  $|AC| = |BD| = 1 - x$ ,  $|BC| = 1 - 2x$ .

Для того, чтобы точка В была «выгодной» как на данном, так и на следующем этапе (шаге), она должна делить отрезок AD в таком же отношении, как и AC:  $AB/AD = BC/AC$ . При этом в силу симметрии аналогичным свойством будет обладать и точка С:  $CD/AD = BC/BD$ . В обозначениях координаты  $x$  эти пропорции принимают вид:  $x/1 = (1 - 2x)/(1 - x)$ . Решим эту пропорцию:

$$x(1 - x) = 1 - 2x \Rightarrow x - x^2 = 1 - 2x \Rightarrow x - x^2 - 1 + 2x = 0 \Rightarrow -x^2 + 3x - 1 = 0 \Rightarrow x^2 - 3x + 1 = 0, D = b^2 - 4ac = 9 - 4 = 5.$$

Корни этого уравнения равны:

$$x_1 = \frac{-b - \sqrt{D}}{2a} = \frac{3 - \sqrt{5}}{2} \cong 0,382, x_2 = \frac{-b + \sqrt{D}}{2a} = \frac{3 + \sqrt{5}}{2} > 1$$

не приемлем, т.е. уравнение имеет один корень.

О точке, которая расположена на расстоянии  $\frac{3 - \sqrt{5}}{2}$  длины от одного из концов отрезка, говорят, что она осуществляет «золотое сечение» данного отрезка.

Очевидно, что каждый отрезок имеет две такие точки, расположенные симметрично относительно его середины.

Итак, алгоритм метода «золотого сечения» заключается в следующем (см. также рис.7.6). На исходном отрезке  $[a, b]$  выбираются две точки  $x_1$  и  $x_2$ , так, чтобы выполнялось приведенное выше соотношение «золотого сечения» этого отрезка. Вычисляются значения целевой функции в этих точках –  $f(x_1)$  и  $f(x_2)$ . Они сравниваются, и из дальнейшего рассмотрения исключается отрезок, прилегающий к точке, дающей большее значение целевой функции (здесь отрезок  $[x_2, b]$ ). Т.е. исходный отрезок  $[a, b]$  «стягивается» до отрезка  $[a, b_1]$ . Для этого нового отрезка находится его середина, и по отношению к ней симметрично оставшейся точке  $x_1$  ставится точка  $x_3$ . Для нее рассчитывается значение целевой функции  $f(x_3)$  и сравнивается с  $f(x_1)$ . Из дальнейшего рассмотрения опять исключается отрезок, прилегающий к точке с большим значением целевой функции, здесь это отрезок  $[a, x_3]$ . Текущий отрезок «стягивается» до нового отрезка, здесь это  $[a_1, b_1]$  и т.д.

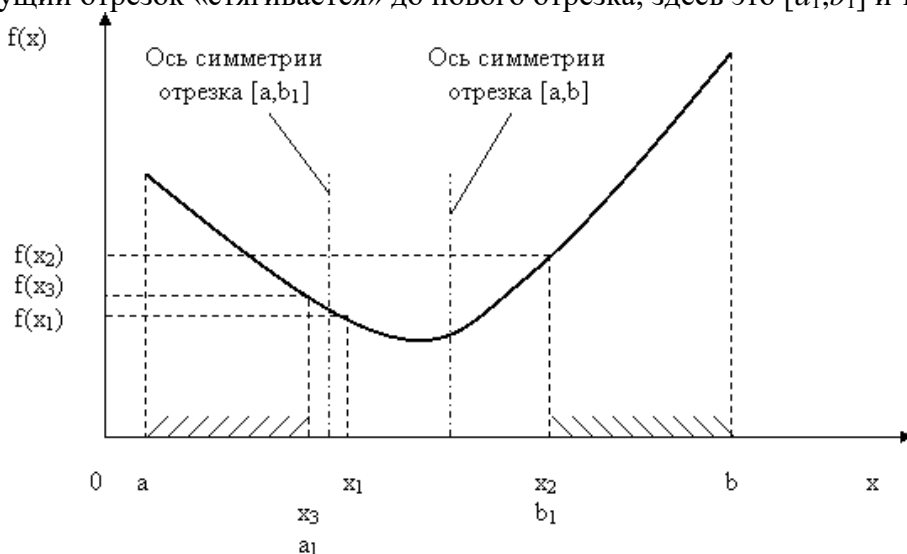


Рис.7.6. Иллюстрация алгоритма метода «золотого сечения»

Метод «золотого сечения» прост, эффективен и широко применяется в практической оптимизации.

## Лекция 8. Численные методы решения задач нелинейного программирования (поиск экстремума функции n – переменных)

### Метод линеаризации (приведения задачи нелинейного программирования к задаче линейного программирования)

Данный метод строго не относится к численным методам решения задач оптимизации. Но он эффективен и часто используется для решения практических задач. Рассмотрим суть данного метода на примере, который решался в лекции 5. Напомним формулировку задачи:

найти  $x_{1opt}$  и  $x_{2opt}$ . Целевая функция  $y = x_2 / x_1 \rightarrow \max$ , ограничения:

$$1 \leq x_1 \leq 8,$$

$$2 \leq x_2 \leq 12,$$

$$x_1 \cdot x_2 \geq 10.$$

**1 этап.** Приводим данную задачу к задаче линейного программирования. Для этого проводим логарифмирование ограничений и целевой функции:

$$\lg 1 \leq \lg x_1 \leq \lg 8;$$

$$\lg 2 \leq \lg x_2 \leq \lg 12;$$

$$\lg x_1 + \lg x_2 \geq \lg 10.$$

После вычислений получим:

$$0 \leq \lg x_1 \leq 0,903; \tag{8.1}$$

$$0,301 \leq \lg x_2 \leq 1,079; \tag{8.2}$$

$$\lg x_1 + \lg x_2 \geq 1. \tag{8.3}$$

После логарифмирования целевой функции:

$$\lg y = \lg x_2 - \lg x_1 \rightarrow \max.$$

Далее задача решается с применением симплекс - алгоритма или графо – аналитически (см. рис.8.1 и вычисления, сопровождающие построения). Для построения области допустимых решений (ОДР) в логарифмических координатах работаем с ограничениями (8.1) – (8.3). Ограничения (8.1) и (8.2) – это ограничения, графически представляющие собой прямые линии, параллельные соответственно осям " $0 - \lg x_2$ " и " $0 - \lg x_1$ ". Причем, левая ограничительная линия в ограничении (8.1) совпадает с осью " $0 - \lg x_2$ ". Ограничение (8.3) представляет собой прямую линию, наклонную под углом 45 градусов к осям, имеющая координаты пересечения осей «0-1». Для нахождения точки касания линии, соответствующей целевой функции, сначала строим «произвольную» линию для целевой функции, приравняв ее выражение к произвольному числу в данном масштабе. Приравняем выражение для целевой функции к числу «1,2»:  $\lg x_2 - \lg x_1 = 1,2 \Rightarrow \lg x_2 = 1,2 + \lg x_1$ .

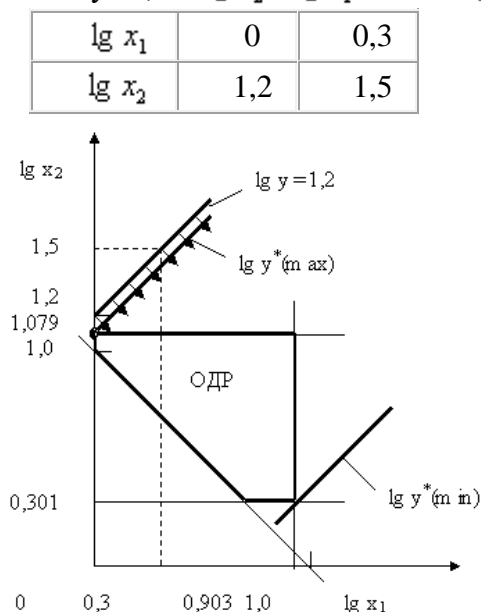


Рис.8.1. Графическая иллюстрация задачи оптимизации методом линейризации

Далее строим линию, параллельную данной линии и касающуюся границы ОДР. Находим координаты точки касания:

$$\lg x_{1ОЛТ} = 0 \Rightarrow x_{1ОЛТ} = 1; \lg x_{2ОЛТ} = 1,079 \Rightarrow x_{2ОЛТ} = 12.$$

Если целевая функция стремится к минимуму, т.е.  $\lg y = \lg x_2 - \lg x_1 \rightarrow \min$ , то прямая линия, соответствующая ей, коснется границы ОДР в точке с координатами:

$$\lg x_{1ОЛТ} = 0,903 \Rightarrow x_{1ОЛТ} = 8; \lg x_{2ОЛТ} = 0,301 \Rightarrow x_{2ОЛТ} = 2.$$

### Метод покоординатного спуска в задачах без ограничений

Это задача **безусловной минимизации**, т.е. задачи минимизации целевой функции  $y = f(x_1, x_2, \dots, x_n)$  на всем пространстве переменных (на всем евклидовом пространстве). Если требуется решить задачу максимизации, то выражение целевой функции умножают на (-1) и снова решается задача минимизации.

Суть данного метода заключается в построении последовательности точек  $x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(n)}$ , монотонно уменьшающих значение целевой функции  $f(x^{(0)}) \geq f(x^{(1)}) \geq f(x^{(2)}) \geq \dots \geq f(x^{(n)})$ .

Согласно этому методу направления спуска выбирается параллельно координатным осям, т.е. сначала спуск осуществляется вдоль первой оси  $OX_1$  затем вдоль второй оси  $OX_2$  и т.д. до последней оси  $OX_n$ .

Пусть  $x^{(0)}$  – начальная точка (см. рис. 8.2),  $a$  – некоторое положительное число. Вычисляют значение целевой функции в этой точке –  $f(x^{(0)})$ . Далее вычисляют значение целевой функции при  $x = x^{(0)} + a$  и проверяют выполнение неравенства

$$f(x^{(0)} + a) < f(x^{(0)}). \quad (8.4)$$

Если это неравенство справедливо, то вдоль направления оси  $OX_1$  значение функции  $f$  уменьшилось, и поэтому полагают  $x^{(1)} = x^{(0)} + a$ . Если неравенство (8.4) не выполняется, то делают шаг в противоположном направлении и проверяют выполнение неравенства

$$f(x^{(0)} - a) < f(x^{(0)}). \quad (8.5)$$

В случае выполнения этого неравенства полагают  $x^{(1)} = x^{(0)} - a$ . Если оба неравенства и (8.4), и (8.5) не выполняются, то  $x^{(1)} = x^{(0)}$ .

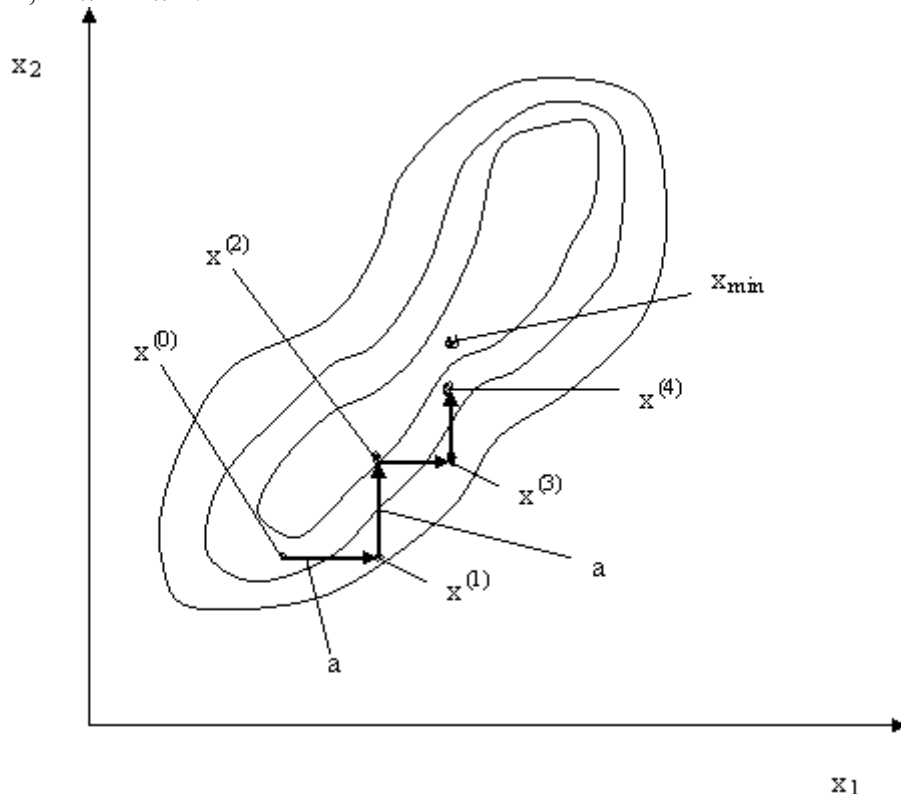




Рис.8.2. Графическая иллюстрация поиска точки минимума методом покоординатного спуска  
 Второй шаг производят вдоль координатной оси  $Ox_2$ . Вычисляют значение функции в точке  $(x^{(1)} + a)$  и сравнивают его с предыдущим значением, т.е. проверяют выполнение неравенства

$$f(x^{(1)} + a) < f(x^{(1)}). \quad (8.6)$$

Если это неравенство выполняется, то полагают  $x^{(2)} = x^{(1)} + a$ . Если оно не выполняется, то делают шаг в противоположном направлении и проверяют выполнение неравенства

$$f(x^{(1)} - a) < f(x^{(1)}). \quad (8.7)$$

В случае выполнения неравенства (8.7) считают, что  $x^{(2)} = x^{(1)} - a$ . Если оба неравенства и (8.6), и (8.7) не выполняются, то принимают  $x^{(2)} = x^{(1)}$ .

Так перебирают все  $n$  – направлений координатных осей. На этом первая итерация закончена. На  $n$  - м шаге будет получена некоторая точка  $x^{(n)}$ . Если  $x^{(n)} \neq x^{(0)}$ , то аналогично, начиная с  $x^{(n)}$  осуществляют вторую итерацию. Если же  $x^{(n)} = x^{(0)}$  (это имеет место, если на каждом шаге ни одно из пары неравенств не окажется выполненным), то величину шага нужно уменьшить, взяв, например,  $a_{n+1} = a_n/2$ , и в следующей итерации использовать новое значение величины шага.

Последующие итерации выполняют аналогично. На практике вычисления прекращают при выполнении какого – либо условия окончания счета, например

$$|f(x)_{(k+1)} - f(x)_{(k)}| < \delta,$$

где  $f(x)_{(k+1)}$  – значение целевой функции на  $(k+1)$  итерации;

$f(x)_{(k)}$  – значение целевой функции на  $k$  –ой итерации;

$\delta$  - некоторое положительное число, характеризующее точность решения исходной задачи минимизации целевой функции.

### Метод покоординатного спуска в задачах с ограничениями

Данный метод распространяется на задачи, с простыми ограничениями типа:

$$a_1 \leq x_1 \leq b_1; \quad (8.8)$$

$$a_2 \leq x_2 \leq b_2; \quad (8.9)$$

$$\dots$$

$$a_n \leq x_n \leq b_n. \quad (8.10)$$

Основные процедуры данного метода аналогичны предыдущему методу. Различие заключается в том, что наряду с проверкой выполнения неравенств  $f(x^{(0)} + a) < f(x^{(0)})$  ( $f(x^{(0)} - a) < f(x^{(0)})$ ),  $f(x^{(1)} + a) < f(x^{(1)})$  ( $f(x^{(1)} - a) < f(x^{(1)})$ ) и т.д. осуществляют проверку выполнения неравенств (8.8) – (8.10). Выполнение или невыполнение этих неравенств приводит к тем же последствиям, что и выполнение или невыполнение неравенств, приведенных выше.

## Лекция 9. Методы решения многокритериальных задач оптимизации

Эта задачи проектирования (оптимизации), в которых используется не один, а несколько критериев. На практике такие задачи возникают, когда проектируемый объект не может быть описан однокритериальной зависимостью, или объединить отдельные критерии в единый критерий не представляется возможным. Такое объединение критериев в единый критерий применяется, и оно будет рассмотрено ниже. Но это объединение, как правило, бывает формальным, искусственным. С математической точки зрения не существует идеального способа, метода решения таких задач. Каждый из них имеет преимущества и недостатки. Рассмотрим некоторые методы решения многокритериальных задач оптимизации.

### Метод поиска Парето – эффективных решений

Рассмотрим его суть на примере использования двух критериев. Критерии при использовании данного метода являются равнозначными.

Пусть имеется множество вариантов решения. По каждому из вариантов определены значения всех критериев. Представим множество оценок вариантов решения в пространстве критериев (рис.9.1).

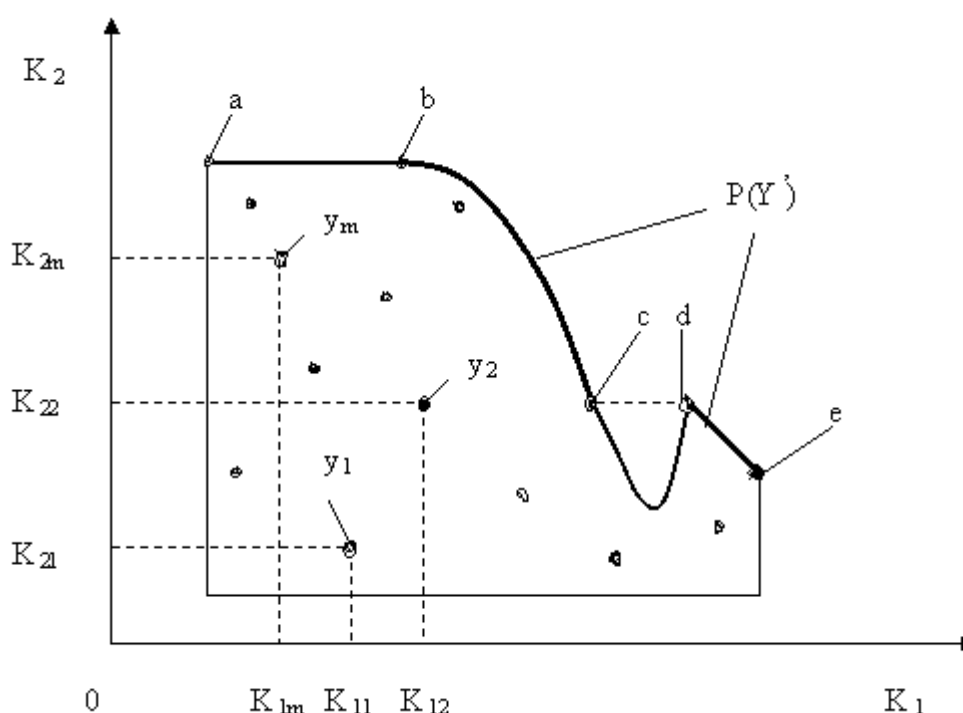


Рис.9.1. Иллюстрация поиска Парето – эффективных решений

На рис.9.1 приняты следующие обозначения:

$K_1$  и  $K_2$  – критерии оценки вариантов решения;

$Y = \{y_1, y_2, \dots, y_m\}$  – множество оценок альтернативных вариантов решения;

$K_{11}, K_{12}, \dots, K_{1m}$  – значения первого критерия для 1, 2, ..., m - го варианта решения;

$K_{21}, K_{22}, \dots, K_{2m}$  – значения второго критерия для 1, 2, ..., m - го варианта решения;

$P(Y)$  – множество Парето – эффективных оценок решений.

**Правило.** Множество Парето – эффективных оценок  $P(Y')$  представляет собой «северо – восточную» границу множества  $Y$  без тех его частей, которые параллельны одной из координатных осей или лежат в «глубоких» провалах .

Для случая, изображенного на рис.9.1, Парето – эффективные оценки состоят из точек кривой (bc), исключая точку (c), и линии (de).

Преимущества метода: 1) Критерии равнозначны; 2) Метод математически объективен.

Недостаток метода: 1) Одно окончательное решение получается только в частном случае, т.е. количество Парето – эффективных решений, как правило, более одного.

**Пример.** Имеется 10 вариантов машин, среди которых для проектируемого объекта необходимо выбрать наилучший. Машины оценены экспертами по двум показателям (критериям): производительности и надежности. Оценивание производилось по 11 - бальной шкале от 0 до 10. Результаты оценки машин приведены в таблице 9.1.

Таблица 9.1 Экспертные оценки машин по критериям производительности и надежности

Критерии	Оценки экспертов (баллы) для станков									
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Производительность (П)	6	4	10	3	10	0	2	4	6	7
Надежность (Н)	6	2	1	7	4	4	10	4	8	2

Представим множество оценок вариантов машин в пространстве критериев (рис.9.2): Парето – эффективными решениями здесь являются варианты машин C<sub>5</sub>, C<sub>7</sub> и C<sub>9</sub>.

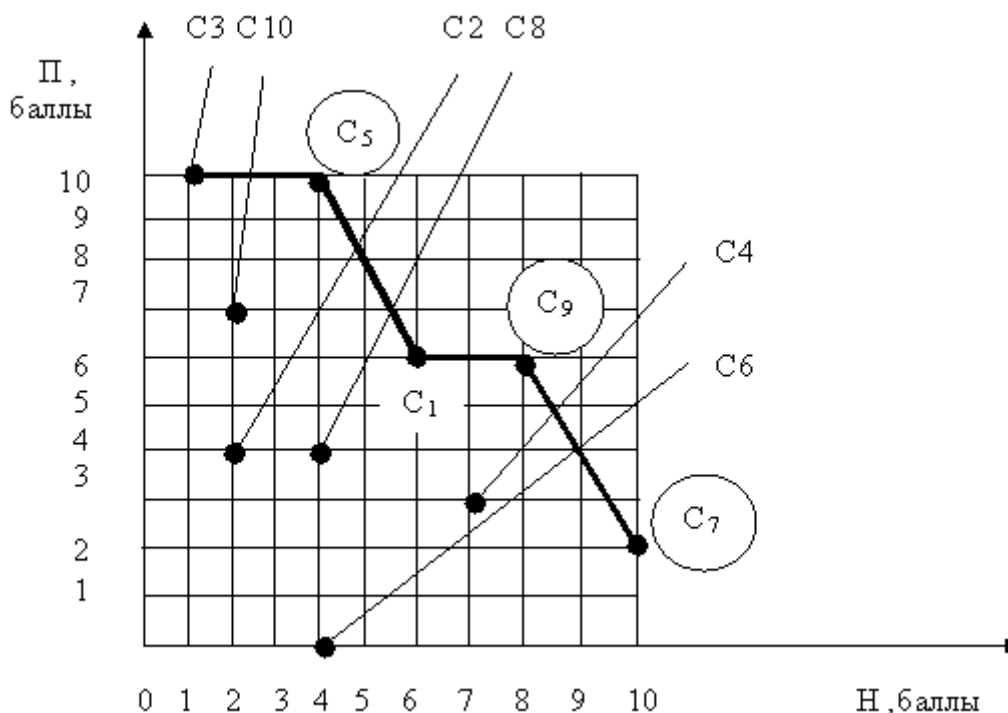


Рис.9.2. Пример поиска Парето – эффективных решений

### Метод решения многокритериальных задач оптимизации с использованием обобщенного (интегрального) критерия

Суть данного метода заключается в том, что частные критерии  $F_i(X), i = \overline{1, n}$  каким - либо образом объединяются в один интегральный критерий  $F(X) = \Phi(F_1(X), F_2(X), \dots, F_n(X))$ , а затем находится максимум или минимум данного критерия.

Если объединение частных критериев производится, исходя из объектной взаимосвязи частных критериев и критерия обобщенного, то тогда оптимальное решение будет корректно. Но такое объединение осуществить крайне сложно или невозможно, поэтому, как правило, обобщенный критерий есть результат чисто формального объединения частных критериев.

В зависимости от того, каким образом частные критерии объединяются в обобщенный критерий различают следующие виды обобщенных критериев:

1. Аддитивный критерий;
2. Мультипликативный критерий;
3. Максиминный (минимаксный) критерий.

### Аддитивный критерий

В них целевая функция получается путем сложения нормированных значений частных критериев. В общем виде целевая функция имеет следующий вид:

$$F(X) = \sum_{i=1}^n C_i \frac{F_i(X)}{F_i^{(0)}(X)} = \sum_{i=1}^n C_i f_i(X) \rightarrow \max(\min),$$

где  $n$  – количество объединяемых частных критериев;

$C_i$  – весовой коэффициент  $i$  – го частного критерия;

$F_i(X)$  – числовое значение  $i$  – го частного критерия;

$F_i^{(0)}(X)$  –  $i$  – й нормирующий делитель;

$f_i(X)$  – нормированное значение  $i$  – го частного критерия.

Частные критерии имеют различную физическую природу и поэтому различную размерность. А значит просто суммировать их некорректно. В связи с этим в предыдущей формуле числовые значения частных критериев делятся на некоторые нормирующие делители, которые назначаются следующим образом:

1. В качестве нормирующих делителей принимаются **директивные** значения параметров или критериев, заданные заказчиком. Считается, что значения параметров, заложенные в техническом задании, являются оптимальными или наилучшими.

2. В качестве нормирующих делителей принимаются максимальные (минимальные) значения критериев, достигаемые в области допустимых решений.

Размерности самих частных критериев и соответствующих нормирующих делителей одинаковы, поэтому в итоге обобщенный аддитивный критерий получается безразмерной величиной.

**Пример.** Определить оптимальный вариант машины с использованием обобщенного (интегрального) аддитивного критерия. Частными критериями, с помощью которых оценены варианты машины, являются ее производительность и надежность (наработка на отказ). Оба критерия «работают» на максимум, т.е. наилучшими вариантами машины являются те из них, которые обеспечивают наибольшую ее производительность и надежность. Исходные данные для решения задачи приведены в таблице 9.2.

Таблица 9.2. Исходные данные для определения оптимального варианта исполнения машины

Критерий $F_i$	Весовой коэффициент $C_i$	Значения критериев для вариантов исполнения машины		
		Вариант 1	Вариант 2	Вариант 3
Производительность ( $F_1$ ), шт/ч	0,6	1000	2000	4000
Надежность (наработка на отказ) ( $F_2$ ), ч.	0,4	1500	1000	500

Целевая функция на основе аддитивного критерия запишется следующим образом:

$$F(X) = C_1 \frac{F_1(X)}{F_1^{(0)}(X)} + C_2 \frac{F_2(X)}{F_2^{(0)}(X)} \rightarrow \max.$$

В качестве нормирующих делителей в данной задаче примем наилучшие (максимальные) значения частных критериев:

$$F_1^{(0)}(X) = 4000 \text{ шт/ч}, F_2^{(0)}(X) = 1500 \text{ шт/ч}.$$

Значения обобщенного аддитивного критерия рассчитываются для каждого варианта машины:  
Вариант 1.

$$F(X) = 0,6(1000/4000) + 0,4(1500/1500) = 0,55.$$

Вариант 2

$$F(X) = 0,6(2000/4000) + 0,4(1000/1500) = 0,558.$$

Вариант 3

$$F(X) = 0,6(4000/4000) + 0,4(500/1500) = 0,732.$$

**Оптимальным является 3 вариант машины**, т.к. ему соответствует максимальное значение обобщенного аддитивного критерия.

Один из недостатков этого метода заключается в том, что весовые коэффициенты назначает проектировщик. Разные проектировщики могут назначать разные весовые коэффициенты. Пусть, например,  $C_1 = 0,4$ ;  $C_2 = 0,6$ . Определим теперь значения аддитивных критериев для вариантов машины:

Вариант 1.

$$F(X) = 0,4 \cdot 0,25 + 0,6 \cdot 1 = 0,7;$$

Вариант 2.

$$F(X) = 0,4 \cdot 0,5 + 0,6 \cdot 0,67 = 0,602;$$

Вариант 3.

$$F(X) = 0,4 \cdot 1 + 0,6 \cdot 0,33 = 0,598.$$

Т.е. при таком изменении значений весовых коэффициентов **оптимальным уже будет 1 вариант машины**.

**Преимущество** данного метода: как правило, всегда удастся определить единственный оптимальный вариант решения.

**Недостатки:**

1. Трудности (субъективизм) в определении весовых коэффициентов.
2. Аддитивный критерий не вытекает из объектной роли частных критериев и поэтому выступает как формальный математический прием.
3. В аддитивном критерии происходит взаимная компенсация частных критериев, т.е. уменьшение одного из них может быть компенсировано увеличением другого критерия.

### Мультипликативный критерий

Целевая функция здесь записывается следующим образом:

$$F(X) = \prod_{i=1}^n C_i F_i(X) \rightarrow \max(\min),$$

где  $\Pi$  – знак произведения;

$C_i$  - весовой коэффициент  $i$ -го частного критерия;

$F_i(X)$  - числовое значение  $i$ -го частного критерия.

**Преимущества** мультипликативного критерия:

1. Не требуется нормирование частных критериев.
2. Практически всегда определяется одно оптимальное решение.

**Недостатки:**

1. Трудности (субъективизм) в определении весовых коэффициентов.
2. Перемножение разных размерностей.
3. Взаимная компенсация значений частных критериев.

### Максиминный (минимаксный) критерий

Эти критерии работают по принципу компромисса, который основывается на идее равномерности. Сущность принципа максимина заключается в следующем. При проектировании сложных систем, при наличии большого числа частных критериев установить между ними аналитическую взаимосвязь очень сложно. Поэтому стараются найти такие значения переменных (параметров)  $X = \{x_1, x_2, \dots, x_m\}$ , при которых нормированные значения всех частных критериев равны между собой:

$$C_i f_i(X) = K,$$

где  $C_i$  - весовой коэффициент  $i$ -го частного критерия;

$f_i(X)$  – нормированное значение  $i$ -го частного критерия;

$K$  – константа.

При большом количестве частных критериев из-за сложных взаимосвязей добиться выполнения указанного выше соотношения очень сложно. Поэтому на практике так варьируют значениями переменных проектирования  $x_1, x_2, \dots, x_m$ , при которых последовательно «подтягиваются» те нормированные критерии, численные значения которых в исходном решении оказались наименьшими. Т.к. эта операция производится в области компромисса, подтягивание «отстающего» критерия неизбежно приводит к снижению значений части остальных критериев. Но при проведении ряда шагов можно добиться определенной степени **уравновешивания противоречивых частных критериев, что и является целью принципа максимина.**

Формально принцип максимина формулируется следующим образом: выбрать такой набор переменных  $X^{(0)} \in X$ , при котором реализуется максимум из минимальных нормированных значений частных критериев, т.е.  $F(X^{(0)}) = \max \min f_i(X)$ .

Такой принцип выбора  $X^{(0)}$  иногда носит название гарантированного результата. Он заимствован из теории игр, где является основным принципом.

Если частные критерии необходимо минимизировать, то самым отстающим критерием является тот, который принимает максимальное значение. В этом случае применяют **принцип минимакса:**

$$F(X^{(0)}) = \min \max f_i(X)$$

### **Основные принципы выбора критериев оптимальности**

Выбор критериев – сложная задача, т.к. цели при проектировании любого объекта, как правило, противоречивы (обеспечение минимальной стоимости и максимальной надежности, максимальной производительности и минимальной энергоемкости и т.д.).

Если требуется оптимизировать один из показателей качества проектируемого объекта при соблюдении ограничительных требований на остальные показатели, то нужно сформировать один частный критерий. Задача оптимизации при этом сводится к задаче максимизации (минимизации) данного критерия с учетом заданных ограничений.

При наличии нескольких критериев выбирают:

а) **аддитивный критерий**, если существенное значение имеют абсолютные значения критериев при выбранном векторе параметров  $X$ ;

б) **мультипликативный критерий**, если существенную роль играет изменение абсолютных значений частных критериев при вариации вектора  $X$ ;

в) **максиминный (минимаксный) критерий**, если стоит задача достижения равенства нормированных значений противоречивых (конфликтных) частных критериев.

## Лекция 10 Математическое моделирование силового взаимодействия в зоне резания при изготовлении деталей на станках

Данная лекция открывает цикл лекций, посвященный математическому моделированию физических процессов в технологических системах.

Методы расчета сил резания при обработке деталей на металлорежущих станках рассматриваются в курсе «Резание металлов». Здесь в начале лекции мы лишь вспомним некоторые известные силовые зависимости. Итак, например, для расчета сил резания при точении используются следующие формулы:

$$\begin{aligned}P_X &= C_{P_X} t^{X_{P_X}} s^{Y_{P_X}} v^{Z_{P_X}} K_{ОБЩ}; \\P_Y &= C_{P_Y} t^{X_{P_Y}} s^{Y_{P_Y}} v^{Z_{P_Y}} K_{ОБЩ}; \\P_Z &= C_{P_Z} t^{X_{P_Z}} s^{Y_{P_Z}} v^{Z_{P_Z}} K_{ОБЩ},\end{aligned}$$

где  $P_X, P_Y, P_Z$  – соответственно, осевая, радиальная и тангенциальная составляющие силы резания

$C_{P_X}, C_{P_Y}, C_{P_Z}$  – константы (удельные силы резания), учитывающие влияние на силы резания всех прочих параметров, не входящих в формулы (геометрии инструмента, вида обрабатываемого материала, смазочно - охлаждающей технологической среды (СОТС) и т.д.);

$t, s, v$  - глубина, подача и скорость резания;

$x, y, z$  - показатели степени, выражающие влияние соответствующего параметра режима резания на силы резания;

$K_{ОБЩ}$  - коэффициент, учитывающий дополнительно влияние на силы резания свойств обрабатываемого, инструментального материала и т.д.

При сверлении для определения силовых параметров используются следующие формулы:

$$\begin{aligned}M_{KP} &= C_m D^{q_m} s^{y_m} K_p; \\P_0 &= C_p D^{q_p} s^{y_p} K_p.\end{aligned}$$

где  $M_{KP}, P_0$  – соответственно, крутящий момент и осевая сила резания;

$C_m, C_p$  - константы, зависящие от свойств обрабатываемого материала и вида обработки;

$D$  - диаметр обрабатываемого отверстия;

$s$  - подача;

$q, y$  - показатели степени, определяющие степень влияния соответствующего параметра на крутящий момент или осевую силу резания;

$K_p$  - коэффициент, зависящий от свойств обрабатываемого материала.

Эти формулы (математические модели) используются для практических расчетов. Они получены эмпирическим путем, т.е. на основании проведения экспериментов и обработки их результатов. Имеются и теоретически полученные уравнения, но они сложны и для практических расчетов не используются. Наша дальнейшая задача в данной лекции освоить порядок проведения силовых экспериментов и обработки их результатов с целью получения математических моделей.

### Порядок проведения силовых экспериментов и аппроксимации результатов измерений (получения математических моделей)

В качестве измерительных приборов при проведении силовых экспериментов для измерения сил резания используются динамометры. Их схемы и принципы работы приведены в специальной литературе и рассматриваются в соответствующих разделах курса «Резание металлов».

При установлении силовых зависимостей, как правило, проводят однофакторные эксперименты. В ходе таких экспериментов варьируют только тем параметром (фактором), влияние которого изучают, а все остальные факторы за время опыта остаются постоянными. Например, устанавливают влияние подачи на силу резания. Принимают разные значения подачи и измеряют соответствующее им значения силы резания. Все остальные параметры, а именно, обрабатываемый и инструментальный материалы, геометрические параметры резца, вид СОТС, глубину, скорость и т.д. оставляют неизменными.

**Конечная цель эксперимента** - установление функциональной связи (зависимости) между варьируемым параметром и силой резания и описание этой зависимости математической формулой. После установления всех частных зависимостей получают общую зависимость способом, приведенным ниже.

Каждый эксперимент состоит из ряда опытов. В каждом опыте принимаются разные значения исследуемого параметра. Для обеспечения требуемой точности эксперимента в каждом опыте проводится определенное количество повторений или дублей. Например, для обеспечения уровня надежности (доверительной вероятности) 0,9 количество дублей в каждом опыте должно быть не менее 5. Для обеспечения уровня надежности (доверительной вероятности) 0,95 количество повторений должно быть не менее 7.

Прежде чем математически обработать полученные экспериментальные данные значения силы, полученные в каждом опыте, усредняют. Обозначим значение силы через «у», а значение исследуемого параметра через «х». Т.е., например, «у» - это  $P_x$ , или  $P_y$ , или  $P_z$ , а «х» -  $s$ , или  $t$ , или  $v$ . Среднее значение силы в каждом опыте  $y$  определится следующим образом:

$$y_i = \frac{1}{m} \sum_{j=1}^m y_{j,i}, i = \overline{1, n},$$

где  $i$  – номер текущего опыта;

$m$  – количество повторений в  $i$  – ом опыте;

$j$  – номер текущего дубля или повторения;

$y_j$  – значение силы в  $j$  – м дубле;

$n$  – количество опытов в эксперименте.

После этого все данные сводятся в таблицу:

Номер опыта	1	2	.....	$n$
$x$	$x_1$	$x_2$	.....	$x_n$
$y$	$y_1$	$y_2$	.....	$y_n$

Полученные экспериментальные данные для получения формул подвергают следующей обработке. Сначала выбирают формулу для описания будущей зависимости. Важно правильно выбрать вид аппроксимирующей формулы («аппроксимация» - упрощение в том смысле, что зависимость между функцией «у» и аргументом «х» воплощена и в виде приведенной выше таблицы, но эта зависимость более просто будет описана в виде аппроксимирующей формулы).

Основные требования к аппроксимирующей формуле:

1. Максимальная точность описания устанавливаемой зависимости.
2. Простота
3. Обеспечение быстроты обработки экспериментальных данных.

Многолетние наблюдения показали, что зависимость силы резания от параметров резания хорошо описываются аппроксимирующей функцией вида:

$$y = Cx^k.$$

Задача состоит в том, чтобы определить неизвестные коэффициенты  $C$  и  $k$ . Для упрощения вычислений эту степенную зависимость логарифмируют:

$$\lg y = \lg C + k \lg x.$$

В логарифмических координатах эта функция изображается прямой линией – рис. 10.1.



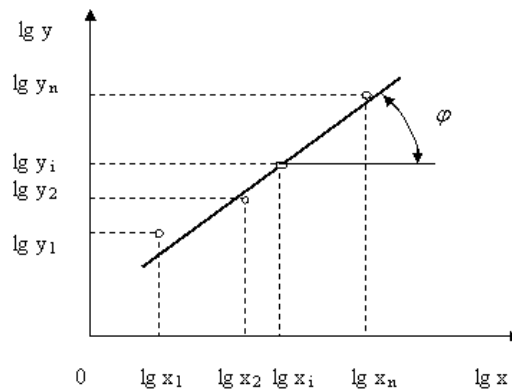


Рис.10.1. Аппроксимация экспериментальных данных в логарифмических координатах

Через опытные точки проводят прямую линию так, чтобы возможно большее число точек равномерно группировалось вокруг нее. Тогда  $k = \operatorname{tg} \varphi$  (тангенсу угла наклона прямой к оси абсцисс). Далее берется опытная точка на прямой, и определяется постоянная  $C$ :  $\lg C = \lg y_i - k \lg x_i$ . После потенцирования находится явное значение постоянной  $C$ . Приведенный способ применяют, тогда, когда зависимость  $y = f(x)$  монотонна и экспериментальные точки плотно группируются вокруг прямой. Только в этом случае возможно обеспечить требуемую точность. Если приведенные выше условия не выполняются, то для обработки экспериментальных данных применяют **метод наименьших квадратов (метод Гаусса)**. Для его использования применяют исходную формулу или ее опять логарифмируют. В последнем случае после логарифмирования проще решать уравнения. В настоящее время при реализации метода наименьших квадратов на компьютере надобность в логарифмировании исходной формулы практически отпала.

#### Аналитическая обработка экспериментальных данных методом наименьших квадратов

Для понимания сути данного метода рассмотрим сначала рис.10.2. На нем в обычных и логарифмических координатах изображены опытные значения (точки) и, соответственно, аппроксимирующая кривая и аппроксимирующая прямая.

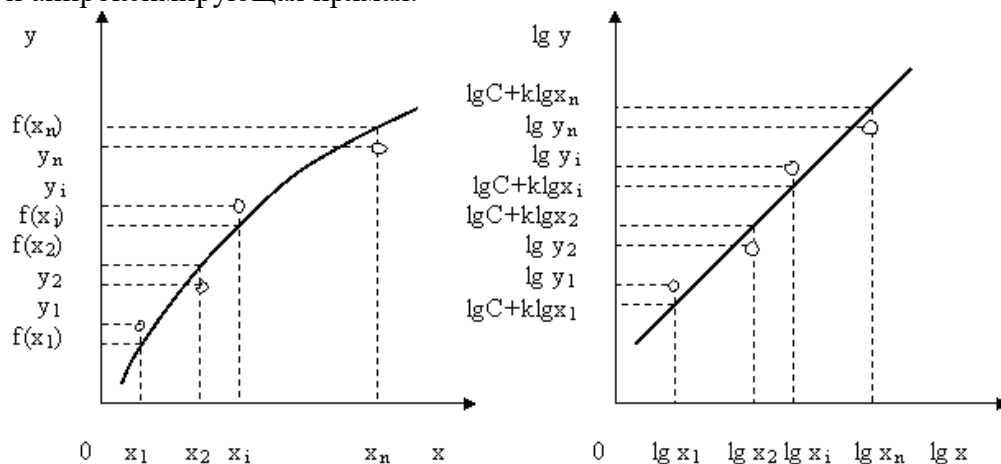


Рис.10.2. Иллюстрация реализации метода наименьших квадратов

На данном рисунке  $y_1, y_2, \dots, y_n$  – экспериментальные значения,  $f(x_1), f(x_2), \dots, f(x_n)$  – расчетные значения аппроксимирующей функции. Аналогичные значения, но только в логарифмическом измерении, приведены на правой части рисунка.

В основе метода наименьших квадратов лежит следующее положение: наилучшее приближение аппроксимирующей функции  $y = f(x)$  к экспериментальным данным будет в том случае, когда сумма квадратов отклонений расчетных значений  $f(x_1), f(x_2), \dots, f(x_n)$  от экспериментальных данных  $y_1, y_2, \dots, y_n$ , является минимальной, т.е.

$$S = \sum_{i=1}^n [y_i - f(x_i)]^2 = \min$$

или

$$S' = \sum_{i=1}^n [\lg y_i - (\lg C + k \lg x_i)]^2 = \min.$$

Разность  $[y_i - f(x_i)]$  в выражении для  $S$  есть отклонение по ординате  $i$ -ой экспериментальной точки от заменяющей (аппроксимирующей) кривой. Квадраты отклонений берутся, чтобы компенсировать знаки «-» отклонений.

Сумма  $S'$  будет минимальной, если ее частные производные по параметрам  $C$  и  $k$  равны нулю. Произведя дифференцирование и соответствующие преобразования, получают систему нормальных уравнений, которая затем решается для нахождения искомой постоянной  $C$  и показателя степени  $k$ . Существуют компьютерные программы для обработки экспериментальных данных с целью получения аппроксимирующих функций, реализующие метод наименьших квадратов.

Выражение  $y = Cx^k$  является математической моделью объекта исследования (здесь процесса резания). Поэтому после получения численных значений  $C$  и  $k$  необходимо проверить степень соответствия (адекватность) принятой математической модели описываемому объекту. Проверку адекватности производят, например, по  $F$ -критерию Фишера. Если принятая аппроксимирующая функция не удовлетворяет критерию Фишера, то она должна быть заменена другой.

После получения частных зависимостей их объединяют в общую зависимость. **Пример объединения двух частных зависимостей.** Пусть в результате проведения двух серий экспериментов получены следующие зависимости:

$$\begin{aligned} P_Z &= C_t t^x, \\ P_Z &= C_s s^y. \end{aligned}$$

Обе серии экспериментов проводились при неизменных обрабатываемом и инструментальном материалах, скорости резания, СОЖ и т.д. Кроме этого первая серия экспериментов была выполнена при постоянной подаче  $s = s \text{ const}$ , а вторая при постоянной глубине резания  $t = t \text{ const}$ . Общая формула, выражающая одновременно влияние  $t$  и  $s$  на силу  $P_Z$ , имеет вид:

$$P_Z = C_{ts} t^x s^y.$$

Здесь неизвестна постоянная  $C_{ts}$ , которая описывает влияние на силу  $P_Z$  всех факторов процесса резания, оставшихся постоянными при проведении обеих серий экспериментов, т.е. всех факторов, кроме подачи и глубины резания. Общая формула превращается в частные, если в нее подставить соответственно или  $s_{const}$ , при которой проводилась первая серия экспериментов, или  $t_{const}$ , при которой проводилась вторая серия экспериментов:

$$\begin{aligned} C_t &= C_{ts} s^y_{const}, \\ C_s &= C_{ts} t^x_{const}. \end{aligned}$$

Используя данные первой и второй серии опытов, получим:

$$\begin{aligned} C_{ts(1)} &= \frac{C_t}{s^y_{const}}, \\ C_{ts(2)} &= \frac{C_s}{t^x_{const}}. \end{aligned}$$

Вследствие неизбежных погрешностей экспериментов величины  $C_{ts}$ , полученные из первой и второй серии опытов, будут отличаться друг от друга. Поэтому окончательно принимают:

$$C_{ts} = \frac{C_{ts(1)} + C_{ts(2)}}{2}.$$

Подобным же образом объединяют три и более частных зависимости.

Полученные формулы не являются физическими законами. Они получены на основе обработки вполне реальных экспериментальных данных, поэтому имеют определенную область адекватности. Т.е. за пределами этой области их использовать нельзя. Например, если силовая зависимость была получена для подач от 0,1 мм/об до 0,5 мм/об, то при подачах меньше 0,1 или больше 0,5 мм/об эту формулу использовать нельзя. Всякая экстраполяция за пределами области адекватности может привести к существенным ошибкам.

## Лекция 11. Математическое моделирование упругих деформаций в технологической системе

При обработке деталей на металлорежущих станках сам станок, приспособление, инструмент и элементы его крепления, обрабатываемая заготовка представляют собой упругую систему, которая называется **упругой системой СПИД**.

При обработке сила резания изменяется под действием переменных условий обработки, а именно неравномерности глубины резания, колебания твердости обрабатываемого материала и т.д. Колебания силы резания приводят к упругим деформациям деталей системы СПИД и смещению этих деталей по отношению друг к другу за счет наличия зазоров в соединениях. Кроме этого упругая система СПИД обладает не бесконечной жесткостью, причем эта жесткость также является случайной величиной, т.е. может изменять свое значение от наименьшего до наибольшего. Все это вместе взятое оказывает влияние на формирование точности обработки.

Формирование упругих перемещений и смещения элементов в технологической системе и влияния их на точность обработки могут быть представлены следующей схемой – см. рис. 11.1.

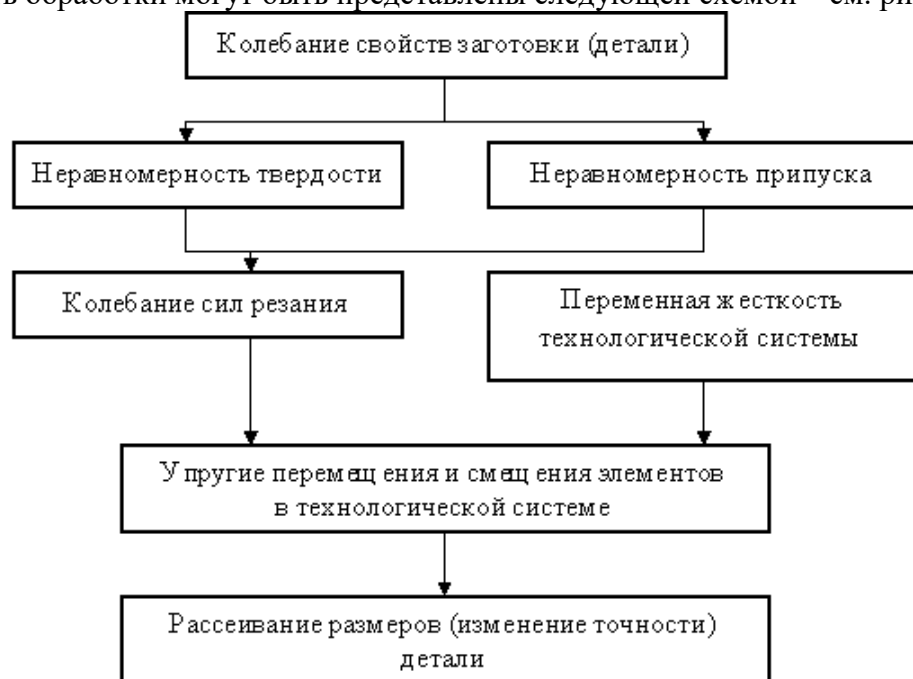


Рис.11.1. Схема формирования упругих перемещений и смещения элементов в технологической системе и влияния их на точность обработки

**Жесткость упругой системы СПИД** – это отношение составляющей силы резания, направленной по нормали к обрабатываемой поверхности, к смещению лезвия инструмента, установленного на размер в данном направлении.

Таким образом, жесткость системы  $j_c$  равна:

$$j_c = \frac{F_y}{y}, \text{ Н/мм},$$

где  $F_y$  – радиальная составляющая силы резания;

$y$  – смещение режущей кромки инструмента (деформация упругой системы).

Аналогично выражается жесткость отдельных элементов технологической системы: для суппорта –  $j_{\text{суп}}$ , для передней бабки –  $j_{\text{пб}}$  и т.д.

Жесткость может быть выражена также отношением приращения силы резания  $\Delta F_y$  к приращению перемещения  $\Delta y$ :

$$j_c = \frac{\Delta F_y}{\Delta y}, \text{ Н/мм}.$$

В практических расчетах иногда используют величину, обратную жесткости – податливость  $\omega$ .

$$\omega = \frac{1}{jc} = \frac{y}{Py}, \text{ мм/Н}$$

Величина деформации упругой системы равна:

$$y = \frac{Py}{jc}, \text{ мм}$$

Для определения **статической жесткости** узла станка производят ряд нагружений, возрастающих от нуля до некоторого предельного значения, которое определяется размерами станка. Для каждого значения нагрузки измеряют смещение узла в направлении действия силы. Затем производят разгрузку этого узла, постепенно уменьшая нагрузку и фиксируя остаточные отжатия. При нагружении и разгрузке строятся зависимости  $y = f(Py)$  – рис.11.2.

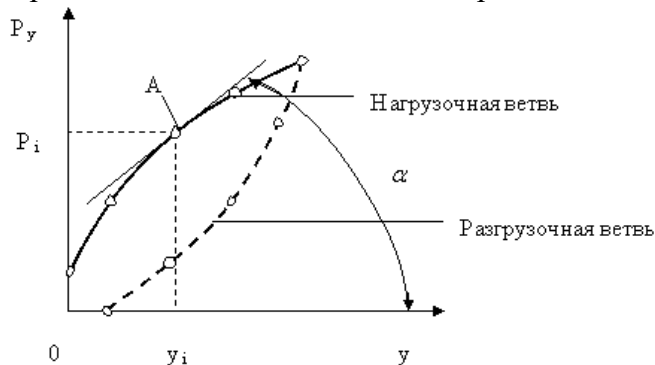


Рис.11.2. Зависимости  $y = f(Py)$  при определении статической жесткости узла станка

При этом нагрузочная и разгрузочная ветви характеристики жесткости обычно не совпадают, т.е. имеет место **гистерезис**. Площадь петли гистерезиса – работа, затраченная на преодоление сил трения, контактных деформаций и т.д. При повторных циклах нагрузки площадь петли гистерезиса уменьшается вследствие выборки зазоров в стыках деталей при первом нагружении и уменьшения пластического деформирования.

Для оценки жесткости в любой точке (например, в точке А) кривой  $Py = f(y)$  необходимо найти производную в данной точке

$$y' = j_A = \frac{dPy}{dy} = tg\alpha.$$

Численно производная равна тангенсу угла наклона касательной к кривой в точке А.

Жесткость станка зависит от его конструкции и от качества сборки. Для многих станков жесткость и податливость регламентированы ГОСТами.

Как показывает практика, статическая жесткость станков в 1,2 – 1,4 раза выше, чем жесткость работающих станков. Поэтому более объективной величиной является динамическая жесткость станков, которая определяется с помощью произвольного метода, который основан на обработке заготовок с переменным припуском и последующих расчетах.

**Пример.** Определение динамической жесткости токарного станка (рис.11.3).

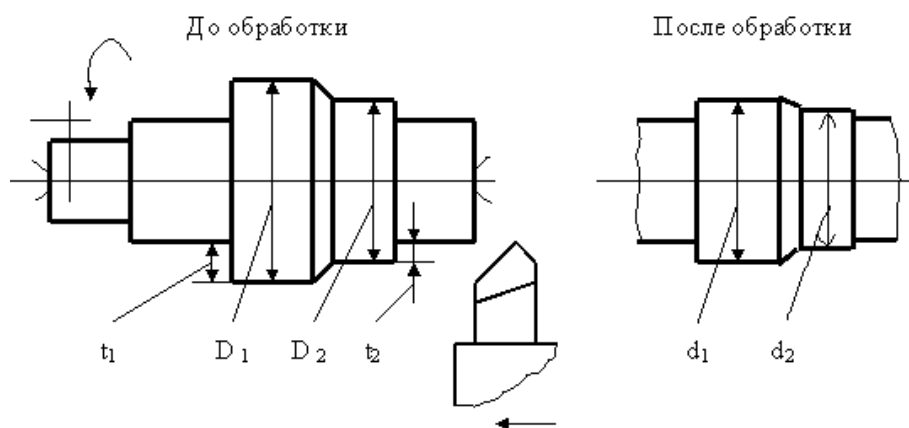


Рис.11.3. Схема испытания токарного станка для определения его динамической жесткости

Пояски диаметрами  $D_1$  и  $D_2$  обрабатываются за один рабочий ход при неизменных условиях (подаче, скорости резания и т.д.). Уступ  $\Delta_{\text{дет}} = d_1 - d_2$ , образовавшийся вследствие разных глубин резания и, следовательно, разных отжатий в технологической системе, будет меньше  $\Delta_{\text{заз}} = D_1 - D_2$  (т.е.  $\Delta_{\text{дет}} < \Delta_{\text{заз}}$ ).

Отношение  $\varepsilon = \Delta_{\text{заз}} / \Delta_{\text{дет}}$  называется уточнением. Оно показывает, во сколько раз в результате обработки уменьшилась погрешность заготовки. Уточнение характеризует жесткость технологической системы:

$$j = \lambda C_p s^{0,75} \varepsilon,$$

где  $\lambda = P_y / P_x; C_p$  - константа (постоянная силы резания).  $s$  - подача.

Жесткость станка зависит и от степени его нагрева. Так, например, по наблюдениям профессора А.В.Худобина жесткость шпиндельной бабки круглошлифовального неработающего станка составляет 15100 Н/мм. После 30 мин его работы на холостом ходу жесткость достигла 22400 Н/мм, т.е. увеличилась на 44%.

Жесткость узлов станков с вращающимися деталями не одинакова в радиальном направлении. Например, по данным Г. Б. Фикс – Марголина для токарных станков характерны следующие круглограммы жесткости шпиндельной группы ( $j$ ) и упругих перемещений шпинделей ( $y$ ), подобные приведенным на рис.11.4.

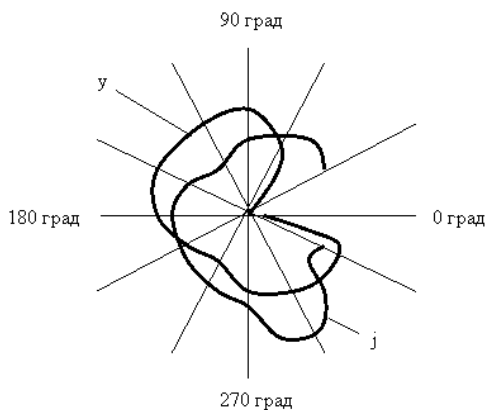


Рис.11.4. Круглограммы жесткости шпиндельной группы ( $j$ ) и упругих перемещений шпинделей ( $y$ ) токарных станков

Учитывая, что жесткость технологической системы переменна, и ее значение зависит от многих факторов, проявление которых невозможно предвидеть заранее, ее характеристику следует отнести к разряду **случайных величин**.

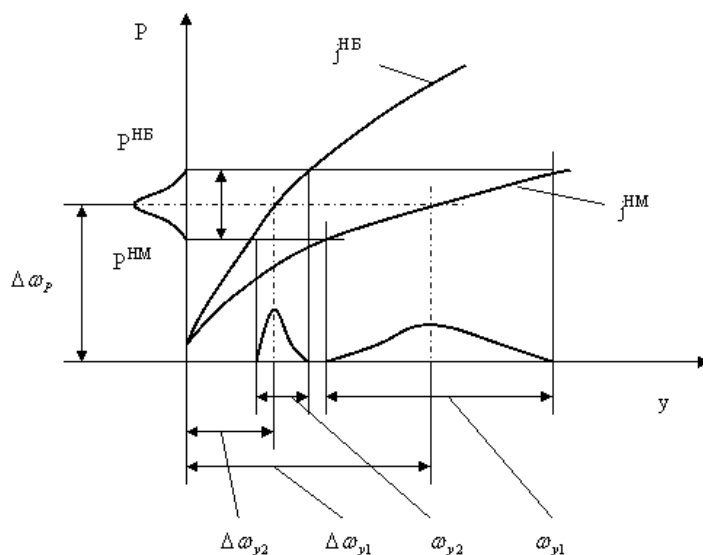


Рис.11.5. Иллюстрация формирования полей рассеяния упругих перемещений в технологической системе

Вследствие рассеяния значений припусков и характеристик свойств материалов заготовок, например, твердости, значения силы резания будут также рассеиваться от  $P^{HM}$  до  $P^{HB}$ . Если при этом учесть, что значения жесткости технологической системы случайны и подлежат рассеянию от  $j^{HM}$  до  $j^{HB}$ , то даже при стабильном рассеянии значений силы резания, значения поля рассеяния  $\omega_y$  и координаты его середины  $\Delta \omega_y$  не будут оставаться постоянными – рис.11.5. Участок диаграммы между кривыми  $j^{HM}$  и  $j^{HB}$  на этом рисунке представляет собой поле  $\omega_j$  рассеяния значений жесткости технологической системы.

## Лекция 12. Математическое моделирование точности обработки деталей на станках

### Основные факторы, определяющие погрешность обработки деталей

Наличие и формирование погрешности обработки деталей на металлорежущих станках определяются следующими факторами:

1. недостаточная жесткость станка;
2. колебание силы резания из – за неравномерности твердости заготовки и припусков на обработку;
3. неточность станков;
4. погрешности управляющих программ (для станков с ЧПУ, наиболее существенные – погрешности аппроксимации);
5. погрешность установки (базирования и закрепления) заготовки на станке;
6. погрешности настройки инструмента и станка на размер;
7. погрешности обработки, вызванные неточностью инструмента и его износом;
8. тепловые деформации;
9. ошибки рабочего (наладчика, станочника) и т.д.

Действие всех этих факторов, влияющих на точность обработки, приводит к возникновению суммарной погрешности обработки.

### Расчетно-аналитический метод определения точности обработки

Согласно этого метода, элементарные погрешности  $\delta_i$  определяемые действием каждого из приведенных выше факторов, принимаются практически независимыми друг от друга. Их суммирование производят по вероятностному методу:

$$\delta_{\Sigma} = \sqrt{\sum_{i=1}^n (k_i * \delta_i)^2},$$

где  $n$  – количество учитываемых погрешностей;

$k_i$  – коэффициент, учитывающий закон распределения  $i$  – ой погрешности ( $k_i = 1,0 - 1,73$ ) для разных законов распределения, например, для нормального закона распределения  $k = 1$ );

$\delta_i$  -  $i$  – я элементарная погрешность обработки.

**Пример:** расчет одной из составляющих погрешностей обработки, вызванной неточностью настройки инструмента на размер для станков с ЧПУ.

Современные приборы для настройки инструментов на размер имеют высокую разрешающую способность: цена деления координатных шкал 1мкм и оптическое увеличение проектора до 30раз. Однако, сколь бы высокой ни была точность исполнения прибора, инструмент всегда настраивается с некоторыми отклонениями. Они складываются из погрешностей самого прибора ( $\delta_1 - \delta_5$ ) и погрешностей установки на станке настроенного на размер инструмента ( $\delta_6 - \delta_8$ ).

По правилам сложения случайных величин погрешность положения вершины настроенного на размер инструмента равна:

$$\delta_{\Sigma} = \sqrt{(k_1 \delta_1)^2 + (k_2 \delta_2)^2 + (k_3 \delta_3)^2 + (k_4 \delta_4)^2 + (k_5 \delta_5)^2 + (k_6 \delta_6)^2 + (k_7 \delta_7)^2 + (k_8 \delta_8)^2},$$

где  $k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8$  – коэффициенты, учитывающие законы распределения погрешностей;

$\delta_1$  - погрешность шкал отчета прибора;

$\delta_2$  - погрешность отсчета размера по шкалам;

$\delta_3$  - неточность совмещения вершины инструмента с перекрестием экрана проектора;

$\delta_4$  - не совмещение начал отсчета шкал и устройства для крепления инструмента;

$\delta_5$  - погрешность от неточности углового расположения на приборе устройства для крепления инструмента;

$\delta_6$  - несовпадение нуля отсчета координат инструмента с теоретическим положением из – за неточности расположения поверхностей, базирующих инструмент на станке;

$\delta_7$  - погрешность от неправильного углового расположения на станке базирующих поверхностей;

$\delta_8$  - погрешность в связи с деформациями элементов, участвующих в зажиме инструмента.

Пусть

$$\delta_1 = \delta_2 = 1 \text{ мм}, \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = \delta_8 = 3 \text{ мм}, \kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \kappa_5 = \kappa_6 = \kappa_7 = \kappa_8 = 1.$$

Тогда

$$\delta_{\Sigma} = \sqrt{1^2 + 1^2 + 3^2 + 3^2 + 3^2 + 3^2 + 3^2 + 3^2} = 7,48 \text{ мм}$$

В технологических справочниках приведены среднестатистические данные по точности обработки детали на станках. Все они получены статистическими методами на основе обработки результатов экспериментов. Основной недостаток определения точности обработки по среднестатистическим данным состоит в том, что при этом невозможно учесть индивидуальные особенности конкретных станков. Станки даже одной модели отличаются друг от друга по точности из-за различного износа, разной точности сборки и т.д. В идеальном случае необходимо иметь данные по точности каждого станка, причем эти данные должны периодически обновляться. Т.е. нужны математические модели, представляющие точность обработки деталей на станках, реальные, учитывающие динамические процессы, протекающие в конкретной технологической системе СПИД.

### Моделирование точности обработки деталей на основе динамических характеристик станков

Материалы этой части лекции взяты из работ к.т.н., доцента кафедры «Технология автоматизированного машиностроения» ИГЭУ Минеева А.С. и его дипломника Таланова П.В. Рассмотрим эти вопросы на примере исследования точности плоскошлифовального станка.

Сначала разрабатывается расчетная схема упругой системы станка, которая представляет собой многомассовую модель. Для этого станок разбивается на ряд узлов, каждый из которых представляет собой отдельную массу. Кроме этого какая-либо деталь станка принимается как деталь условно бесконечной массы, к которой «привязываются» все остальные массы. Расчетная схема упругой системы плоскошлифовального станка представляет собой четырехмассовую модель. В качестве детали условно бесконечной массы принята станина, к которой «привязаны» все остальные массы – см. рис. 12.1.

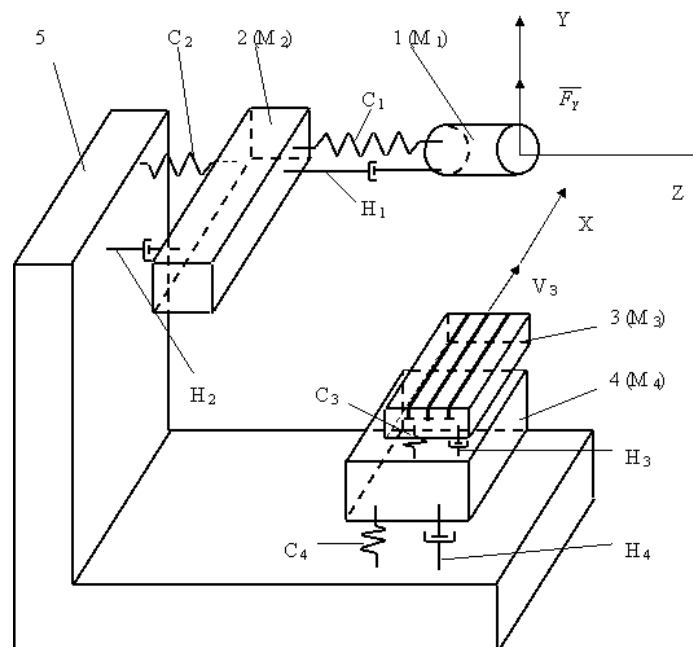


Рис.12.1. Расчетная схема упругой системы плоскошлифовального станка

На данном рисунке приняты следующие обозначения: 1.шпиндель; 2.шпиндельная бабка; 3.продольный стол; 4.поперечный стол; 5.станина;



$M_1, M_2, M_3, M_4$  – приведенные массы шпинделя, шпиндельной бабки, продольного стола, поперечного стола;

$C_1, C_2, C_3, C_4$  – жесткости;  $H_1, H_2, H_3, H_4$  – коэффициенты демпфирования;

$\vec{F}_y$  – вектор радиальной составляющей силы резания;

$V_3$  – скорость продольного перемещения заготовки.

Взаимодействие между отдельными массами описывается, как правило, двумя видами связей:

1. упругими связями, которые характеризуются жесткостью;
2. демпфирующими связями, которые характеризуются коэффициентами демпфирования.

Чтобы лучше понять характер взаимодействия масс, можно привести пример работы подвески автомобиля. В ней пружина осуществляет упругую связь, а амортизатор является демпфером.

Иногда, рассматривая взаимодействие масс, учитывают третью связь, которая называется инерционной, которая, как и две предыдущие также зависит от массы узлов. Все три вида связей зависят и от величины перемещения узла, от его скорости и ускорения. Инерционные связи в данной модели мы рассматривать не будем.

Здесь исследуется механизм образования непрямолинейности, поэтому считают, что данная четырехмассовая упругая система имеет свободу перемещения вдоль оси  $Y$ , следовательно, в качестве возбуждающей силы будет рассматриваться только радиальная составляющая сила резания. Т.к. смещение упругой системы необходимо рассматривать только вдоль оси  $Y$ , приведению выше на рис.12.1 расчетную схему нужно заменять эквивалентной схемой.

Динамическая модель эквивалентной упругой системы плоскошлифовального станка приведена на рис.12.2.

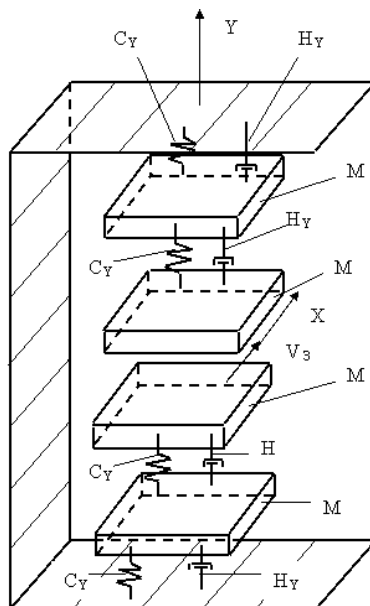


Рис.12.2. Динамическая модель эквивалентной упругой системы плоскошлифовального станка

На данном рисунке приняты следующие обозначения:

$C_{Y1}, C_{Y2}, C_{Y3}, C_{Y4}$  – жесткости в направлении оси  $Y$ ;

$H_{Y1}, H_{Y2}, H_{Y3}, H_{Y4}$  – коэффициенты демпфирования в направлении оси  $Y$ .

Основой построения математической динамической модели является уравнение колебаний одномассовой упругой системы (рис.12.3).

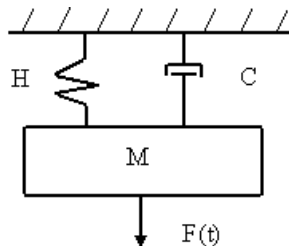


Рис.12.3. Динамическая одномассовая модель

Это уравнение представляет собой дифференциальное уравнение второго порядка:

$$M\ddot{x} + H\dot{x} + Cx = F(t) - U(t).$$

На рис.12.3 и в данном уравнении приняты следующие обозначения:  $M$  – масса,  $H$  – коэффициент демпфирования,  $C$  – жесткость,  $F(t)$  – возбуждающая сила, действующая в данном направлении,  $t$  – время,  $x$  - перемещение,  $\dot{x}$  - скорость,  $\ddot{x}$  - ускорение,  $U(t)$  - «обратная сила».

Смысл понятия «обратная сила» заключается в следующем – см. рис. 12.4 ( на нем буквой «З» обозначена заготовка).

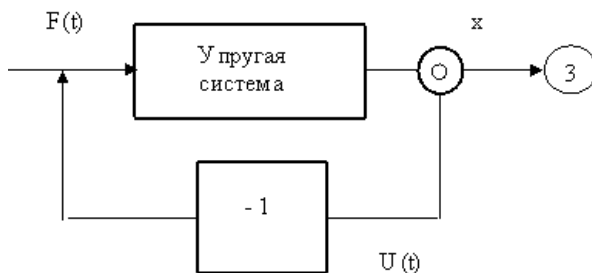


Рис.12.4. Иллюстрация понятия «обратная сила»

Под действием силы  $F(t)$  возникает смещение  $x$  упругой системы станка и, как следствие, уменьшение глубины резания, отсюда сила  $F(t)$  уменьшается на величину  $U(t)$ .

Итак, необходимо определить геометрическую сумму смещений  $\Delta_Y$  эквивалентной упругой системы вдоль оси  $Y$ , которое определяет суммарное смещение оси вращения шпинделя от прямолинейного движения

$$\Delta_Y = \Delta_{Y_{\max}} - \Delta_{Y_{\min}}.$$

Имея данные об упругих перемещениях несущей системы (шпинделя относительно детали) за  $n$  – е количество оборотов шпинделя можно определить величину непрямолинейности поверхности детали, получаемую после шлифования.

При шлифовании поверхностей на плоскошлифовальных станках образуется погрешность профиля плоской поверхности в продольном сечении (непрямолинейность), которая имеет следующую природу – см. рис. 12.5.

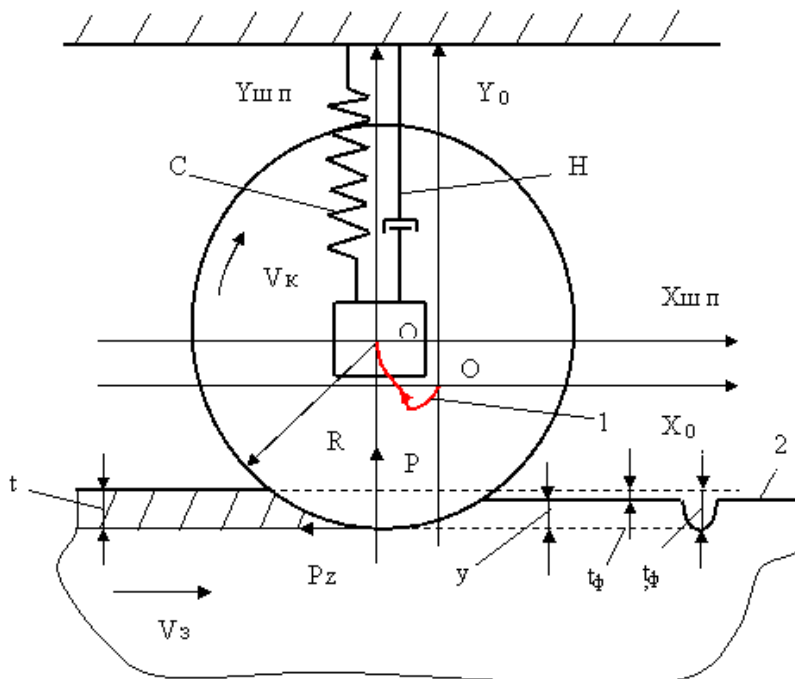


Рис.12.5. Механизм образования непрямолинейности шлифуемой поверхности

На рис.12.5: 1 – траектория упругих приведенных перемещений оси вращения шпинделя, 2 – профиль продольного сечения плоской поверхности.

Деталь рассматривается как неподвижная, а все упругие перемещения узлов станка приводятся к шпинделю. В системе координат шпинделя  $X_{шп}O_1Y_{шп}$  траектория перемещения торца шлифовального круга имеет форму идеальной окружности радиусом  $R$ , центр которой лежит на оси вращения шпинделя.

Однако, траектория перемещения оси шпинделя в координатах детали имеет форму сложной кривой 1. Это вызвано неравномерной жесткостью узлов, наличием упругих люфтов в стыках и т.д.

Таким образом, траектория движения торца шлифовального круга, полученная при его вращении, образуется при сложении двух движений:

1. Вращательного движения круга вокруг оси шпинделя.
2. Перемещения оси шпинделя по траектории 1 (смещение систем координат  $X_{шп}O_1Y_{шп}$  и  $X_oOY_o$ ).

Поэтому траектория движения торца шлифовального круга будет отличаться от идеальной. И это, в свою очередь, определит образование непрямолинейности шлифуемой поверхности.

Шлифовальному кругу сообщается круговое движение  $V_k$ , детали - продольное движение  $V_z$ . В процессе резания возникает сила резания, составляющая  $P_y$  которой вызовет деформацию замкнутой технологической системы станка на величину  $y_1$ , поэтому  $t_\phi = t - y_1$ . Кроме этого на глубину шлифования оказывает влияние и перемещение заготовки вдоль оси  $X$ . Оно имеет **скачкообразный** характер. В момент выхода шлифовального круга из контакта с заготовкой упругая система станка «выбирает» деформацию технологической системы  $y_1$ .

В момент врезания круга в заготовку глубина резания  $t_\phi' = t$ . Все это вместе взятое определяет характер образования непрямолинейности шлифуемой поверхности.

Для реализации имитационного моделирования процесса шлифования был применен программно – методический комплекс МИК (моделирования и конструирования), разработанный в ИГЭУ. При его применении модель динамического процесса представляется в виде структурной модели. Затем с помощью специального языка вводятся параметры модели и осуществляется компьютерный эксперимент.

Таланов П.В. разработал программный комплекс PROGNOZ. Он позволяет удобно вводить и хранить исходные данные, осуществляет связь с МИКом, обрабатывает результаты компьютерного эксперимента. Этот комплекс позволяет прогнозировать точность обработки деталей на плоскошлифовальных станках, а также на станках фрезерно-сверлильно-расточной группы при растачивании отверстий (некруглость, нецилиндричность).

### Лекция 13. Математическое моделирование управления производительностью, себестоимостью и точностью обработки деталей на металлорежущих станках

#### Моделирование связей производительности и точности операций металлообработки с изменением входных параметров. Идея адаптивного управления процессом обработки

Наибольшее влияние на величину поля рассеивания размеров детали оказывает изменение таких входных параметров процесса обработки как размеры припуска и твердость материала заготовки. Это обуславливает соответственное возрастание диапазона изменения усилий резания и вызываемых ими упругих отжати́й, а, следовательно, и поля рассеивания размеров обработанных заготовок (деталей). Как известно, отжатие в упругой технологической системе рассчитывается по следующей формуле:

$$y = \frac{F_y}{j},$$

где  $F_y$  - нормальная (радиальная) составляющая сила резания;

$j$  - жесткость технологической системы.

Погрешность обработки зависит от колебания отжати́я вследствие изменения усилий резания или жесткости системы. Для достижения высокой точности обработки **необходимо обеспечить постоянство величины  $y$** , что возможно

1. при постоянстве составляющей силы резания  $F_y$  (или в целом силы резания  $P$ );
2. при постоянстве отношения  $P/j$ .

Например, при обработке стали и чугуна, сила резания определяется по следующей формуле:

$$P = \sqrt{P_x^2 + P_y^2 + P_z^2} = \sqrt{(C_{px} t^{1.2} s^{0.6} HB^{1.5})^2 + (C_{py} t^{0.9} s^{0.75} HB^2)^2 + (C_{pz} t s^{0.75} HB^{0.6})^2}.$$

Иногда в формулу вводят скорость резания  $v$ .

Сила резания зависит от твердости заготовки  $HB$ , от глубины резания  $t$ , от постоянных  $C_{px}, C_{py}, C_{pz}$ , которые учитывают влияние на силу инструментального материала, геометрии инструмента, вида СОЖ и т.д. и от подачи  $s$ . Для того, чтобы управлять величиной силы, теоретически можно изменять все перечисленные выше параметры. Но практически изменять можно только подачу, т.к. изменить твердость обрабатываемого материала невозможно. Изменять глубину резания можно, но это повлечет за собой ухудшение точности обработки. Изменять по ходу обработки инструментальный материал, геометрию резца тоже практически невозможно. Поэтому единственным способом компенсации силы резания является изменение подачи, что в основном и используется в современных системах адаптивного управления в станках (иногда регулируют и скорость резания).

**Пример:** обрабатывался стальной валик диаметром  $d = 23$  мм и длиной  $l = 400$  мм. Обработка велась резцом с пластиной из твердого сплава Т15К6 и углом в плане  $\varphi = 45^\circ$ , глубина резания  $t = 1$  мм, число оборотов шпинделя  $n = 375$  об/мин. Обработка велась по двум вариантам:

а) подача оставалась постоянной  $s = 0,3$  мм/об ;

б) подача изменялась по длине обработки: при  $l = 50$  и  $100$  мм  $s = 0,53$  мм/об; при  $l = 150$  мм  $s = 0,33$  мм/об; при  $l = 200$  мм  $s = 0,3$  мм/об; при  $l = 250$  мм  $s = 0,33$  мм/об; при  $l = 300$  мм  $s = 0,47$  мм/об; при  $l = 350$  мм  $s = 0,53$  мм/об.

Были получены результаты, представленные на рис.13.1.

При изменении продольной подачи **повысилась не только точность** геометрической формы детали (колебание диаметра по длине детали уменьшилось с  $\Delta_1 = 0,34$  мм до  $\Delta_2 = 0,085$  мм), но и **увеличилась производительность обработки**, т.к. средняя подача выросла с  $0,3$  до  $0,43$  мм/об (в  $1,43$  раза). Производительность по основному времени равна

$$Q = \frac{1}{t_0} = \frac{n s_0}{L_{px}^i},$$

где  $t_0$  - основное время;

$n$  - число оборотов шпинделя;

$s_0$  - подача (на один оборот шпинделя);

$L_{рх}$  - длина рабочего хода;

$i$  - число проходов.

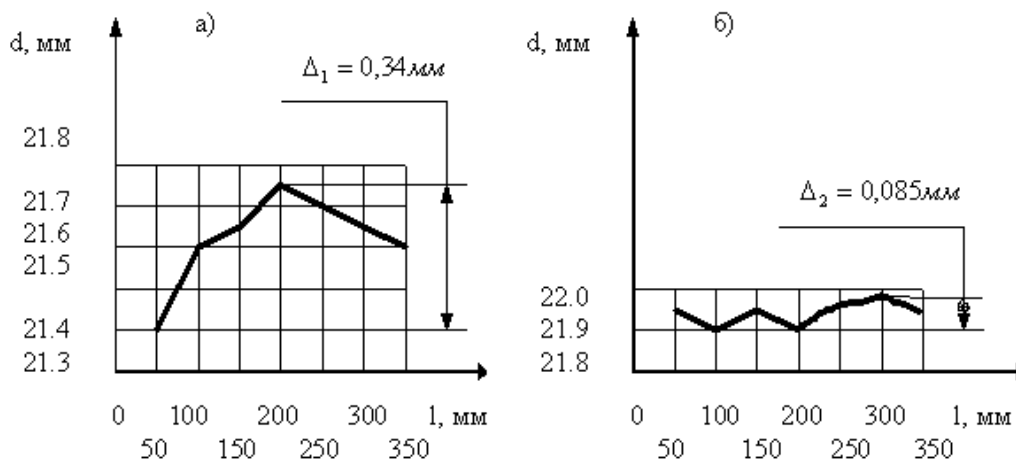


Рис.13.1. Результаты обработки валика

Как видим, производительность по основному времени прямо пропорциональна подаче, следовательно, она также увеличилась в 1,43 раза.

Для регулирования подачи применяются специальные устройства. Ранее они назывались (для универсальных станков) системы автоматического регулирования. Сейчас их называют системами адаптивного управления (САУ).

#### **Моделирование управления производительностью, себестоимостью и точностью обработки деталей на станках с ЧПУ**

Процесс достижения требуемой точности при изготовлении деталей на станках (в том числе и на станках с ЧПУ) включает три этапа:

Этап 1. Этап установки;

Этап 2. Этап статистической настройки;

Этап 3. Этап динамической настройки

**На этапе установки** происходит ориентация и закрепление заготовки на столе – спутнике или на столе станка, установка спутника на станке, автоматическая установка режущего инструмента. В результате формируется размер установки –  $A_y$ .

**При статической настройке** инструмент по программе выводится относительно технологических баз детали на размер  $A_c$ .

**При динамической настройке** в процессе резания в результате деформирования технологической системы СПИД формируется размер динамической настройки  $A_d$ , т.е. размер между режущей кромкой инструмента и базой.

Т.е. у детали получается размер  $A = A_y + A_c + A_d$ . Мы рассмотрим вопросы моделирования управления точностью, производительностью и себестоимостью обработки на третьем этапе, т.е. в процессе обработки детали, т.к. именно размером  $A_d$  управляют в САУ.

Состояние технологической системы во время выполнения процесса обработки заготовки характеризует комплекс значений:

$\delta$  - получаемая точность обработанной детали;

$N$  – потребляемая мощность;

$P$  – сила резания;

$M_{кр}$  – крутящий момент;

$A_d$  – порождаемые нагрузками упругие перемещения;

$T$  – температурный режим системы;

$\mu$  - интенсивность изнашивания режущего инструмента;

$h$  - уровень вибраций;

$S_z$  – характер стружкообразования и т.д.

Эти параметры – переменные (**внутренние параметры**) состояния технологической системы. Они образуют вектор состояния системы

$$\psi = (\delta, N, P, M_{\text{ХР}}, A_{\text{д}}, T, \mu, h, S_t, \dots).$$

Возмущающие факторы (**входные параметры**):

$t$  - колебания силы резания;

$HB$  - изменение твердости материала заготовки;

$j$  - переменная жесткость технологической системы;

$h_{\text{ж}}$  - различная режущая способность инструмента и т.д.

Они образуют вектор

$$\upsilon = (t, HB, j, h_{\text{ж}}, \dots).$$

Воздействие на технологическую систему в процессе регулирования осуществляется через **управляемые параметры**:  $s$  - подачу;  $n$  - частоту вращения шпинделя ( $v$  - скорость резания) и т.д.

Эти параметры, являясь **внутренними параметрами**, здесь выступают в роли **управляемых (управляющих)** параметров. Они образуют вектор управления

$$U = (s, n(v), \dots).$$

На переменные (параметры) состояния и управления накладываются ограничения, например,

- требуемая точность обработки детали;
- допускаемые станком нагрузки (по силе, по мощности и т.д.).

Эти ограничения либо определяют область допустимых значений переменных, либо устанавливают дополнительные зависимости между переменными состояния и управления.

На стадии программирования (разработки управляющей программы для станка с ЧПУ) невозможно заранее и окончательно определить оптимальные режимы резания вследствие неполной определенности входных (возмущающих) параметров  $t$ ,  $HB$  и т.д. Поэтому управлять подачей и скоростью резания необходимо непосредственно на станке. На станке с ЧПУ это осуществляется путем автоматической коррекции программы с целью изменения  $s$  и  $v$  на соответствующем переходе.

При создании САУ важным моментом является выбор источника информации, на основании которой осуществляется корректировка управляющей программы ( $s, v$ ). Прямое измерение размера динамической настройки  $A_{\text{д}}$ , т.е. измерение отклонений между режущими кромками инструмента и базами, определяющими положение заготовки, представляет значительные трудности. Поэтому используют **косвенные методы измерения**  $A_{\text{д}}$ . Они основаны на измерении упругих перемещений в отдельных стыках или в специально встроенных динамометрических устройствах. Их создание основано на выявлении зависимости между размером динамической настройки  $A_{\text{д}}$  и составляющими силы резания  $P = (P_x, P_y, P_z)$ :

$$A_{\text{д}} = aP_x + bP_y + cP_z,$$

где  $a, b, c$  - коэффициенты, характеризующие степень влияния каждой из составляющих силы резания на  $A_{\text{д}}$ .  $P_x, P_y, P_z$  измеряют с помощью динамометрических устройств.

Коэффициенты  $a, b$  и  $c$  определяются следующим образом. Проводят ряд экспериментов, в ходе которых изменяют в определенных пределах подачу и глубину. Для этих разных величин измеряют размер  $A_{\text{д}}$ . Для получения функциональной зависимости  $A_{\text{д}} = f(P_x, P_y, P_z)$  осуществляют аппроксимацию экспериментальных данных методом наименьших квадратов (методом Гаусса):

$$\sum_{i=1}^n [A_{\text{д}i} - (aP_{xi} + bP_{yi} + cP_{zi})]^2 \rightarrow \min.$$

При управлении процессом на этапе динамической настройки решается комплекс задач, связанных с управлением точностью обработки, с управлением режимами резания, обеспечивающими экстремум целевой функции (максимум производительности, минимум себестоимости операции). При этом используют системы **предельного регулирования и системы оптимального управления**.

#### **Адаптивные системы предельного регулирования**

Их сущность заключается в том, что с целью повышения эффективности обработки деталей на станках обеспечивается постоянство протекания технологического процесса относительно таких

регулируемых величин, как, например, сила резания, потребляемая мощность, крутящий момент, параметры точности. В результате в процессе обработки непрерывно поддерживаются наибольшие предельные значения режимов резания ( $v, s$ ), при которых исключается возможность превышения заданных ограничений.

В системах предельного регулирования для **управления процессом черновой и предварительной** обработки, когда стремятся к максимальному использованию мощности оборудования, ограничения относятся главным образом к силовым параметрам нагрузки, действующей в технологической системе. Например, при обработке с постоянной мощностью резания закон изменения продольной подачи определяется выражением

$$s \leq \sqrt[3]{\frac{1,2N_{дв}\eta}{C_x \kappa_x \gamma_x n_x}}$$

где  $N_{дв}$  - мощность двигателя главного привода станка;

$\eta$  - коэффициент полезного действия главного привода станка;

$C_x, \kappa_x, \gamma_x, n_x$  - постоянная силы резания и показатели степени в выражении для составляющей силы резания  $F_x$ .

Однако, когда при допустимой мощности резания одна из переменных состояния, например, сила  $F_y$  или крутящий момент на шпинделе  $M_{шп}$  достигают предельного значения, происходит дополнительное изменение подачи, в результате чего исключается возможность превышения заданных ограничений.

В системах предельного регулирования, обеспечивающих **управление процессом чистовой и получистовой обработки деталей**, налагаемые ограничения, образующие систему неравенств, относятся непосредственно к параметрам, определяющим достижение требуемой точности детали. Например, требование к шероховатости обработанной поверхности для токарных станков (в том числе и с ЧПУ) определяется следующим выражением (см. лабораторные работы):

$$s \leq \frac{[Ra]v^{0,4}}{65},$$

где  $[Ra]$  - допустимая шероховатость обработанной поверхности.

#### Адаптивные системы оптимального управления

С целью повышения эффективности обработки деталей в этих системах поддерживается оптимальное протекание технологического процесса относительно заданной целевой функции.

Несмотря на случайный характер потока возмущающих воздействий, адаптивная система так управляет технологическим процессом, что целевая функция непрерывно поддерживается на максимальном или минимальном уровне в пределах установленных ограничений.

При условии обеспечения требуемой точности детали эффективность технологического процесса определяет себестоимость изготавливаемой детали и производительность обработки, зависящих от основного времени обработки.

**Целевая функция времени**, устанавливающая связь между параметрами режима резания  $s, v, t$  и продолжительностью времени, затрачиваемого на операцию, имеет вид (для токарных станков с ЧПУ):

$$F_T = T_H + E v^{q-1} s^{r-1} t^n + D \pi L / 1000 v s \rightarrow \min,$$

где  $T_H$  - постоянные затраты, которые не зависят от режима обработки, мин;

$$T_H = T_{пз} / a + T_B + T_{об} + T_{отд}, \text{ мин.}$$

Здесь  $T_{пз}$  - подготовительно – заключительное время на операцию, мин;

$a$  - размер партии обрабатываемых деталей;

$T_B$  - вспомогательное время операции, мин;

$T_{об}$  - время на обслуживание без учета времени на замену инструмента, мин;

$T_{отд}$  - время на отдых рабочего, мин;

$E v^{q-1} s^{r-1} t^n$  - затраты времени, связанные с заменой затупившегося инструмента и соответствующей поднастройкой технологической системы;

$$E = T_M D \pi L / C_v^q,$$

где  $T_M$  - время на замену инструмента и соответствующую размерную настройку;

$D, L$  - диаметр и длина обрабатываемого вала;

$C_v^q$  - коэффициент для расчета скорости резания;

$v$  - скорость резания;

$s$  - подача;

$t$  - глубина резания;

$q = 1/m, r = y_v/m, n = x_v/m$ , здесь  $m, x_v, y_v$  - показатели степени в формулах для расчета режимов резания.

Анализ целевой функции времени позволяет вскрыть резервы дополнительного повышения производительности и определить оптимальные режимы резания, обеспечивающие минимальные затраты на выполнение операции.

**Целевая функция стоимости** на примере обработки вала имеет вид:

$$F_C = C_{II} + K v^{q-1} s^{r-1} t^n + C_1 D \pi L / v s \rightarrow \min,$$

где

$$C_{II} = M + C_1 (T_{из} / a + T_B + T_{об} + T_{отд}),$$

$$K = (C_1 T_M + C_{инт}) D \pi L / C_v^q;$$

$$C_1 = C_0 + C_{лр} + C_{зм}.$$

Здесь  $M$  - расходы на материал;

$C_0, C_{лр}, C_{зм}$  - расходы в единицу времени соответственно на эксплуатацию оборудования, приспособления, по зарплате с учетом накладных расходов;

$T_M$  - время на замену инструмента и соответствующую размерную настройку;

$C_{инт}$  - стоимость инструмента за период его эксплуатации.

Первый член выражения определяет постоянные затраты на материал, расходы, связанные с подготовительно – заключительным временем и временем обслуживания. Второй член выражения определяет затраты на режущий инструмент и простои при его замене. Третий член выражения определяет расходы, связанные непосредственно с выполнением процесса резания.

Далее оптимальные значения  $s, v$  находят известными методами (см., например, методические указания к практическим работам).



## Лекция 14 Объемное планирование работы технологических станочных систем

Эта и все последующие лекции посвящены вопросам математического моделирования и оптимизации технологических станочных систем.

### Объемное планирование работы механического участка при достижении максимальной загрузки технологического оборудования

**Постановка задачи.** Имеется  $m$  – станков ( $m$  – групп станков), на которых могут быть изготовлены  $n$  – типов деталей. Трудоемкость обработки  $j$  - ой детали на  $i$  – м станке составляет  $t_{ij}$ , час. Известны фонды времени работы каждого станка (группы станков) –  $B_i$ . Исходные данные для решения задачи представлены в таблице 14.1.

Таблица 14.1. Исходные данные для решения задачи, представленные в общем виде

Станки	Трудоемкость при обработке деталей типа				Фонд времени работы оборудования
	1	2	...	$n$	
1	$t_{11}$	$t_{12}$	...	$t_{1n}$	$B_1$
2	$t_{21}$	$t_{22}$	...	$t_{2n}$	$B_2$
...	...	...	...	...	...
$m$	$t_{m1}$	$t_{m2}$	...	$t_{mn}$	$B_m$

**Требуется определить** количество деталей каждого наименования ( $x_1, x_2, \dots, x_n$ ), при обработке которых достигается максимальная загрузка оборудования участка.

**Математическая модель** для решения задачи запишется:

**Ограничения:**

$$t_{11}x_1 + t_{12}x_2 + \dots + t_{1n}x_n \leq B_1; \quad (14.1)$$

$$t_{21}x_1 + t_{22}x_2 + \dots + t_{2n}x_n \leq B_2; \quad (14.2)$$

$$\dots$$

$$t_{m1}x_1 + t_{m2}x_2 + \dots + t_{mn}x_n \leq B_m; \quad (14.3)$$

$$\forall x_j \geq 0 (j = \overline{1, n}). \quad (14.4)$$

**Управляемые параметры:**

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

**Целевая функция:**

$$y = \sum_{i=1}^m \sum_{j=1}^n t_{ij}x_j \rightarrow \max. \quad (14.5)$$

Задача решается методом линейного программирования. При этом следует иметь в виду следующее. Количество ограничений вида (14.1) - (14.3) в математической модели должно строго равняться количеству станков (групп станков) участка. При решении задачи с помощью компьютера количество станков (групп станков), а также типов деталей практически не ограничено и определяется только возможностями компьютера и соответствующей программы.

При решении задачи вручную с применением графо-аналитического метода количество типов станков (групп станков) также не ограничено, но их увеличение естественным образом приведет к увеличению времени расчетов. Количество же типов деталей не должно превышать двух, т.к. в противном случае невозможно будет на плоскости выполнить необходимые графические построения.

**Пример.** Исходные данные для примера приведены в таблице 14.2.

Таблица 14.2. Исходные данные для решения задачи

Станки	Трудоемкость обработки, час., деталей типа		Фонд времени работы оборудования, час.
	D <sub>1</sub>	D <sub>2</sub>	
Токарные	0,1	0,3	200
Фрезерные	0,5	0,9	700
Сверлильные	0,3	0,2	330

Обозначим через  $x_1$  количество деталей типа D<sub>1</sub>, через  $x_2$  количество деталей типа D<sub>2</sub>.

Математическая модель для решения данной задачи запишется :

Ограничения (по фонду времени работы оборудования):

$$0,1x_1 + 0,3x_2 \leq 200; \quad (14.6)$$

$$0,5x_1 + 0,9x_2 \leq 700; \quad (14.7)$$

$$0,3x_1 + 0,2x_2 \leq 330; \quad (14.8)$$

$$x_1 \geq 0; \quad (14.9)$$

$$x_2 \geq 0. \quad (14.10)$$

Целевая функция (суммарное время работы всех групп оборудования):

$$F(x_1, x_2) = 0,1x_1 + 0,3x_2 + 0,5x_1 + 0,9x_2 + 0,3x_1 + 0,2x_2 = 0,9x_1 + 1,4x_2 \rightarrow \max \quad (14.11)$$

Требуется найти значения  $x_1$  и  $x_2$ , удовлетворяющие заданным ограничениям (14.6) – (14.10) и обеспечивающие максимум целевой функции (14.11). Параметры  $x_1$  и  $x_2$  являются **управляемыми параметрами** в математической модели.

Решим задачу графо – аналитическим методом (см. лекцию 6). Графическая иллюстрация решения задачи приведена на рис. 14.1.

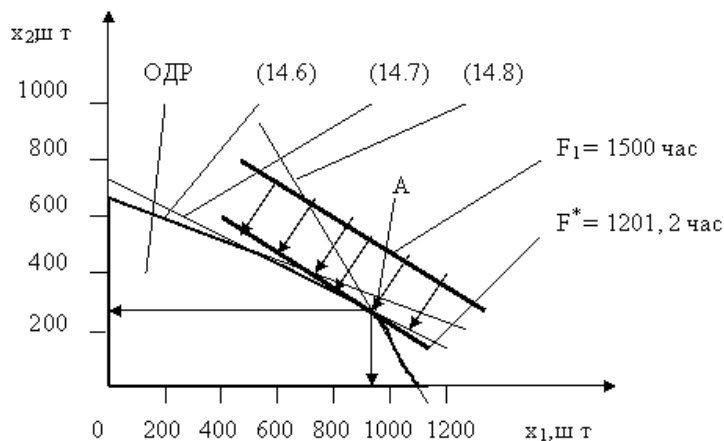


Рис.14.1. Графическая иллюстрация решения задачи

Вычисления для построения ограничений (14.6) – (14.8):

$$0,1x_1 + 0,3x_2 \leq 200 \rightarrow 0,1x_1 + 0,3x_2 = 200 \Rightarrow x_2 = (200 - 0,1x_1) / 0,3.$$

$x_1$	0	1000
$x_2$	667	333

$$0,5x_1 + 0,9x_2 \leq 700 \rightarrow 0,5x_1 + 0,9x_2 = 700 \Rightarrow x_2 = (700 - 0,5x_1) / 0,9.$$

$x_1$	0	1000
$x_2$	778	222

$$0,3x_1 + 0,2x_2 \leq 330 \rightarrow 0,3x_1 + 0,2x_2 = 330 \Rightarrow x_2 = (330 - 0,3x_1) / 0,2.$$

$x_1$	500	1000
$x_2$	900	150

Направления допустимости ограничений (14.6) – (14.8) – «вниз – влево».

Ограничения (14.9) и (14.10) – это оси координат. Направления их допустимости – «вправо» и «вверх».

Для нахождения точки касания границы ОДР прямой линией, определяющей целевую функцию, построим сначала произвольную прямую для целевой функции, приравняв ее выражение к произвольному числу в пределах масштаба построений, например, к 1500:

$$F_1 = 0,9x_1 + 1,4x_2 = 1500 \Rightarrow x_2 = (1500 - 0,9x_1) / 1,4.$$

$x_1$	500	1000
$x_2$	750	428

Проведя прямую линию, параллельную данной, находим точку касания ее границы ОДР – это точка А. Для нахождения ее координат (точки пересечения ограничений 14.7 и 14.8) решаем следующую систему уравнений:

$$\begin{cases} 0,5x_1 + 0,9x_2 = 700 \\ 0,3x_1 + 0,2x_2 = 330 \end{cases}$$

$$x_2 = (330 - 0,3x_1) / 0,2 \Rightarrow 0,5x_1 + 0,9((330 - 0,3x_1) / 0,2) = 700 \Rightarrow 0,5x_1 + 1485 - 1,35x_1 = 700 \Rightarrow 0,85x_1 = 785 \Rightarrow x_1 \cong 924 \text{шт.}$$

$$x_2 = (330 - 0,3 \cdot 924) / 0,2 = 264 \text{шт.}$$

Т.е. окончательно  $x_{1\text{опт}} = 924 \text{шт.}$ ,  $x_{2\text{опт}} = 264 \text{шт.}$

Максимальное значение целевой функции (максимальная загрузка оборудования участка) при оптимальных значениях искомых параметров составит:

$$F^* = 0,9 \cdot 924 + 1,4 \cdot 264 = 1201,2 \text{час.}$$

### Задача о минимальной загрузке оборудования

Эта и последующие задачи в данной лекции приводятся на уровне постановки задачи и формирования математической модели для ее решения. Все они решаются методами линейного программирования.

Имеется  $m$  станков, на которых могут быть изготовлены  $n$  типов деталей. Производительность  $i$ -го станка при изготовлении детали  $j$ -го типа составляет  $C_{ij}$ . Величины плановых заданий  $A_j$  на изготовление  $j$ -ой детали и ресурс времени  $B_i$  работы  $i$ -го станка приведены в таблице 14.3.

Таблица 14.3 Исходные данные для решения задачи

Станки	Тип деталей				Ресурс времени
	1	2	...	$n$	
1	$c_{11}$	$c_{12}$	...	$c_{1n}$	$B_1$
2	$c_{21}$	$c_{22}$	...	$c_{2n}$	$B_2$
...	...	...	...	...	...
$m$	$c_{m1}$	$c_{m2}$	...	$c_{mn}$	$B_m$
Требуемое Количество Деталей	$A_1$	$A_2$	...	$A_n$	-

Требуется, учитывая ресурсы времени работы каждого станка распределить задания между станками таким образом, чтобы общее время работы всех станков было минимальным.

Пусть  $t_{ij}$  - время изготовления  $j$ -ой детали  $i$ -м станком. Составим ограничения по ресурсу времени для каждого станка:

$$\sum_{j=1}^n t_{ij} \leq B_i, i = \overline{1, m}. \quad (14.12)$$

Условия выполнения плановых заданий имеют вид:

$$\sum_{i=1}^m c_{ij} t_{ij} = A_j, j = \overline{1, n}. \quad (14.13)$$

Решение поставленной задачи состоит в минимизации линейной целевой функции (суммарного времени)

$$\sum_{i=1}^m \sum_{j=1}^n t_{ij} \rightarrow \min \quad (14.14)$$

при ограничениях (14.12), (14.13) и условии, что все переменные  $t_{ij} \geq 0$ .

### Задача об оптимальном распределении деталей по станкам

Пусть некоторая машина состоит из  $k$  различных видов деталей, которые мы пронумеруем числами  $1, 2, \dots, k$ . Имеется  $m$  типов различных станков, причем количество станков  $i$ -го типа равно  $a_i, i = 1, 2, \dots, m$ . Детали могут быть изготовлены на станках разного типа. Производительность станка  $i$ -го типа при изготовлении  $j$ -ой детали составляет  $c_{ij}$ . После изготовления детали поступают на сборку. Требуется закрепить станки за деталями так, чтобы в единицу времени получать максимальное количество машин.

Пусть  $x_{ij}$  - количество станков  $i$ -го типа, на которых можно изготовить  $j$ -ю деталь. Очевидно, что количество станков  $i$ -го типа, изготавливающих детали  $1, 2, \dots, k$  видов, не должно превышать заданное число  $a_i$ :

$$\sum_{j=1}^k x_{ij} \leq a_i, i = \overline{1, m}. \quad (14.15)$$

Общее количество деталей  $j$ -го вида, изготовленное на станках за единицу времени, составляет  $\sum_{i=1}^m c_{ij} x_{ij}, j = \overline{1, k}$ . В каждой машине имеется ровно одна деталь с номером  $j, j = 1, 2, \dots, k$ . Поэтому, для того чтобы не было изготовлено лишних и не было дефицитных деталей, должны выполняться условия комплектности:

$$\sum_{i=1}^m c_{ij} x_{ij} = \sum_{i=1}^m c_{il} x_{il}, j \neq l; j, l = \overline{1, k}. \quad (14.16)$$

Общее количество комплектов деталей, необходимых для сборки машины, равно общему количеству какой-либо одной детали, имеющей, например, номер 1. Поэтому решение задачи заключается в максимизации линейной функции

$$\sum_{i=1}^m c_{i1} x_{i1} \rightarrow \max \quad (14.17)$$

при ограничениях (14.15), (14.16) с дополнительным условием, что все переменные  $x_{ij} \geq 0$ .

Найденные оптимальные значения  $x_{ij}^{(0)}$  этой задачи не обязательно целые числа. Например,  $x_{11}^{(0)} = 2,5$  означает, что на двух станках первого типа в течение единицы времени будут изготавливать деталь с номером 1, тогда как третий станок того же типа будет работать лишь половину указанного времени.

### Задача о производстве продукции при ограниченных запасах сырья

Из  $n$  видов сырья производится  $m$  различных типов продукции. Стоимость реализации изготовленной продукции  $i$ -го типа составляет  $c_i, i = \overline{1, m}$ . Запас сырья  $j$ -го вида на планируемый период равен  $z_j, j = \overline{1, n}$ . Потребность в сырье  $i$ -го типа составляет  $p_{ij}$ . Исходные данные для решения задачи приведены в таблице 14.4.

Таблица 14.4 Исходные данные для решения задачи

Тип продукции	Вид сырья				Стоимость продукции
	1	2	...	$n$	
1	$p_{11}$	$p_{12}$	...	$p_{1n}$	$c_1$
2	$p_{21}$	$p_{22}$	...	$p_{2n}$	$c_2$
...	...	...	...	...	...
$m$	$p_{m1}$	$p_{m2}$	...	$p_{mn}$	$c_m$
Запас сырья	$z_1$	$z_2$	...	$z_n$	-

Требуется для каждого типа продукта  $i = 1, 2, \dots, m$  определить такой объем производства  $x_i$ , чтобы обеспечить максимальную стоимость реализации изготовленной продукции при условии, что не будут превышены запасы имеющегося сырья.

Ограничения по запасам сырья имеют вид:

$$\sum_{i=1}^m p_{ij} x_i \leq z_j, j = \overline{1, n}. \quad (14.18)$$

Задача заключается в том, чтобы определить оптимальные значения параметров (переменных)  $x_1, x_2, \dots, x_m$ , обращающих в максимум стоимость продукции, т.е. целевую функцию

$$y = \sum_{i=1}^m c_i x_i \rightarrow \max$$

при ограничениях (14.18) и дополнительных условиях

## Лекция 15. Основы теории массового обслуживания

Теория массового обслуживания составляет один из разделов теории вероятностей. В этой теории рассматриваются **вероятностные** задачи и математические модели (до этого нами рассматривались детерминированные математические модели). Напомним, что:

**Детерминированная математическая модель** отражает поведение объекта (системы, процесса) с позиций **полной определенности** в настоящем и будущем.

**Вероятностная математическая модель** учитывает влияние случайных факторов на поведение объекта (системы, процесса) и, следовательно, оценивает будущее с позиций вероятности тех или иных событий.

Т.е. здесь как, например, в теории игр задачи рассматриваются **в условиях неопределенности**.

Рассмотрим сначала некоторые понятия, которые характеризуют «стохастическую неопределенность», когда неопределенные факторы, входящие в задачу, представляют собой случайные величины (или случайные функции), вероятностные характеристики которых либо известны, либо могут быть получены из опыта. Такую неопределенность называют еще «благоприятной», «доброкачественной».

### Понятие случайного процесса

Строго говоря, случайные возмущения присущи любому процессу. Проще привести примеры случайного, чем «неслучайного» процесса. Даже, например, процесс хода часов (вроде бы это строгая выверенная работа – «работает как часы») подвержен случайным изменениям (уход вперед, отставание, остановка). Но до тех пор, пока эти возмущения несущественны, мало влияют на интересующие нас параметры, мы можем ими пренебречь и рассматривать процесс как детерминированный, неслучайный.

Пусть имеется некоторая система  $S$  (техническое устройство, группа таких устройств, технологическая система – станок, участок, цех, предприятие, отрасль промышленности и т.д.). В системе  $S$  протекает **случайный процесс**, если она с течением времени меняет свое состояние (переходит из одного состояния в другое), причем, заранее неизвестным случайным образом.

**Примеры:** 1. Система  $S$  – технологическая система (участок станков). Станки время от времени выходят из строя и ремонтируются. Процесс, протекающий в этой системе, случаен.

2. Система  $S$  – самолет, совершающий рейс на заданной высоте по определенному маршруту. Возмущающие факторы – метеоусловия, ошибки экипажа и т.д., последствия – «болтанка», нарушение графика полетов и т.д.

### Марковский случайный процесс

Случайный процесс, протекающий в системе, называется **Марковским**, если для любого момента времени  $t_0$  вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент  $t_0$  и не зависят от того, когда и как система пришла в это состояние.

Пусть в настоящий момент  $t_0$  система находится в определенном состоянии  $S_0$ . Мы знаем характеристики состояния системы в настоящем  $t_0 \rightarrow S_0$  и все, что было при  $t < t_0$  (предысторию процесса). Можем ли мы предугадать (предсказать) будущее, т.е. что будет при  $t > t_0$ ? В точности – нет, но какие-то вероятностные характеристики процесса в будущем найти можно. Например, вероятность того, что через некоторое время  $\tau$  система  $S$  окажется в состоянии  $S_1$  или останется в состоянии  $S_0$  и т.д.

**Пример.** Система  $S$  – группа самолетов, участвующих в воздушном бою. Пусть  $x$  – количество «красных» самолетов,  $y$  – количество «синих» самолетов. К моменту времени  $t_0$  количество сохранившихся (не сбитых) самолетов соответственно –  $x_0, y_0$ . Нас интересует вероятность того, что в момент времени  $(t_0 + \tau)$  численный перевес будет на стороне «красных». Эта вероятность зависит от того, в каком состоянии находилась система в момент времени  $t_0$ , а не от того, когда и в какой последовательности погибали сбитые до момента  $t_0$  самолеты.

На практике Марковские процессы в чистом виде обычно не встречаются. Но имеются процессы, для которых влиянием «предистории» можно пренебречь. **И при изучении таких процессов можно применять Марковские модели** (в теории массового обслуживания рассматриваются и не Марковские системы массового обслуживания, но математический аппарат, их описывающий, гораздо сложнее).

В исследовании операций большое значение имеют Марковские случайные процессы с дискретными состояниями и непрерывным временем.

Процесс называется **процессом с дискретным состоянием**, если его возможные состояния  $S_1, S_2, \dots$  можно заранее определить, и переход системы из состояния в состояние происходит «скачком», практически мгновенно.

Процесс называется **процессом с непрерывным временем**, если моменты возможных переходов из состояния в состояние не фиксированы заранее, а неопределенны, случайны и могут произойти в любой момент.

Далее рассматриваются только процессы с дискретным состоянием и непрерывным временем.

**Пример.** Технологическая система (участок)  $S$  состоит из двух станков, каждый из которых в случайный момент времени может выйти из строя (отказаться), после чего мгновенно начинается ремонт узла, тоже продолжающийся заранее неизвестное, случайное время. Возможны следующие состояния системы:

$S_0$  - оба станка исправны;

$S_1$  - первый станок ремонтируется, второй исправен;

$S_2$  - второй станок ремонтируется, первый исправен;

$S_3$  - оба станка ремонтируются.

Переходы системы  $S$  из состояния в состояние происходят практически мгновенно, в случайные моменты выхода из строя того или иного станка или окончания ремонта.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой – **графом состояний**. Вершины графа – состояния системы. Дуги графа – возможные переходы из состояния в состояние. Для нашего примера граф состояний приведен на рис.15.1.

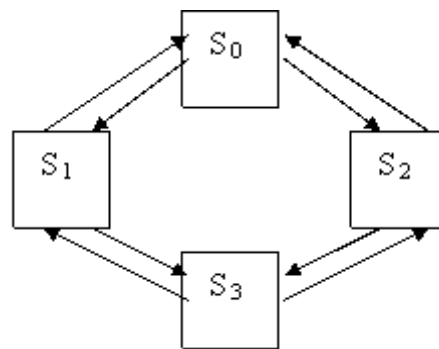


Рис.15.1. Граф состояний системы

Примечание. Переход из состояния  $S_0$  в  $S_3$  на рисунке не обозначен, т.к. предполагается, что станки выходят из строя независимо друг от друга. Вероятностью одновременного выхода из строя обоих станков мы пренебрегаем.

### Потоки событий

**Поток событий** – последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени.

В предыдущем примере – это поток отказов и поток восстановлений. Другие примеры: поток вызовов на телефонной станции, поток покупателей в магазине и т.д.

Поток событий можно наглядно изобразить рядом точек на оси времени  $O t$  – рис. 15.2.

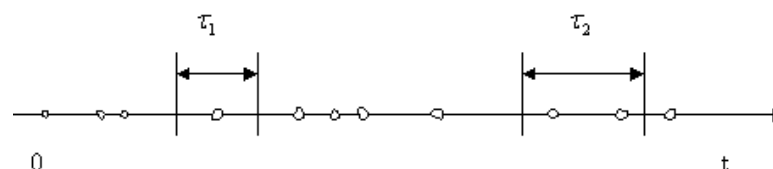


Рис.15.2. Изображение потока событий на оси времени

Положение каждой точки случайно, и здесь изображена лишь какая-то одна реализация потока.

**Интенсивность потока событий** ( $\lambda$ ) – это среднее число событий, приходящееся на единицу времени.

Рассмотрим некоторые свойства (виды) потоков событий.

Поток событий называется **стационарным**, если его вероятностные характеристики не зависят от времени. В частности, интенсивность  $\lambda$  стационарного потока постоянна. Поток событий неизбежно имеет сгущения или разрежения, но они не носят закономерного характера, и среднее число событий, приходящееся на единицу времени, постоянно и от времени не зависит.

Поток событий называется **поток без последствий**, если для любых двух непересекающихся участков времени  $\tau_1$  и  $\tau_2$  (см. рис.15.2) число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой. Другими словами, это означает, что события, образующие поток, появляются в те или иные моменты времени **независимо друг от друга** и вызваны каждое своими собственными причинами.

Поток событий называется **ординарным**, если события в нем появляются поодиночке, а не группами по несколько сразу.

Поток событий называется **простейшим (или стационарным пуассоновским)**, если он обладает сразу тремя свойствами: 1) стационарен, 2) ординарен, 3) не имеет последствий.

Простейший поток имеет наиболее простое математическое описание. Он играет среди потоков такую же особую роль, как и закон нормального распределения среди других законов распределения. А именно, при наложении достаточно большого числа независимых, стационарных и ординарных потоков (сравнимых между собой по интенсивности) получается поток, близкий к простейшему.

Для простейшего потока с интенсивностью  $\lambda$  интервал  $T$  между соседними событиями имеет так называемое **показательное (экспоненциальное) распределение** с плотностью

$$f(t) = \lambda e^{-\lambda t},$$

где  $\lambda$  - параметр показательного закона.

Для случайной величины  $T$ , имеющей показательное распределение, математическое ожидание  $m_T$  есть величина, обратная параметру, а среднее квадратичное отклонение  $\sigma_T$  равно математическому ожиданию

$$m_T = \sigma_T = 1/\lambda$$

### Уравнения Колмогорова для вероятностей состояний. Финальные вероятности состояний

Рассматривая Марковские процессы с дискретными состояниями и непрерывным временем, подразумевается, что все переходы системы  $S$  из состояния в состояние происходят под действием простейших потоков событий (потоков вызовов, потоков отказов, потоков восстановлений и т.д.). Если все потоки событий, переводящие систему  $S$  из состояния в состояние простейшие, то процесс, протекающий в системе, будет Марковским.

Итак, на систему, находящуюся в состоянии  $S_i$ , действует простейший поток событий. Как только появится первое событие этого потока, происходит «перескок» системы из состояния  $S_i$  в состояние  $S_j$  (на графе состояний по стрелке  $S_i \rightarrow S_j$ ).

Для наглядности на графе состояний системы у каждой дуги проставляют интенсивности того потока событий, который переводит систему по данной дуге (стрелке).  $\lambda_{ij}$  - интенсивность потока событий, переводящий систему из состояния  $S_i$  в  $S_j$ . Такой граф называется **размеченным**. Для нашего примера размеченный граф приведен на рис.15.3.

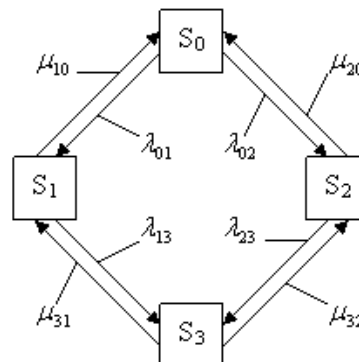


Рис.15.3. Размеченный граф состояний системы



На этом рисунке  $\lambda_{ij}$  - интенсивности потока отказов;  $\mu_{ij}$  - интенсивности потока восстановлений.

Предполагаем, что среднее время ремонта станка не зависит от того, ремонтируется ли один станок или оба сразу. Т.е. ремонтом каждого станка занят отдельный специалист.

Пусть система находится в состоянии  $S_0$ . В состояние  $S_1$  ее переводит поток отказов первого станка. Его интенсивность равна

$$\lambda_{01} = 1 / T_{\text{ср. раб. 1, ед. времени}}^{-1},$$

где  $T_{\text{ср. раб. 1}}$  - среднее время безотказной работы первого станка.

Из состояния  $S_1$  в  $S_0$  систему переводит поток «окончаний ремонтов» первого станка. Его интенсивность равна

$$\mu_{10} = 1 / T_{\text{ср. рем. 1, ед. времени}}^{-1},$$

где  $T_{\text{ср. рем. 1}}$  - среднее время ремонта первого станка.

Аналогично вычисляются интенсивности потоков событий, переводящих систему по всем дугам графа. Имея в своем распоряжении размеченный граф состояний системы, строится **математическая модель** данного процесса.

Пусть рассматриваемая система  $S$  имеет  $n$  - возможных состояний  $S_1, S_2, \dots, S_n$ . Вероятность  $i$ -го состояния  $p_i(t)$  - это вероятность того, что в момент времени  $t$  система будет находиться в состоянии  $S_i$ . Очевидно, что для любого момента времени сумма всех вероятностей состояний равна единице:

$$\sum_{i=1}^n p_i(t) = 1.$$

Для нахождения всех вероятностей состояний  $p_i(t)$  как функций времени составляются и решаются **уравнения Колмогорова** – особого вида уравнения, в которых неизвестными функциями являются вероятности состояний. Правило составления этих уравнений приведем здесь без доказательств. Но прежде, чем его приводить, объясним понятие **финальной вероятности состояния**. Что будет происходить с вероятностями состояний при  $t \rightarrow \infty$ ? Будут ли  $p_1(t), p_2(t), \dots$  стремиться к каким-либо пределам? Если эти пределы существуют и не зависят от начального состояния системы, то они называются **финальными вероятностями состояний**.

$$\lim_{t \rightarrow \infty} p_i(t) = p_i, i = \overline{1, n},$$

где  $n$  - конечное число состояний системы.

**Финальные вероятности состояний** – это уже не переменные величины (функции времени), а постоянные числа. Очевидно, что

$$\sum_{i=1}^n p_i = 1.$$

**Финальная вероятность состояния**  $S_i$  – это по – существу среднее относительное время пребывания системы в этом состоянии.

Например, система  $S$  имеет три состояния  $S_1, S_2$  и  $S_3$ . Их финальные вероятности равны соответственно 0,2; 0,3 и 0,5. Это значит, что система в предельном стационарном состоянии в среднем 2/10 времени проводит в состоянии  $S_1$ , 3/10 – в состоянии  $S_2$  и 5/10 – в состоянии  $S_3$ .

**Правило составления системы уравнений Колмогорова:** в каждом уравнении системы в **левой его части** стоит финальная вероятность данного состояния  $p_i$ , умноженная на суммарную интенсивность всех потоков, **ведущих из данного состояния**, а в **правой его части** – сумма произведений интенсивностей всех потоков, **входящих в  $i$ -е состояние**, на вероятности тех состояний, из которых эти потоки исходят.

Пользуясь этим правилом, напомним систему уравнений для нашего примера:

$$\begin{cases} (\lambda_{01} + \lambda_{02})p_0 = \mu_{10}p_1 + \mu_{20}p_2 \\ (\mu_{10} + \lambda_{13})p_1 = \lambda_{01}p_0 + \mu_{31}p_3 \\ (\mu_{20} + \lambda_{23})p_2 = \lambda_{02}p_0 + \mu_{32}p_3 \\ (\mu_{31} + \mu_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2 \end{cases}$$

Эту систему четырех уравнений с четырьмя неизвестными  $p_0, p_1, p_2, p_3$ , казалось бы, можно вполне решить. Но эти уравнения однородны (не имеют свободного члена), и, значит, определяют неизвестные только с точностью до произвольного множителя. Однако можно воспользоваться нормировочным условием

$$p_0 + p_1 + p_2 + p_3 = 1$$

и с его помощью решить систему. При этом одно (любое) из уравнений можно отбросить (оно вытекает как следствие из остальных).

**Продолжение примера.** Пусть значения интенсивностей потоков равны:

$$\lambda_{01} = \lambda_{23} = 1; \lambda_{13} = \lambda_{02} = 2; \mu_{10} = \mu_{32} = 2; \mu_{20} = \mu_{31} = 3.$$

Четвертое уравнение отбрасываем, добавляя вместо него нормировочное условие:

$$\begin{cases} 3p_0 = 2p_1 + 3p_2 \\ 4p_1 = p_0 + 3p_3 \\ 4p_2 = 2p_0 + 2p_3 \\ p_0 + p_1 + p_2 + p_3 = 1 \end{cases}$$

$$p_0 = 6/15 = 0,4; p_1 = 3/15 = 0,2; p_2 = 4/15 \cong 0,27; p_3 = 2/15 \cong 0,13.$$

Т.е. в предельном, стационарном режиме система  $S$  в среднем 40 % времени будет проводить в состоянии  $S_0$  (оба станка исправны), 20 % - в состоянии  $S_1$  (первый станок ремонтируется, второй работает), 27 % - в состоянии  $S_2$  (второй станок ремонтируется, первый работает), 13% - в состоянии  $S_3$  (оба станка ремонтируются). Знание этих финальных вероятностей может помочь оценить среднюю эффективность работы системы и загрузку ремонтных органов.

Пусть система  $S$  в состоянии  $S_0$  (полностью исправна) приносит в единицу времени доход 8 условных единиц, в состоянии  $S_1$  – доход 3 условные единицы, в состоянии  $S_2$  – доход 5 условных единиц, в состоянии  $S_3$  – не приносит дохода. Тогда в предельном, стационарном режиме средний доход в единицу времени будет равен  $W = 0,4 * 8 + 0,2 * 3 + 0,27 * 5 = 5,15$  условных единиц.

Станок 1 ремонтируется долю времени, равную  $p_1 + p_3 = 0,2 + 0,13 = 0,33$ . Станок 2 ремонтируется долю времени, равную  $p_2 + p_3 = 0,27 + 0,13 = 0,4$ . Возникает **задача оптимизации**. Пусть мы можем уменьшить среднее время ремонта первого или второго станка (или обоих), но это нам обойдется в определенную сумму. Спрашивается, окупит ли увеличение дохода, связанное с ускорением ремонта, повышенные расходы на ремонт? Нужно будет решить систему четырех уравнений с четырьмя неизвестными.

### Задачи теории массового обслуживания

Примеры систем массового обслуживания (СМО): телефонные станции, ремонтные мастерские, билетные кассы, справочные бюро, станочные и другие технологические системы, системы управления гибких производственных систем и т.д.

Каждая СМО состоит из какого – то количества обслуживающих единиц, которые называются **каналами обслуживания** (это станки, транспортные тележки, роботы, линии связи, кассиры, продавцы и т.д.). Всякая СМО предназначена для обслуживания какого – то **потока заявок** (требований), поступающих в какие – то случайные моменты времени.

Обслуживание заявки продолжается какое – то, вообще говоря, случайное время, после чего канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времени обслуживания приводит к тому, что в какие – то периоды времени на входе СМО скапливается излишне большое количество заявок (они либо становятся в очередь, либо покидают СМО необслуженными). В другие же периоды СМО будет работать с недогрузкой или вообще простаивать.

Процесс работы СМО – случайный процесс с дискретными состояниями и непрерывным временем. Состояние СМО меняется скачком в моменты появления каких-то событий (прихода новой заявки, окончания обслуживания, 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. Эти показатели описывают способность СМО справляться с потоком заявок. Ими могут быть: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналов; среднее число заявок в очереди; среднее время ожидания обслуживания и т.д.

Математический анализ работы СМО очень облегчается, если процесс этой работы Марковский, т.е. потоки событий, переводящие систему из состояния в состояние – простейшие. Иначе математическое описание процесса очень усложняется и его редко удается довести до конкретных аналитических зависимостей. На практике не Марковские процессы с приближением приводятся к Марковским. Приведенный далее математический аппарат описывает Марковские процессы.

### **Классификация систем массового обслуживания**

Первое деление (по наличию очередей):

1. СМО с отказами;
2. СМО с очередью.

**В СМО с отказами** заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем не обслуживается.

**В СМО с очередью** заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь и ожидает возможности быть обслуженной.

**СМО с очередями подразделяются** на разные виды в зависимости от того, как организована очередь – **ограничена или не ограничена**. Ограничения могут касаться как длины очереди, так и времени ожидания, «дисциплины обслуживания».

Итак, например, рассматриваются следующие СМО:

- СМО с нетерпеливыми заявками (длина очереди и время обслуживания ограничено);
- СМО с обслуживанием с приоритетом, т.е. некоторые заявки обслуживаются вне очереди и т.д.

Кроме этого СМО делятся на открытые СМО и замкнутые СМО.

**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 исправно и ждет наладки.

Классификация СМО далеко не ограничивается приведенными разновидностями, но этого достаточно.

## Лекция 16. Математические модели простейших систем массового обслуживания

Ниже будут рассмотрены примеры простейших систем массового обслуживания (СМО). Понятие «простейшие» не означает «элементарные». Математические модели этих систем применимы и успешно используются в практических расчетах.

### Одноканальная СМО с отказами

**Дано:** система имеет один канал обслуживания, на который поступает простейший поток заявок с интенсивностью  $\lambda$ . Поток обслуживаний имеет интенсивность  $\mu$ . Заявка, заставшая систему занятой, сразу же покидает ее.

**Найти:** абсолютную и относительную пропускную способность СМО и вероятность того, что заявка, пришедшая в момент времени  $t$ , получит отказ.

Система при любом  $t > 0$  может находиться в двух состояниях:  $S_0$  – канал свободен;  $S_1$  – канал занят. Переход из  $S_0$  в  $S_1$  связан с появлением заявки и немедленным началом ее обслуживания. Переход из  $S_1$  в  $S_0$  осуществляется, как только очередное обслуживание завершится (рис.16.1).

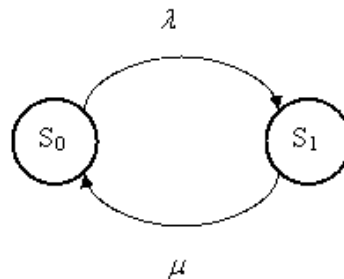


Рис.16.1. Граф состояний одноканальной СМО с отказами

Выходные характеристики (характеристики эффективности) этой и других СМО будут даваться без выводов и доказательств.

**Абсолютная пропускная способность** (среднее число заявок, обслуживаемых в единицу времени):

$$A = \frac{\lambda\mu}{\lambda + \mu}, \text{ шт/ед.времени,}$$

где  $\lambda$  – интенсивность потока заявок (величина, обратная среднему промежутку времени между поступающими заявками -  $\overline{t_3}$ );  $\mu$  – интенсивность потока обслуживаний (величина, обратная среднему времени обслуживания  $\overline{t_{об}}$ )

**Относительная пропускная способность** (средняя доля заявок, обслуживаемых системой):

$$Q = \frac{\mu}{\lambda + \mu}.$$

**Вероятность отказа** (вероятность того, что заявка покинет СМО необслуженной):

$$P_{отк} = \frac{\lambda}{\lambda + \mu}.$$

Очевидны следующие соотношения:  $Q = 1 - P_{отк}$  и  $P_{отк} = 1 - Q$ .

**Пример.** Технологическая система состоит из одного станка. На станок поступают заявки на изготовление деталей в среднем через 0,5 часа ( $\overline{t_3} = 0,5\text{ч.}$ ). Среднее время изготовления одной детали равно  $\overline{t_{об}} = 0,6\text{ч.}$  Если при поступлении заявки на изготовление детали станок занят, то она (деталь) направляется на другой станок. Найти абсолютную и относительную пропускную способности системы и вероятность отказа по изготовлению детали.

Решение.

$$\lambda = 1/\overline{t_3} = 1/0,5 = 2\text{ч}^{-1}, \mu = 1/\overline{t_{об}} = 1/0,6 \cong 1,67\text{ч}^{-1},$$

$$A = \frac{\lambda\mu}{\lambda + \mu} = \frac{2 \cdot 1,67}{2 + 1,67} = 0,91\text{дет/ч}; Q = \frac{\mu}{\lambda + \mu} = \frac{1,67}{2 + 1,67} = 0,455 \cong 0,46.$$

Т.е. в среднем примерно 46 % деталей обрабатываются на этом станке.

$$P_{отк} = \frac{\lambda}{\lambda + \mu} = \frac{2}{2 + 1,67} \cong 0,54$$

Т.е. в среднем примерно 54 % деталей направляются на обработку на другие станки.

### N – канальная СМО с отказами (задача Эрланга)

Это одна из первых задач теории массового обслуживания. Она возникла из практических нужд телефонии и была решена в начале 20 века датским математиком Эрлангом.

**Дано:** в системе имеется  $n$  – каналов, на которые поступает поток заявок с интенсивностью  $\lambda$ . Поток обслуживаний имеет интенсивность  $\mu$ . Заявка, заставшая систему занятой, сразу же покидает ее.

**Найти:** абсолютную и относительную пропускную способность СМО; вероятность того, что заявка, пришедшая в момент времени  $t$ , получит отказ; среднее число заявок, обслуживаемых одновременно (или, другими словами, среднее число занятых каналов).

**Решение.** Состояние системы  $S$  (СМО) нумеруется по максимальному числу заявок, находящихся в системе (оно совпадает с числом занятых каналов):

- $S_0$  – в СМО нет ни одной заявки;
- $S_1$  – в СМО находится одна заявка (один канал занят, остальные свободны);
- $S_2$  – в СМО находится две заявки (два канала заняты, остальные свободны);
- ...
- $S_n$  – в СМО находится  $n$  – заявок (все  $n$  – каналов заняты).

Граф состояний СМО представлен на рис. 16.2.

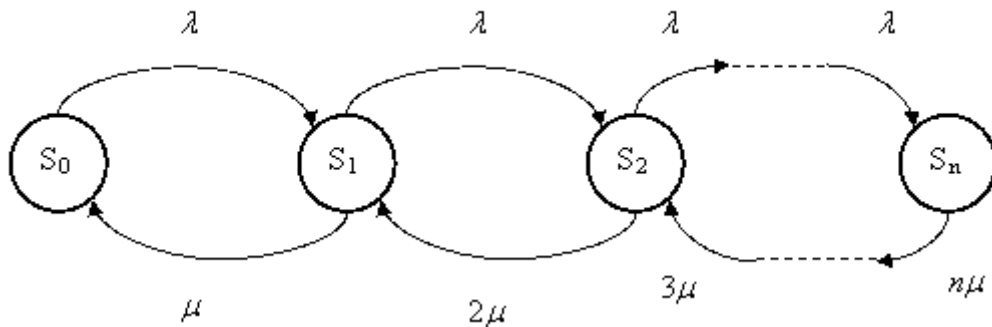


Рис.16.2. Граф состояний для  $n$  – канальной СМО с отказами

Почему граф состояний размечен именно так? Из состояния  $S_0$  в состояние  $S_1$  систему переводит поток заявок с интенсивностью  $\lambda$  (как только приходит заявка, система переходит из  $S_0$  в  $S_1$ ). Если система находилась в состоянии  $S_1$  и пришла еще одна заявка, то она переходит в состояние  $S_2$  и т.д.

Почему такие интенсивности у нижних стрелок (дуг графа)? Пусть система находится в состоянии  $S_1$  (работает один канал). Он производит  $\mu$  обслуживаний в единицу времени. Поэтому дуга перехода из состояния  $S_1$  в состояние  $S_0$  нагружена интенсивностью  $\mu$ . Пусть теперь система находится в состоянии  $S_2$  (работают два канала). Чтобы ей перейти в  $S_1$ , нужно, чтобы закончил обслуживание первый канал, либо второй. Суммарная интенсивность их потоков равна  $2\mu$  и т.д.

Выходные характеристики (характеристики эффективности) данной СМО определяются следующим образом.

**Абсолютная пропускная способность:**

$$A = \lambda \left[ 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{P_0}{n!} \right], \text{ шт / в. времени,}$$

где  $n$  – количество каналов СМО;

$P_0$  – вероятность нахождения СМО в начальном состоянии, когда все каналы свободны (финальная вероятность нахождения СМО в состоянии  $S_0$ );

Для того, чтобы написать формулу для определения  $P_0$ , рассмотрим рис.16.3.

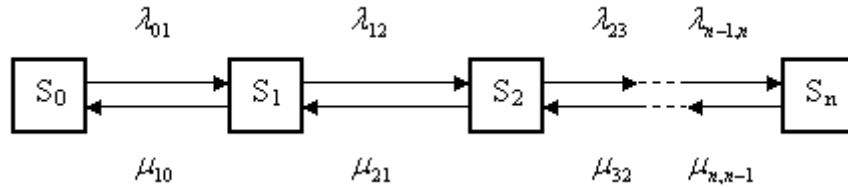


Рис.16.3. Граф состояний для схемы «гибели и размножения»

Граф, представленный на этом рисунке, называют еще графом состояний для схемы «гибели и размножения». Напишем сначала для  $p_0$  общую формулу (без доказательства):

$$p_0 = \left( 1 + \frac{\lambda_{01}}{\mu_{10}} + \frac{\lambda_{12}\lambda_{01}}{\mu_{21}\mu_{10}} + \frac{\lambda_{23}\lambda_{12}\lambda_{01}}{\mu_{32}\mu_{21}\mu_{10}} + \dots + \frac{\lambda_{n-1,n} \dots \lambda_{23}\lambda_{12}\lambda_{01}}{\mu_{n,n-1} \dots \mu_{32}\mu_{21}\mu_{10}} \right)^{-1}.$$

Кстати, остальные финальные вероятности состояний СМО запишутся следующим образом. Вероятность того, что СМО находится в состоянии  $S_1$ , когда один канал занят:

$$p_1 = \frac{\lambda_{01}}{\mu_{10}} p_0.$$

Вероятность того, что СМО находится в состоянии  $S_2$ , т.е. когда два канала заняты:

$$p_2 = \frac{\lambda_{12}}{\mu_{21}} p_1 = \frac{\lambda_{12}\lambda_{01}}{\mu_{21}\mu_{10}} p_0.$$

Вероятность того, что СМО находится в состоянии  $S_n$ , т.е. когда все каналы заняты.

$$p_n = \frac{\lambda_{n-1,n} \dots \lambda_{23}\lambda_{12}\lambda_{01}}{\mu_{n,n-1} \dots \mu_{32}\mu_{21}\mu_{10}} p_0.$$

Теперь для  $n$  – канальной СМО с отказами

$$p_0 = \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{1 \times 2 \mu^2} + \frac{\lambda^3}{1 \times 2 \times 3 \mu^3} + \dots + \frac{\lambda^n}{n! \mu^n} \right)^{-1}.$$

При этом

$$p_1 = \frac{\lambda}{\mu} p_0;$$

$$p_2 = \frac{\lambda^2}{2 \mu^2} p_0;$$

$$p_n = \frac{\lambda^n}{n! \mu^n} p_0.$$

Относительная пропускная способность:

$$Q = 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!}.$$

Напомним, что это средняя доля заявок, обслуживаемых системой. При этом

$$A = \lambda Q;$$

$$Q = 1 - P_{отк}.$$

**Вероятность отказа:**

$$P_{отк} = \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!}.$$

Напомним, что это вероятность того, что заявка покинет СМО необслуженной. Очевидно, что  $P_{отк} = 1 - Q$ . Среднее число занятых каналов (среднее число заявок, обслуживаемых одновременно):

$$\bar{k} = \frac{\lambda}{\mu} \left[ 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!} \right]$$

При этом

$$\bar{k} = \frac{\lambda}{\mu} Q.$$

**Пример.** Имеется технологическая система (участок), состоящая из трех одинаковых станков. В систему поступают для обработки детали в среднем через 0,5 часа ( $\bar{t}_3$ ). Среднее время изготовления одной детали  $\bar{t}_{об} = 0,6$  ч. Если при поступлении заявки на изготовление детали все станки заняты, то деталь направляется на другой участок таких же станков. Найти финальные вероятности состояний системы и характеристики (показатели эффективности) данной СМО.

$$\lambda = \frac{1}{\bar{t}_3} = \frac{1}{0,5} = 2$$

т.е. в среднем две заявки на обработку деталей в час.

$$\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{0,6} \cong 1,67$$

Граф состояний системы представлен на рис.16.4.

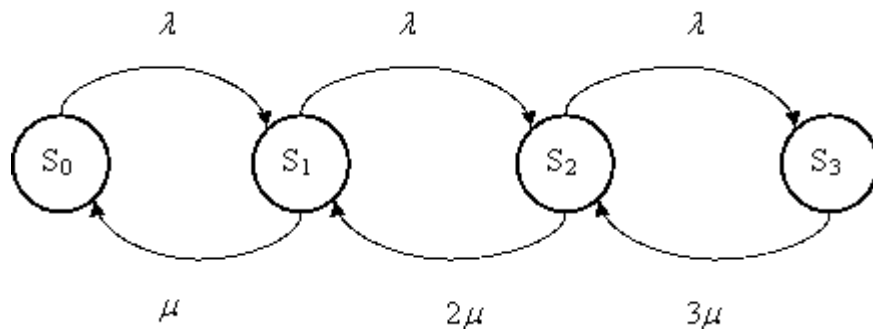


Рис.16.4.Граф состояний для рассматриваемого примера

Возможные состояния системы:

S<sub>0</sub> – в СМО (на участке) нет ни одной заявки;

S<sub>1</sub> – в СМО (на участке) одна заявка;

S<sub>2</sub> – в СМО (на участке) две заявки;

S<sub>3</sub> – в СМО (на участке) три заявки (заняты все три станка).

Вероятность того, что все станки свободны:

$$p_0 = \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{2 \cdot 3\mu^3} \right)^{-1} = \left( 1 + \frac{2}{1,67} + \frac{2^2}{2 \cdot 1,67^2} + \frac{2^3}{2 \cdot 3 \cdot 1,67^3} \right)^{-1} = \frac{1}{3,21} \cong 0,31.$$

Вероятность того, что один станок занят:

$$p_1 = \frac{\lambda}{\mu} p_0 = \frac{2}{1,67} 0,31 \cong 0,37.$$

Вероятность того, что два станка заняты:

$$p_2 = \frac{\lambda^2}{2\mu^2} p_0 = \frac{2^2}{2 \cdot 1,67^2} 0,31 \cong 0,22.$$

Вероятность того, что все три станка заняты:

$$p_3 = \frac{\lambda^3}{3! \mu^3} p_0 = \frac{2^3}{2 \cdot 3 \cdot 1,67^3} 0,31 \cong 0,09.$$

$$A = \lambda \left[ 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!} \right] = 2 \left[ 1 - \left( \frac{2}{1,67} \right)^3 \frac{0,31}{1 \cdot 2 \cdot 3} \right] \cong 1,82 \text{дет/ч.}$$

$$Q = 1 - \left(\frac{\lambda}{\mu}\right)^n \frac{p_0}{n!} = 1 - \left(\frac{2}{1,67}\right)^3 \frac{0,31}{1 \cdot 2 \cdot 3} \cong 0,91; P_{отк} = \left(\frac{\lambda}{\mu}\right)^n \frac{p_0}{n!} = 1 - Q \cong 0,09.$$

$$\bar{k} = \frac{\lambda}{\mu} \left[ 1 - \left(\frac{\lambda}{\mu}\right)^n \frac{p_0}{n!} \right] = \frac{2}{1,67} 0,91 \cong 1,09.$$

Т.е. в среднем в этой системе обрабатывается 1,82 дет/ч (примерно 91 % направляемых деталей), при этом примерно 9 % деталей направляется для обработки на другие участки. Одновременно в среднем работает в основном один станок ( $\bar{k} = 1,09$ ). Но из-за случайных характеристик потока заявок иногда работают одновременно все три станка ( $p_3 = 0,09$ ), отсюда 9 % отказов.

### Возможные постановки задач оптимизации п – канальных СМО с отказами

1. Определить оптимальное число каналов, обеспечивающее минимум затрат на систему, при условии достижения требуемого уровня ее безотказной работы.

**Пример.** Пусть  $\frac{\lambda}{\mu} = 1, P_{отк} \leq 0,03$  (м.е.  $\leq 3\%$ ). Целевая функция (затраты на СМО) запишется:  $y = cn \rightarrow \min$ , где  $c - const$ . Найти:  $n_{опт}$ .

**Решение:**

$$P_{отк} = \left(\frac{\lambda}{\mu}\right)^n \frac{p_0}{n!};$$

$$\frac{\lambda}{\mu} = 1 \Rightarrow P_{отк} = \frac{p_0}{n!}$$

$$P_{отк} \leq 0,03 \Rightarrow \frac{p_0}{n!} \leq 0,03$$

или

$$\frac{n!}{p_0} \geq 33$$

По-другому можно записать:

$$n! \left( 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} \right) \geq 33$$

Последнее равенство начинает выполняться при  $n_{опт} = 4$ , т.к.

$$n = 1 \rightarrow 1 \left( 1 + \frac{1}{1} \right) = 2 < 33 \quad ; \quad n = 2 \rightarrow 1 \cdot 2 \left( 1 + \frac{1}{1} + \frac{1}{1 \cdot 2} \right) = 5 < 33 ;$$

$$n = 3 \rightarrow 1 \cdot 2 \cdot 3 \left( 1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} \right) = 16 < 33 ;$$

$$n = 4 \rightarrow 1 \cdot 2 \cdot 3 \cdot 4 \left( 1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} \right) \cong 65 > 33$$

2. Определить оптимальное число каналов, обеспечивающее максимум прибыли от эксплуатации СМО в единицу времени.

Содержание каждого канала в единицу времени обходится в какую-то сумму. Чем больше каналов, тем больше затраты на эксплуатацию СМО. Вместе с тем, чем больше каналов (при  $\lambda$  и  $\mu - const$ ), тем больше доля обслуживаемых заявок. А каждая обслуженная заявка дает определенный (пусть постоянный) доход в единицу времени. При увеличении числа каналов растут доходы  $D$ , но растут и расходы на эксплуатацию СМО –  $R$ . Чтобы решить эту задачу, необходимо найти оптимальное число каналов  $n_{опт}$ , обеспечивающее максимум целевой функции  $P = D - R \rightarrow \max$ , т.е. нужно максимизировать прибыль в единицу времени.



## Лекция 17 Основы теории производительности и надежности автоматических и автоматизированных станочных систем

### Основные понятия о производительности и надежности автоматических линий

**Производительность автоматической линии (АЛ)** – это количество деталей, изготовленных в единицу времени.

Однако без простоев АЛ работать не может, следовательно, к понятию производительности нужно подходить дифференцированно. Различают следующие виды производительности АЛ:

1. Цикловую.
2. Потенциальную.
3. Фактическую (реальную)

**Цикловая производительность** определяется по следующей формуле:

$$Q_{\text{ц}} = \frac{60}{T_{\text{ц}}} = \frac{60}{t_o + t_e}, \frac{\text{шт}}{\text{час}},$$

где  $T_{\text{ц}}$  - время цикла, мин;

$t_o$  - основное время, мин;

$t_e$  - вспомогательное время на отвод и подвод инструментов, транспортирование деталей от позиции к позиции и т.д., мин.

При учете (расчете) цикловой производительности условно предполагают, что инструменты работают «бесконечно», оборудование не ремонтируется. В действительности это не так, и необходимо учитывать так называемые «внецикловые» простои АЛ. К ним относятся:

$\sum T_{\text{н}}$  - простои во время замены, регулировки и подналадки инструментов в плановый период (например, в месяц);

$\sum T_{\text{об}}$  - простои во время ремонта, регулировки и отладки различных механизмов АЛ в плановый период.

Если величину этих потерь времени отнести к одной детали, т.е. разделить на количество деталей  $Q$ , обработанных в плановый период, то можно записать формулу для **потенциальной производительности**:

$$Q_{\text{п}} = \frac{60}{T_{\text{ц}} + \sum t_{\text{н}} + \sum t_{\text{об}}}, \frac{\text{шт}}{\text{час}},$$

где  $\sum t_{\text{н}} = \frac{\sum T_{\text{н}}}{Q}$ , мин;  $\sum t_{\text{об}} = \frac{\sum T_{\text{об}}}{Q}$ , мин.

При реальной эксплуатации АЛ могут иметь место потери времени по **организационным причинам** из-за отсутствия заготовок, электроэнергии, рабочего и т.д. -  $\sum t_{\text{орг}}$ . Учет этих потерь позволяет судить о **фактической** или **реальной производительности**:

$$Q_{\text{ф}} = \frac{60}{T_{\text{ц}} + \sum t_{\text{н}} + \sum t_{\text{об}} + \sum t_{\text{орг}}}, \frac{\text{шт}}{\text{час}}.$$

Очевидно, что выполняется следующее соотношение:  $Q_{\text{ф}} < Q_{\text{п}} < Q_{\text{ц}}$ .

Эффективность АЛ, ее технический уровень характеризует **коэффициент технического использования**:

$$\eta_{\text{т.и}} = \frac{Q_{\text{п}}}{Q_{\text{ц}}} = \frac{T_{\text{ц}}}{T_{\text{ц}} + \sum t_{\text{н}} + \sum t_{\text{об}}}.$$

Разделим числитель и знаменатель на  $T_{\text{ц}}$ , получим следующее выражение:

$$\eta_{\text{т.и}} = \frac{T_{\text{ц}}/T_{\text{ц}}}{T_{\text{ц}}/T_{\text{ц}} + \sum t_{\text{н}}/T_{\text{ц}} + \sum t_{\text{об}}/T_{\text{ц}}} = \frac{1}{1 + B_{\text{н}} + B_{\text{об}}},$$

где  $B_M = \sum t_M / T_{Ц}$  - удельная, т.е. отнесенная к одной минуте работы линии длительность замены и наладки инструментов;

$B_{ОБ} = \sum t_{ОБ} / T_{Ц}$  - удельная длительность устранения отказов механизмов линии.

Формула для потенциальной производительности может быть записана следующим образом:

$$Q_{П} = Q_{Ц} \eta_{Г.М} = \frac{Q_{Ц}}{1 + B_M + B_{ОБ}}$$

Следует заметить, что коэффициент технического использования  $\eta_{Г.М} < 1$ . Величина  $(1 - \eta_{Г.М})$  характеризует долю времени, в течение которого АЛ простаивает из-за плановых и внеплановых ремонтов механизмов, их замены и подналадки.

**Коэффициент общего использования АЛ** определяется:

$$\eta_{\Sigma} = \frac{Q_{\Phi}}{Q_{Ц}}$$

Он показывает, какую часть действительного времени АЛ работает. Величина  $(1 - \eta_{\Sigma})$  характеризует долю времени, соответствующую простоям по техническим и организационным причинам.

$$\eta_{\Sigma} = \frac{T_{Ц}}{T_{Ц} + \sum t_M + \sum t_{ОБ} + \sum t_{ОРГ}} = \frac{1}{1 + B_M + B_{ОБ} + B_{ОРГ}}$$

где  $B_{ОРГ}$  - удельные, т.е. отнесенные к одной минуте работы потери времени по организационным причинам.

Формула для фактической производительности теперь запишется следующим образом:

$$Q_{\Phi} = Q_{Ц} \eta_{\Sigma} = \frac{Q_{Ц}}{1 + B_M + B_{ОБ} + B_{ОРГ}}$$

Коэффициент технического использования  $\eta_{Г.М}$  АЛ, существенно влияя на ее производительность, зависит от количества узлов, сложности компоновки, **надежности каждого узла и инструмента, т.е. от надежности АЛ в целом.**

**Надежность АЛ** – ее свойство выполнять функции, сохраняя во времени значения установленных эксплуатационных показателей в заданных пределах, соответствующих заданным режимам и условиям использования, технического обслуживания, ремонта.

**Функция АЛ** – ее средняя производительность при условии обработки деталей в соответствии с техническими требованиями чертежа и запланированной себестоимостью обработки.

Надежность АЛ характеризуется долговечностью, безотказностью и ремонтпригодностью оборудования.

**Долговечность** – это свойство АЛ длительно (с перерывами на ремонт) сохранять свою работоспособность при определенных режимах работы и условиях эксплуатации.

Количественно она оценивается средним сроком службы между ремонтами.

**Безотказность** – это свойство линии непрерывно сохранять работоспособность в течении некоторого времени.

Количественно она оценивается:

- вероятностью безотказной работы;
- интенсивностью отказов;
- наработкой на отказ и т.д.

В частности, наработка на отказ – это среднее значение времени работы между двумя последовательно возникающими отказами.

**Ремонтпригодность** – это приспособленность узлов и механизмов линии к предупреждению и обнаружению причин возникновения отказов, повреждений и устранению их последствий путем ремонтов и технического обслуживания.

Количественно она оценивается средним временем восстановления работоспособности.

## Расчет производительности гибких производственных систем

Применение в машиностроении гибких производственных систем (ГПС) – важнейший путь повышения эффективности производства. Данные системы требуют значительных капиталовложений, поэтому важным моментом при их внедрении является правильная оценка их эффективности, которая характеризуется рядом показателей. Одним из важнейших показателей эффективности ГПС является ее производительность

Производительность ГПС зависит от производительности станков, входящих в ее состав, а также от производительности таких ее составляющих, как транспортно – накопительная система, автоматизированная система инструментального обеспечения и т.д.

Производительность следует рассчитывать как часовую, суточную, так и по полному годовому фонду времени, что позволит учитывать степень автоматизации, степень гибкости и возможности данной ГПС работать в малолюдном («безлюдном») режиме. При круглосуточной, ежедневной, круглогодичной работе полный годовой фонд времени составляет 8760 ч. Фактический годовой фонд времени работы заданной ГПС будет меньше и будет определяться надежностью ее работы, способностью длительное время (ночная и вечерняя смены или несколько праздничных дней подряд) работать в «безлюдном» режиме, а также суммарным объемом простоев всех видов в расчете на год.

Отношение фактического времени работы ГПС к полному (установленному) годовому фонду времени дает качественную оценку возможностей той или иной ГПС, ее совершенства, является показателем того, насколько решены в ней проблемы использования возможностей гибкого производства:

$$k = \frac{F_{\text{ф}}}{F},$$

где  $k$  - коэффициент использования годового времени, или коэффициент годовой загрузки ГПС, или коэффициент совершенства гибкого производства;

$F_{\text{ф}}$  - фактическое годовое время работы ГПС, ч;

$F$  - годовой фонд времени, равный 8760 ч.

В свою очередь

$$F_{\text{ф}} = d_y h_y - \sum t_i,$$

где  $d_y$  - установленное (планируемое) количество рабочих дней в году с учетом субботах, выходных и праздничных дней;

$h_y$  - количество установленных часов работы ГПС в сутки (16, 20 или 24 ч. в сутки, из которых какое-то количество в «безлюдном» режиме);

$\sum t_i$  - суммарные фактические простои станков и всей ГПС в расчете на год.

Чем ближе значение  $k$  к единице, тем совершеннее ГПС.

Фактическое годовое время работы ГПС в часах и составляет ее производительность в станко – часах, т.е. по суммарной станкочасности обработки всех деталей за год

$$F_{\text{ф}} = Q_{\text{ГПС}}.$$

Среднестручная годовая производительность может быть рассчитана следующим образом:

$$Q_{\text{шт}} = \frac{60 F_{\text{ф}} N}{t_{\text{шт. ср}}},$$

где  $N$  - количество станков в ГПС;

$t_{\text{шт. ср}}$  - среднее штучное время по всем деталям, обрабатываемым в ГПС, мин.

Среднестручная производительность за сутки  $Q_{\text{сут}}$ , час  $Q_{\text{час}}$  или в минуту  $Q_{\text{мин}}$  может быть рассчитана соответственно по формулам:

$$Q_{\text{сут}} = \frac{Q_{\text{шт}}}{d_y};$$

$$Q_{\text{час}} = \frac{Q_{\text{шт}}}{d_y h_y};$$

$$Q_{\text{факт}} = \frac{Q_{\text{шт}}}{60d_y h_y}$$

Фактическая производительность ГПС за любой период времени определяется с учетом суммарных простоев отдельных станков в системе и всей ГПС ( $\sum t_i$ ) с учетом всех потерь времени (простоев), которые происходят по разным причинам.

Учитывая возможность работы ГПС круглосуточно и круглогодично, при расчете потерь времени необходимо исходить из годового фонда времени, а не одной смены или одних суток. При этом следует рассматривать простои не одного станка а системы станков и других агрегатов ГПС и связанных с ними участков работы (участка сборки приспособлений и установки заготовок, участка настройки инструментов на размер вне станка, участка контроля деталей и т.д.). Это позволит определить сумму потерь всего годового фонда времени в одной ГПС.

Кроме этого, следует учитывать, что меняется состав различных видов потерь. С повышением степени автоматизации ГПС отдельные потери переходят из внецикловых видов потерь в цикловые. Например при работе станка типа «обрабатывающий центр» (ОЦ) в ГПС смена инструмента между технологическими переходами, а также из – за потери им стойкости при наличии автоматических устройств слежения за стойкостью (периодический замер режущей кромки инструмента щупом – пробой) является частью цикла, вместе с другими холостыми ходами составляет цикловые потери и входит в штучное программируемое время обработки детали на ОЦ в ГПС. К внецикловым потерям, связанным с работой, в ГПС следует относить, например, отдельные виды простоев из – за поломки инструмента, когда нет автоматических устройств слежения за стойкостью инструмента или требуется вмешательство оператора для восстановления работы станка.

Итак, при расчете производительности ГПС следует учитывать следующие потери времени:

1. Цикловые потери рабочего времени (замена инструмента в шпинделе, ускоренный подвод и отвод инструмента, координатное перемещение стола и т.д.);
2. Простои по вине инструмента (замена инструмента из – за поломок, из – за затупления, регулировка без замены – в случае отсутствия автоматизации и необходимости вмешательства оператора);
3. Простои по вине оборудования (регулировка и ремонт механической части оборудования, ремонт гидравлических систем, отказ системы управления и т.д.);
4. Простои по организационным причинам (отсутствие электроэнергии, вспомогательных материалов);
5. Простои, связанные с браком (браком предыдущих операций, браком материалов, выявленных во время обработки, браком обработки и т.д.);
6. Простои, связанные с переналадкой (замена оснастки станка, замена комплекта инструмента).

Состав и виды потерь рабочего времени при различных сроках наблюдения меняются и зависят от разных факторов. При наблюдении работы оборудования в течении 1 ч производительность процесса обработки зависит от производительности станка, технологии, дифференциации или концентрации обработки, смены инструментов, заготовок, возможно, от отказов станков. При наблюдении в течении суток добавляются простои, связанные со сменой задания, установленными перерывами (обед, сдача смены) и случайными простоями (отсутствие заготовок, рабочих и т.д.).

При наблюдении за год выявляются простои, связанные с переналадкой оборудования для выпуска новой продукции, неравномерностью загрузки оборудования, неплановым ремонтом, нехваткой рабочих в период отпусков и т.д. Если же наблюдать работу оборудования ГПС за весь период его эксплуатации, то добавляются простои, связанные с его ремонтом, модернизацией, перестройкой на новую продукцию и др.

## Лекция 18. Производительность и надежность автоматических и автоматизированных станочных систем

### Производительность и надежность заблокированных автоматических линий

Математические модели, описывающие производительность и надежность автоматических линий (АЛ), рассмотрим на примере математических моделей заблокированных АЛ. В заблокированных АЛ все элементы соединены последовательно в смысле надежности, и отказ любого элемента приводит к отказу всей системы.

Прежде, чем разрабатывать математические модели АЛ, нужно выяснить, по какому закону распределены потоки отказов и потоки восстановлений (см. лекции по «теории массового обслуживания») механизмов и инструментов линии, т.к. это будет определять, какие теоретические положения будут заложены в основу модели.

Впервые изучение вероятностных характеристик АЛ было начато в ЭНИМСе. При этом исходили из предположения, что потоки отказов и потоки восстановлений являются пуассоновскими, а периоды времени безотказной работы  $\tau_{РАБ}$  и восстановления  $\tau_{В}$  АЛ распределены по экспоненциальному закону. Вспомним, что для простейшего (пуассоновского) потока с интенсивностью  $\lambda$  интервал  $T$  между соседними событиями имеет показательное (экспоненциальное) распределение с плотностью  $f(t) = \lambda e^{-\lambda t}$ , где  $\lambda$  - параметр показательного закона (интенсивность потока событий).

Позднее были проведены исследования этих потоков. Причем, в отличие от ранее выполненных работ потоки отказов и восстановлений оборудования и инструментов исследовали отдельно, т.к. природа этих потоков различна. Были получены следующие основные результаты.

Среднестатистически для всех исследованных объектов экспоненциальный закон достаточно точно характеризует распределение случайных периодов восстановления оборудования. А для отдельных станков и инструментов экспериментальные законы распределения времени  $\tau_{РАБ}$  (работы),  $\tau_{РЕМ}$  (ремонта),  $\tau_{СМ.И}$  (смены инструментов) отличаются от теоретического экспоненциального – рис.18.1.

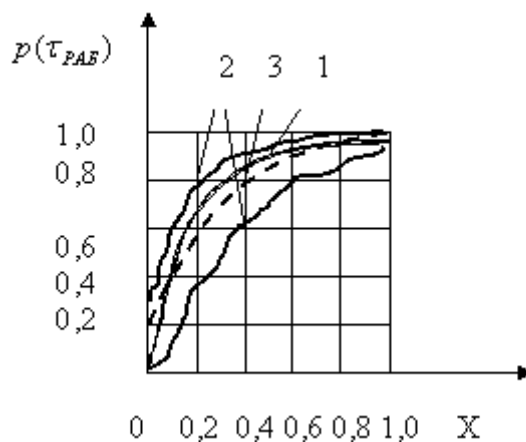


Рис.18.1. Кривые распределения времени безотказной работы оборудования АЛ

На этом рисунке:  $P(\tau_{РАБ})$  - частота;  $X$  - относительные единицы; 1 – теоретическая кривая; 2 – экспериментальные кривые; 3 – среднестатистическая кривая. Расхождение между кривыми (1) и (3) – 7 – 10 %, а для отдельных станков расхождение между экспериментальными и теоретическими кривыми еще более значительное. Поэтому, показательный закон применяют в математических моделях, используемых для первоначальной оценки надежности станочных систем на ранней стадии их проектирования, когда не известны еще конструкции узлов и инструментов. Для точной оценки функционирования конкретных станков и линий следует использовать математические модели с применением тех законов распределения, которые будут соответствовать реальным условиям эксплуатации оборудования.

Мы рассмотрим первый вид математических моделей для простейших (пуассоновских) потоков, тем более, что нам известны соответствующие положения теории массового обслуживания, построенные на основе Марковских процессов.

Пусть заблокированная станочная система состоит из  $m$  – узлов и  $n$  – инструментов. Т.к. при такой компоновке все ее элементы соединены последовательно в смысле надежности и отказ любого элемента приводит к отказу всей системы, то вероятность отказа линии определится

$$F(t) = 1 - (1 - P_1)(1 - P_2) \dots (1 - P_{m+n}),$$

где  $P_1, P_2, \dots, P_{m+n}$  - вероятность того, что 1, 2, ...,  $(m + n)$  – й элемент АЛ находится в работоспособном состоянии.

Выделим три состояния заблокированной системы ( АЛ ) – рис.18.2.

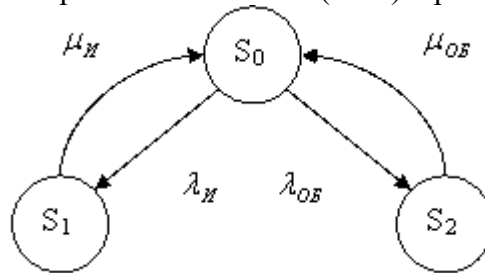


Рис.18.2. Граф состояний заблокированной АЛ

На этом рисунке приняты следующие обозначения:  $S_0$  - все элементы работоспособны;  $S_1$  - отказ по вине инструментов;  $S_2$  - отказ по вине механизмов (оборудования);  $\lambda_M, \lambda_{OB}$  – интенсивности потоков отказов инструментов и оборудования;  $\mu_M, \mu_{OB}$  - интенсивности потоков восстановлений инструментов и оборудования.

В соответствие с положениями теории массового обслуживания вероятность того, что АЛ находится в состоянии  $S_0 \rightarrow p_0, S_1 \rightarrow p_1, S_2 \rightarrow p_2$  определится:

$$p_0 = \frac{1}{1 + \lambda_M / \mu_M + \lambda_{OB} / \mu_{OB}};$$

$$p_1 = \frac{(\lambda_M / \mu_M)}{1 + \lambda_M / \mu_M + \lambda_{OB} / \mu_{OB}};$$

$$p_2 = \frac{(\lambda_{OB} / \mu_{OB})}{1 + \lambda_M / \mu_M + \lambda_{OB} / \mu_{OB}}.$$

Вероятность пребывания линии в исправном состоянии характеризуется **коэффициентом готовности** -  $\eta_r$ . Отношение  $\lambda / \mu = B$  - это удельная длительность настройки элемента или группы элементов. Таким образом

$$p_0 = \eta_r = \frac{1}{1 + B_M + B_{OB}};$$

$$p_1 = \frac{B_M}{1 + B_M + B_{OB}};$$

$$p_2 = \frac{B_{OB}}{1 + B_M + B_{OB}}.$$

Следует иметь в виду, что коэффициент готовности не равен коэффициенту технического использования АЛ (см. лекцию 17). Коэффициент технического использования по сравнению с коэффициентом готовности дополнительно учитывает плановые простои (ремонт) АЛ.

Если известны интенсивности  $\lambda$  и  $\mu$  для каждого инструмента и узла, то

$$p_0 = \frac{1}{1 + \sum_{i=1}^n (\lambda_{iM} / \mu_{iM}) + \sum_{j=1}^m (\lambda_{jOB} / \mu_{jOB})};$$

$$p_{iM} = \frac{(\lambda_{iM} / \mu_{iM})}{1 + \sum_{i=1}^n (\lambda_{iM} / \mu_{iM}) + \sum_{j=1}^m (\lambda_{jOB} / \mu_{jOB})};$$

$$P_{j_{OB}} = \frac{(\lambda_{j_{OB}} / \mu_{j_{OB}})}{1 + \sum_{i=1}^n (\lambda_{i_{M}} / \mu_{i_{M}}) + \sum_{j=1}^m (\lambda_{j_{OB}} / \mu_{j_{OB}})}$$

Для расчета коэффициента готовности АЛ  $\eta_r = P_0$  путем длительных статистических исследований в производственных условиях были получены значения  $\lambda_i$  и  $B_i$  для различных элементов АЛ. Примеры их значений приведены в таблице 18.1.

Таблица 18.1 Значения  $\lambda_i$  и  $B_i$  для некоторых элементов автоматических линий

Наименования элементов АЛ	$\lambda_i \cdot 10^3, 1/\text{цикл}$	$B_i \cdot 10^3, \text{мин}/\text{цикл}$
Приспособления-спутники для фиксации и зажима заготовок	0,1	0,5
Борштанга с вращающейся или скользящей втулкой и плавающим патроном	0,04	0,01
и т.д.	...	...

Это дает возможность при проектировании АЛ выбирать такие технические средства (механизмы) и инструменты, которые обеспечат требуемый уровень надежности линии ( $\eta_r$ ). Зная  $\eta_r$  ( $\eta_{r.M}$ ) и цикловую производительность, можно рассчитать потенциальную производительность ( $Q_{п} = Q_{ц} \eta_{r.M}$ ) и далее фактическую производительность  $Q_{ф}$ .

### Производительность и надежность гибких производственных систем

На количественные значения производительности гибких производственных систем (ГПС) важное влияние оказывает надежность входящих в ее состав элементов (станков, транспортно – накопительных систем, систем инструментального обеспечения и т.д.). Рассмотрим способы оценки надежности ГПС на примере количественной оценки надежности одной из важнейших ее составляющих - автоматизированной системы инструментального обеспечения (АСИО). Материалы этого раздела взяты из кандидатской диссертации автора данного лекционного курса. Тема диссертации: «Разработка метода выбора структурно – компоновочных решений автоматизированных систем инструментального обеспечения ГПС на ранней стадии проектирования (на примере ГПС для обработки корпусных деталей)»

Для ГПС, состоящих из большого числа взаимосвязанных и взаимозависимых устройств, необходим расчет ее показателей надежности. Отсутствие количественной оценки надежности проектируемой сложной технической системы, к числу которых относится и ГПС, может привести к созданию неработоспособной либо неэффективной системы.

На этапе проектирования для оценки надежности элементов системы и системы в целом целесообразно использовать коэффициент готовности:

$$\eta_r = \frac{T_p}{T_p + T_{нпл}}$$

где  $T_p$  - время работы элемента системы за данный период эксплуатации;

$T_{нпл}$  - время неплановых ремонтов по устранению внезапных отказов оборудования.

Для оперативной системы инструментального обеспечения станка  $T_{нпл} = T_{нпл}^I + T_{нпл}^{II}$ , для других устройств АСИО  $T_{нпл} = T_{нпл}^I$ . Здесь  $T_{нпл}^I$  и  $T_{нпл}^{II}$  - время непланового ремонта по устранению, соответственно, отказов устройств АСИО и поломок режущих инструментов. Т.е.

$$\eta_r = \frac{T_p}{T_p + T_{нпл}^I + T_{нпл}^{II}}$$

Разделим числитель и знаменатель дроби в предыдущем выражении на  $T_p$ .

$$\eta_r = \frac{1}{1 + K_{нпл}^I + K_{нпл}^{II}}$$



Здесь  $K_{НПЛ}' = T_{НПЛ}' / T_p$  - удельная длительность устранения отказов (восстановлений) устройств АСИО;

$K_{НПЛ}'' = T_{НПЛ}'' / T_p$  - удельная длительность замены отказавших инструментов.

Коэффициент готовности является комплексным показателем оценки надежности. Он характеризует одновременно безотказность и ремонтпригодность системы.

Например, для определения уровня численных значений  $K_{НПЛ}'$  устройств автоматической смены инструментов (УАСИ) воспользуемся статистическими данными наблюдений за работой станков типа «обрабатывающий центр», проведенных на Ивановском заводе тяжелого станкостроения.

В таблице 18.1 приведена часть данных по простоям станков, вызванных выходом из строя элементов УАСИ. Исследовались 3 станка модели ИР500ПМФ4 (ИР800ПМФ4) и 2 станка модели ИР320ПМФ4. Нарботка станков на отказ составила 8300 час. Длительность обследования – 2 года.

Таблица 18.1 Данные по простоям станков типа «обрабатывающий центр» по причинам выхода из строя механизмов УАСИ

Сводный перечень отказов (наименование отказавших систем, узлов и деталей)	Причина отказа	Число отказов	Суммарное время восстановления, час	Число отказавших станков
1. Манипулятор	Сломан сухарь	1	20	1
2. Манипулятор	Пробило манжету гидроцилиндра	1	6	1
3. Механизм зажима инструмента шпиндельного устройства	Износ тарельчатых пружин	4	99	3
4. ----- " -----	Вышла из строя Цанга	1	36	1
5. ----- " -----	Сломались Кулачки	1	2,5	1
6. Инструментальный магазин	Магазин не выходит в позицию	1	8	1

Для определения удельной длительности восстановления нескольких станков воспользуемся рекомендациями ЭНИМСа:

$$K_{НПЛ}' = \frac{\sum_{i=1}^m t_{Bi}}{\sum_{j=1}^N t_j}$$

В числителе дроби в этой формуле – суммарное время на обнаружение, поиск причины и устранение последствий  $m$  отказов (сбоев), в знаменателе – суммарная наработка  $N$  станков за время наблюдений. По данным таблицы 18.1  $K_{НПЛ}' = 171,5 / 5 * 8300 \cong 0,004$ . Для разных периодов наблюдений, разного количества обследованных станков приведенных выше моделей  $K_{НПЛ}'$  колеблется от 0,002 до 0,007.

Значение  $K_{НПЛ}'' = 0,24 K_{НПЛ}'$  (приводится без вывода). Отсюда  $K_{НПЛ}'' \cong 0,0005 - 0,0017$ . При условии использования на станках ГПС эффективных средств обнаружения поломок инструментов величина  $K_{НПЛ}''$  будет уменьшаться.

Приведенные выше величины  $K_{НПЛ}'$  и  $K_{НПЛ}''$  могут служить ориентирами при определении  $\eta_r$  для УАСИ станков ГПС. С накоплением опыта эксплуатации ГПС будут уточняться эти величины для различных моделей станков, а также значения  $\eta_r$  для других устройств АСИО.

Вид формул для расчета коэффициента готовности АСИО в целом зависит от структуры системы. Если, например, АСИО ГПС устроена так, что каждый станок имеет автономное УАСИ, то коэффициент готовности определится следующим образом:



$$\eta_{Г(АСИО)} = 1 - (1 - \eta_{Г(УАСИ1)})(1 - \eta_{Г(УАСИ2)}) \dots (1 - \eta_{Г(УАСИn)}),$$

где  $\eta_{Г(УАСИ1)}, \eta_{Г(УАСИ2)}, \dots, \eta_{Г(УАСИn)}$  - коэффициенты готовности, соответственно, УАСИ 1, 2, ..., n - го станков ГПС.

А если, например, АСИО ГПС устроена так, что имеется центральный магазин инструментов (ЦМИ) и техническое средство транспортирования инструментов (ТСТИ), общие для всех станков, то коэффициент готовности такой АСИО определится:

$$\eta_{Г(АСИО)} = \eta_{Г(ЦМИ)} \eta_{Г(ТСТИ)},$$

где  $\eta_{Г(ЦМИ)}, \eta_{Г(ТСТИ)}$  - коэффициенты готовности, соответственно, ЦМИ и ТСТИ.

Для других АСИО коэффициенты готовности могут быть определены по аналогии.

Величина коэффициента готовности для АСИО в целом зависит от коэффициентов готовности входящих в ее состав элементов, а также от структурных решений системы. Возможность организации автономных потоков инструментов, например, в том случае, когда каждый станок ГПС имеет автономное УАСИ, позволяет значительно повысить надежность АСИО и ГПС в целом. Это выгодно отличает такие системы, например, от заблокированных автоматических линий, состоящих из элементов, соединенных последовательно. Выход из строя любого, даже одного такого элемента, приводит к потере работоспособности всей линии.

## Лекция 19. Оперативно – календарное планирование в технологических системах на основе теории расписаний

### Элементы (основы) теории расписаний

Качество функционирования современного производства во многом определяется решениями, принимаемыми на этапах календарного планирования и оперативного управления. Особенно это актуально в связи с созданием современных автоматизированных производств – гибких производственных систем (ГПС). Системы оперативно – календарного планирования современных производств строятся в том числе и на достижениях так называемой «теории расписаний».

**Теория расписаний** – это наука, занимающаяся исследованиями детерминированных обслуживающих систем на предмет оптимизации расписаний их функционирования.

Примеры таких систем:

- цех, участок, на станках которых осуществляется обработка деталей;
- ВУЗ, где преподаватели обучают студентов и т.д.

В любом случае имеется конечное множество **требований** (деталей, преподавателей и т.д.)  $N = \{1, 2, \dots, n\}$  и конечное множество **приборов** (станков, групп студентов и т.д.)  $M = \{1, 2, \dots, m\}$ .

Предполагается, что  $i$  – е требование на каждой стадии его обслуживания  $q$  (например, на каждой операции технологического процесса) может быть обслужено любым из приборов  $L \in M$  (но не более, чем одним одновременно). Предполагается также, что каждый прибор одновременно может обслуживать не более одного требования.

В теории расписаний рассматриваются различные системы обслуживания:

- **системы поточного типа**, в которых каждое требование  $i \in N$  сначала обслуживается приборами первой группы, затем второй группы и т.д. пока не будет обслужено приборами последней  $r$  – ой группы;
- **системы с различными порядками (маршрутами) прохождения приборов требованиями** и т.д.

В частности, в последних системах с последовательными приборами для каждого требования  $i \in N$  задается своя, специфическая для этого требования последовательность  $L^i = (L^i_1, L^i_2, \dots, L^i_n)$  его обслуживания приборами. Требование  $i$  сначала обслуживается прибором  $L^i_1$ , затем  $L^i_2$  и т.д. пока не будет обслужено прибором  $L^i_n$ . Последовательности обслуживания могут быть различными для разных требований и могут содержать повторение приборов.

В любом случае, если требование  $i$  на стадии  $q$  должно или может быть обслужено прибором  $L = L^i_q$ , то предполагается заданной **длительность**  $t_{iL} \geq 0$  его обслуживания прибором. Запись  $t_{iL} = 0$ , как правило, означает, что по условию задачи требование  $i$  на стадии  $q$  прибором  $L$  не обслуживается.

Наряду с величинами  $t_{iL}$  могут быть заданы также: момент  $d_i \geq 0$  поступления требования  $i$  в систему; директивный срок  $D_i > 0$ , к которому необходимо завершить обслуживание требования.

Процесс функционирования обслуживающей системы может быть описан путем задания **расписания** (календарного плана, временного графика и т.п.).

**Расписание** – некоторая совокупность указаний относительно того, какие именно требования какими именно приборами обслуживаются в каждый момент времени.

Расписание рассматривается как совокупность  $\{S_1(t), S_2(t), \dots, S_m(t)\}$  кусочно–постоянных непрерывных слева функций, каждая из которых задана на интервале  $0 \leq t \leq \infty$  и принимает значения  $0, 1, \dots, n$ .

Если  $S_L(t') = i \neq 0$  (здесь  $i$  – номер требования), то в момент времени  $t'$  прибор  $L \in M$  обслуживает требование  $i \in N$ . Если  $S_L(t') = 0$ , то в момент времени  $t'$  прибор  $L$  простаивает.

При задании расписания должны соблюдаться все условия и ограничения, вытекающие из постановки рассматриваемой задачи, т.е. расписание должно быть **допустимым**.

**Пример.** На рис.19.1 приведен график расписания  $S_i(t), i \in N$  обслуживания требований  $N = \{1,2,3,4\}$  приборами  $M = \{1,2,3\}$  при различных маршрутах обслуживания требований. Все длительности обслуживания равны «1».<sup>1</sup>

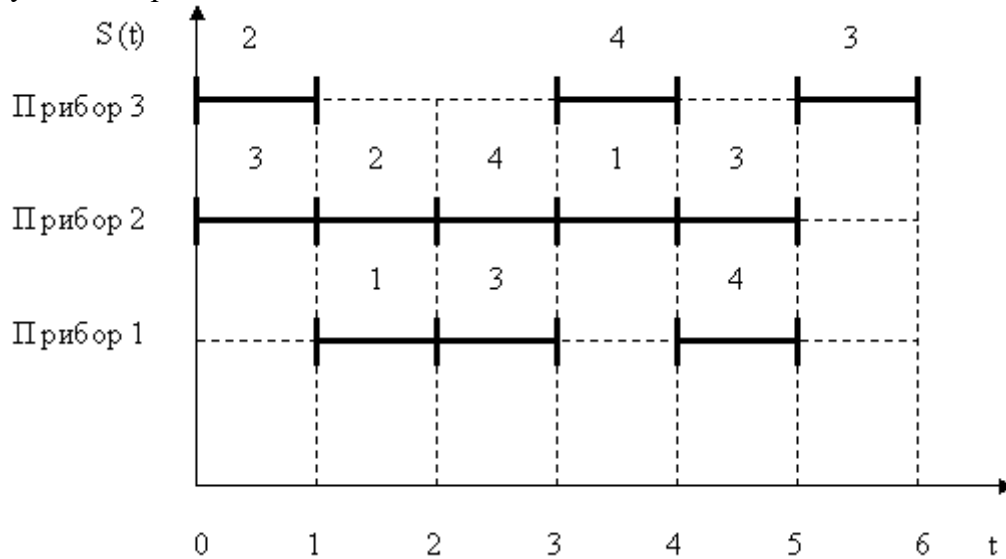


Рис.19.1. График расписания обслуживания требований  $N = \{1,2,3,4\}$  приборами  $M = \{1,2,3\}$

Здесь  $L^1 = (1,2)$ , т.е. первое требование обслуживается первым и вторым приборами,  $L^2 = (3,2)$  – второе требование обслуживается третьим и вторым приборами,  $L^3 = (2,1,2,3)$  – третье требование обслуживается вторым, первым, снова вторым и третьим приборами,  $L^4 = (2,3,1)$  – четвертое требование обслуживается вторым, третьим и первым приборами.  $d_1 = 1$  – момент поступления требования 1 в систему,  $d_2 = d_3 = 0$  – моменты поступления требований 2 и 3 в систему,  $d_4 = 2$  – момент поступления требования 4 в систему.  $D_1 = 4$  – директивный срок завершения обслуживания требования 1,  $D_2 = 2$  – директивный срок завершения обслуживания требования 2,  $D_3 = 6$  – директивный срок завершения обслуживания требования 3,  $D_4 = 5$  – директивный срок завершения обслуживания требования 4.

Прибор 1 во временном интервале  $(1,2]$  обслуживает требование 1, в интервале  $(2,3]$  – требование 3, в интервале  $(4,5]$  – требование 4. Прибор 2 в интервале  $(0,5]$  без простоев обслуживает требования 3, 2, 4, 1, 3 и т.д. Это расписание допустимо, т.е. каждый прибор одновременно обслуживает не более одного требования и  $i$ -е требование обслуживается одновременно не более, чем одним прибором.

Если существует несколько допустимых расписаний, то естественно необходимо выбрать лучшее из них. В теории расписаний **качество расписания** во многих случаях оценивают следующим образом. Каждое (допустимое) расписание  $S$  однозначно определяет вектор  $\bar{t}(S) = (\bar{t}_1(S), \bar{t}_2(S), \dots, \bar{t}_n(S))$  **моментов завершения обслуживания требований**. Задается некоторая действительная неубывающая по каждой из переменных функция  $F(x) = F(x_1, x_2, \dots, x_n)$ . Качество расписания  $S$  оценивается значением этой функции при  $x = \bar{t}(S)$ . Из двух расписаний лучшим считается то, которому соответствует меньшее значение  $F(x)$ . Расписание, которому соответствует **наименьшее значение  $F(x)$**  (среди всех допустимых расписаний), называется **оптимальным**.

В частности, при построении **оптимального по быстродействию** расписания  $F(x) = \max\{x_i\}, i = 1, 2, \dots, n$ . В этом случае  $F(\bar{t}(S)) = \bar{t}_{\max}(S)$ , где  $\bar{t}_{\max}(S) = \max\{\bar{t}_i(S)\}, i = 1, 2, \dots, n$ .

При построении расписания с **наименьшим суммарным временем обслуживания**

$$F(x) = \sum_{i=1}^n x_i, \text{ при этом } F(\bar{t}(S)) = \sum_{i=1}^n \bar{t}_i(S).$$

При построении расписания с **наименьшим временем смещения** моментов завершения обслуживания требований  $i$  относительно сроков  $D_i$  функция  $F(x) = \max\{x_i - D_i\}, i = 1, 2, \dots, n$ . При этом  $F(\bar{t}(S)) = L_{\max}(S)$ , где  $L_{\max}(S) = \max\{\bar{t}_i(S) - D_i\}, i = 1, 2, \dots, n$ .

Оптимальное расписание может быть найдено в результате перебора конечного множества возможных вариантов. Основная трудность при этом состоит в том, что число таких вариантов очень велико и растет, по меньшей мере, экспоненциально с ростом размерности задачи. Известны так называемые эвристические алгоритмы формирования расписаний, алгоритмы на основе методов линейного и динамического программирования. Задачи составления расписаний для некоторых сложных систем обслуживания до сих пор не решены (NP – трудные задачи).

### **Формирование расписания работы оборудования методами линейного и динамического программирования**

Эта методика разработана в лаборатории исследования операций Ленинградского (ныне Санкт-Петербургского) государственного университета под руководством профессора И.В. Романовского.

#### **Исходные данные для решения задачи:**

1. Количество рассматриваемых видов деталей  $M$ . Виды деталей нумеруются числами  $m = 1, 2, \dots, M$ .

2. Количество групп однотипного оборудования  $I$ . Группы оборудования нумеруются числами  $i = 1, 2, \dots, I$ .

3. Технологические маршруты (ТМ) обработки деталей. ТМ не содержат внешних операций, т.е. операций, которые выполняются на другом оборудовании.

Для каждого вида деталей  $m$  ( $m = \overline{1, M}$ ) задаются:

- количество операций в ТМ –  $J^m$ , номера операций в ТМ обозначаются через  $j$  ( $j = 1, 2, \dots, J^m$ );
- продолжительность обработки одной детали на операции  $j$  –  $t_j^m$  (при обработке деталей  $m$ ),  $j = \overline{1, J^m}$ ;
- номер группы оборудования –  $i_j^m$ , на котором выполняется операция  $j$ ,  $j = \overline{1, J^m}$ .

4. План выпуска деталей различных видов – вектор  $D = (d^1, d^2, \dots, d^M)$ .

5. Стоимость  $\gamma^m$  пролеживания деталей вида  $m$  ( $m = \overline{1, M}$ ) в единицу времени.

Пусть отрезок планирования разбит на  $S$  частей, которые для простоты будем называть сутками и нумеровать числами  $s = 1, 2, \dots, S$ . Для каждого сутка должны быть заданы следующие величины:

6. Продолжительность суток  $\sigma_s, s = 1, 2, \dots, S$ .

7. Фонд времени групп оборудования  $i$  в сутки  $s$  –  $f_{is}, i = 1, 2, \dots, I; s = 1, 2, \dots, S$ .

План выпуска деталей каждого вида разбивается на партии обработки. Обозначим число партий обработки  $P$  и введем для них единую нумерацию, не зависящую от вида деталей:  $p = 1, 2, \dots, P$ . Будем считать известной функцию  $m(p)$ , которая по номеру партии  $p, p = \overline{1, P}$  дает номер вида детали  $m, m = \overline{1, M}$  (в частном случае, если запускается в обработку по одной партии каждого вида деталей, то  $P = M$  и  $P = M$ ).

Количество деталей в партии обработки  $p$  обозначим  $x_p, p = \overline{1, P}$ ; тогда должны иметь место равенства

$$\sum_{m(p)=m} x_p = d^m, m = \overline{1, M}, \quad (19.1)$$

т.е. сумма размеров всех партий для каждой детали равна плану ее выпуска.

**Задача формулируется следующим образом:** найти число партий обработки  $P$ , количество деталей  $x_p$  в партиях обработки  $p = 1, 2, \dots, P$  и расписания работы оборудования (обработки каждой партии), оптимальные с точки зрения некоторого критерия.

Задача допускает выбор используемого критерия в широких пределах. Ниже в качестве критерия будет использоваться **стоимость пролеживания деталей**.

**Постановка задачи в терминах линейного программирования:** обозначим  $Q = S \cdot I$ , где  $S$  – количество суток,  $I$  – количество групп оборудования. Рассмотрим  $Q$  – мерный вектор  $F$ , который определяется следующим образом:

Вектор F	$f_{11}$	$f_{12}$	...	$f_{1S}$	$f_{21}$	$f_{22}$	...	$f_{2S}$	...	$f_{I1}$	$f_{I2}$	...	$f_{IS}$
Номер суток $s$	1	2	...	$S$	1	2	...	$S$	...	1	2	...	$S$
Номер группы оборудования $i$	1				2				...	I			

Первые компоненты вектора  $F$  есть фонды времени первой группы оборудования в  $1, 2, \dots, S$  сутки отрезка планирования, вторые – фонды времени второй группы оборудования в те же сутки и т.д., например,  $f_{22}$  – фонд времени второй группы оборудования во вторые сутки.

Если записать вектор  $F$  как  $F = (F_1, \dots, F_q, \dots, F_Q)$ , то  $F_q = f_{is}$ , когда  $q = (i-1)S + s$ ,  $i = [(q-1)/S] + 1, s = q - (i-1)S = q - [(q-1)/S]S$ , где  $[a]$  – целая часть числа  $a$ . Эти равенства позволяют по номеру группы оборудования  $i$  и номеру суток  $s$  определить номер  $q$  компоненты вектора  $F_q$  и наоборот – по порядковому номеру  $q$  компоненты вектора  $F_q$  соответствующие ему номера групп оборудования  $i$  и суток  $s$  у компоненты вектора  $f_{is}$ .

**Пример.**  $S = 30$  суток.

1) Нужно определить порядковый номер компоненты вектора  $F_q$  для второй группы оборудования на третьи сутки его работы

$$(i = 2, s = 3). q = (i-1)S + s = (2-1)30 + 3 = 33$$

2) Найти  $i, s$ , если  $q = 33$ .

$$i = [(33-1)/30] + 1 = 1 + 1 = 2.$$

$$s = 33 - [(33-1)/30]30 = 33 - 1 \cdot 30 = 3$$

Для каждой партии с номерами  $p, p = \overline{1, P}$  рассмотрим вектор  $r_p$ , структура которого аналогична структуре вектора  $F$ :  $r_p = (r_{1p}, \dots, r_{qp}, \dots, r_{Ip})$ . Вектор  $r_p$  – это элементарное расписание обработки партии  $p$ . Его компоненты определяются следующим образом.

Если операция  $j$  по обработке детали из партии  $p$  выполняются в сутки  $s$  на группе оборудования  $i = i_j^{m(p)}$ , то в компоненту  $r_{qp}$  вектора с порядковым номером  $q(i, s)$ , вычисленным по приведенным выше формулам для  $q, i, s$ , заносят время  $t_j^{m(p)}$  обработки этой детали на операции  $j$ . Все остальные компоненты  $r_{qp}$  такие, что в сутки  $s$  на группе оборудования  $i$  не обрабатываются детали партии  $p$ , полагаются равными нулю.

Элементарные расписания являются расписанием обработки одной детали партии  $p$ . Расписание  $R_p$  для всей партии  $p$  получим, умножив вектор элементарного расписания  $r_p$  на число деталей в партии  $x_p$ :  $R_p = r_p x_p$ . Каждая компонента  $R_{qp}, q = \overline{1, Q}$  вектора расписания  $R_p$  равна продолжительности обработки партии  $p$  на группе оборудования  $i(q)$  в сутки  $s(q)$ :  $R_{qp} = r_{qp} x_p$ .

Общая продолжительность использования группы оборудования  $i(q)$  в сутки  $s(q)$  всеми партиями  $p, p = \overline{1, P}$  равна

$$\sum_{p=1}^P R_{qp} = \sum_{p=1}^P r_{qp} x_p.$$

Эта величина не должна превышать фонда времени  $F_q$  группы оборудования  $i(q)$  в сутки  $s(q)$

$$\sum_{p=1}^P r_{qp} x_p \leq F_q, q = \overline{1, Q}. \quad (19.2)$$

Вычислим стоимость  $C_p$  пролеживания одной детали партии  $p$  при ее обработке по элементарному расписанию  $r_p$ . Время пролеживания  $T_{sp}$  одной детали партии  $p$  в сутки  $s$  здесь в этой методике определяется следующим образом. Если детали партии обрабатывались в сутки  $s$ , то считается, что  $T_{sp} = 0$ , если не обрабатывались, то  $T_{sp}$  равно продолжительности суток  $s - \sigma_s$ . Допустим, что полная обработка деталей партии  $p$  по расписанию  $r_p$  начинается в сутки  $s_1$  и заканчивается в сутки  $s_2$ . Тогда время пролеживания  $T_p$  одной детали партии  $p$  по расписанию можно определить как сумму времени пролеживания по всем суткам, на протяжении которых деталь находилась в обработке:

$$T_p = \sum_{s=s_1}^{s_2} T_{sp}.$$

Стоимость пролеживания  $c_p$  одной детали партии  $p$  при ее обработке по расписанию  $r_p$

$$c_p = \gamma^{m(p)} T_p, p = \overline{1, P},$$

а стоимость пролеживания всех деталей партии  $p$  при ее обработке по расписанию  $R_p$  будет равна  $c_p x_p, p = \overline{1, P}$ . Таким образом, стоимость пролеживания всех партий равна

$$C_{pp} = \sum_{p=1}^P c_p x_p. \quad (19.3)$$

В результате приходим к следующей **задаче линейного программирования**: определить целочисленные неизвестные  $P, x_p, p = \overline{1, P}$ , минимизирующие стоимость пролеживания всех партий  $C_{pp}$  (19.3) при ограничениях (19.1) и (19.2). Количество ограничений равно  $M + Q = M + S \cdot I$ .

Для решения задачи на ЭВМ симплекс – методом используют специальные алгоритмы. При этом задачу предварительно записывают в матричной форме (в данной лекции не рассматривается). Элементарные расписания формируются методами динамического программирования. **Динамическое программирование** – методы решения оптимизационных задач, в основе которых лежит идея разбиения исходной задачи на последовательный ряд более простых задач. Основная область приложения динамического программирования – многошаговые процессы, т.е. процессы, протекающие во времени (дискретном или непрерывном).

## Лекция 20. Логические модели процессов и объектов

### Формализация технологического цикла

Полный технологический цикл изготовления готовой штучной продукции всегда представляет собой совокупность отдельных технологических операций, сменяющих друг друга в определенной последовательности. Причинами смены операций могут быть команды человека-оператора или автоматического устройства, выдающего их после получения сигналов от датчиков об окончании предыдущей операции в соответствии с заложенной в него программой.

В то же время очень редко можно обеспечить нормальную работу агрегата, ориентируясь на «жесткую» программу, не способную адаптироваться к неожиданным ситуациям, возникающим в технологическом цикле. Так, если на какой-либо операции становится очевидным появление брака, то оператор или автоматическое устройство следующей командой должны предусмотреть не продолжение обработки, а останов агрегата и уборку бракованной детали. Аналогичная ситуация возникает при поломке оборудования, превышении допустимых значений параметров процесса, несоответствии параметров исходной заготовки техническим условиям [11].

При управлении технологическим циклом необходимо формировать дискретную последовательность команд исполнительным элементам технологического объекта управления (приводам). Формирование команд осуществляется управляющим устройством, называемым дискретным автоматом (рис. 20.1), на основе логического анализа ситуации, о которой сообщают различные датчики положения детали, завершения или качества протекания очередной технологической операции, по командным и оповестительным входам.

Только зная, как и при каких условиях должна формироваться нужная последовательность состояния объекта управления, можно сформулировать задание на синтез управляющего устройства.

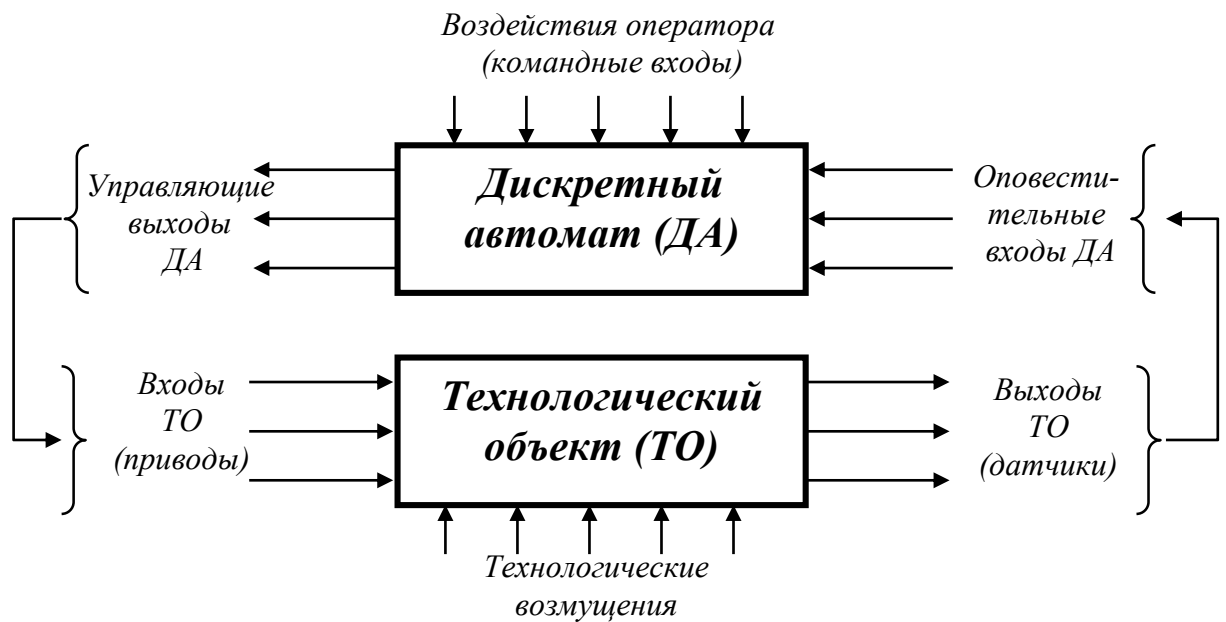


Рис. 20.1. Структура управления технологическим циклом при помощи дискретного автоматического устройства

Существуют различные формы представления моделей дискретных последовательностей операций, то есть моделей технологического цикла. Они могут представляться в виде таблиц, циклограмм, графов, формул и т. д. Предполагая, что все технологические последовательности в конечном счете представляют собой повторяющиеся циклы, следует выделить два существенно отличных вида моделей: **комбинационные и последовательностные**.

Для удобства деления цикла на отдельные элементы вводится понятие технологического такта или состояния, то есть конечного интервала времени, когда объект работает с неизменной

комбинацией включенных / отключенных командных (кнопки, ключи), оповестительных (датчики) и исполнительных (электро-, гидроприводы, электромагниты, муфты) элементов.

Общая последовательность формализации технологического цикла состоит из следующих этапов:

1. Составление содержательного описания, в котором в произвольной повествовательной форме описывается технологический цикл при нормальном его ходе и аварийных ситуациях;
2. Разбиение цикла на такты, характеризуемые неизменным состоянием исполнительных приводов и контролируемых параметров;
3. Анализ переходов от одного такта к другому при нормальных и аварийных ситуациях для выявления причин переходов, то есть выявление изменения состояния командных и исполнительных органов, вызывающих переход;
4. Установление причинно-следственных и логических ситуационных связей между входами и выходами объекта управления, обусловленных требованиями технологии;
5. Составление формализованного графического представления алгоритма функционирования в виде таблицы, циклограммы, графика и т. п.

### Комбинационные модели

В комбинационных моделях ход цикла определяется состоянием входов и выходов объекта управления только в данном такте. При формализации таких моделей часто применяются **таблицы истинности**, отражающие однозначное соответствие дискретных состояний входов и выходов объекта управления.

Активное (включенное) или пассивное (отключенное) состояние исполнительного элемента (входа) или уровень контролируемого выхода (высокий, низкий) может обозначаться любыми символами. Обычно для этих целей используются дискретные величины 1 и 0.

При числе входов  $n$  возможны  $N = 2^n$  сочетаний комбинаций их единичного и нулевого уровней. Поскольку **последовательность смены комбинаций в данном случае роли не играет**, в таблице истинности их удобно располагать в виде кодов натурального ряда двоичных чисел, то есть, чередуя 0 и 1 для первого входа через одно состояние, для второго – через два, для третьего – через три и т. д. Особо следует отметить, что не все комбинации состояний входов (исполнительных приводов) и датчиков реально могут иметь место. Рассмотрим один из возможных примеров.

**Пример.** Произвести сортировку деталей на три группы по размеру, равному  $b$ ,  $2b$  и больше  $3b$ , открыв заслонки бункеров-накопителей, предназначенных для отбора деталей каждого типа. Контроль размеров деталей можно осуществить тремя датчиками  $d_1, d_2, d_3$ , установленными поперек роликового транспортера (рис. 20.2). Обозначив бункеры  $B_1, B_2$  и  $B_3$ , выделим ситуации (см. табл. 20.1):

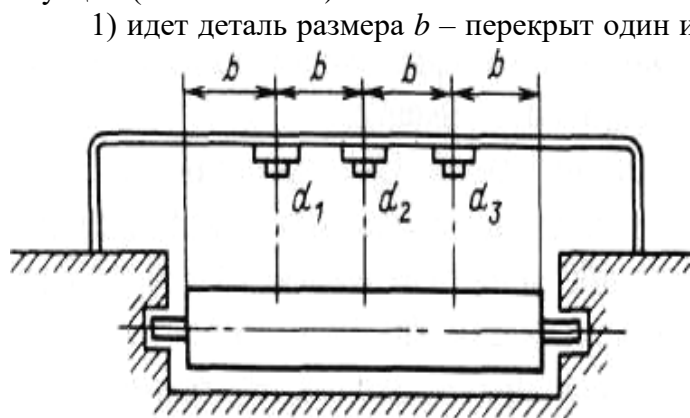


Рис. 20.2. Установка датчиков для контроля размеров детали

закрыт или открыт бункер.

1) идет деталь размера  $b$  – перекрыт один из датчиков ( $d_1$  или  $d_2$ , или  $d_3$ ), открыт бункер  $B_1$  (такты 4, 2, 1);

2) идет деталь  $2b$  – перекрыты два датчика ( $d_1, d_2$  или  $d_2, d_3$ ), открыт бункер  $B_2$  (такты 6, 3);

3) идет деталь размером более  $3b$  – перекрыты все три датчика, открыт бункер  $B_3$  (такт 7).

Таблица истинности составляется для всех возможных комбинаций командных датчиков. Число таких комбинаций равно  $N = 2^n = 8$ , где  $n = 3$  – число командных входов (число датчиков). Состояния датчиков обозначаются двоичным кодом натурального ряда чисел, что позволяет упростить заполнение таблицы. Выходные сигналы 0 или 1 соответственно означают,



Таблица 20.1

Таблица истинности сортировочного автомата

№ комбинаций (тактов)	Состояние					
	Входов			Выходов		
	$d_1$	$d_2$	$d_3$	$B_1$	$B_2$	$B_3$
0	0	0	0	0	0	0
1	0	0	1	1	0	0
2	0	1	0	1	0	0
3	0	1	1	0	1	0
4	1	0	0	1	0	0
5	1	0	1	0	0	0
6	1	1	0	0	1	0
7	1	1	1	0	0	1

Незаполненные клетки состояний выходов (такт 5) соответствуют нереальной ситуации, когда деталь перекрывает датчики  $d_1$  и  $d_3$  и не перекрывает  $d_2$ ; она может возникнуть лишь в результате неисправности датчика  $d_2$ . Эти клетки заполняются нулями для предотвращения аварийной ситуации (все бункеры закрыты, звучит сигнал).

Логическая функция работы бункера  $B_1$  имеет вид:

$$B_1 = \overline{x_1} \overline{x_2} x_3 + \overline{x_1} x_2 \overline{x_3} + \overline{x_1} x_2 x_3 \quad (20.1)$$

### Последовательностные модели

В отличие от комбинационных моделей *при составлении последовательностных моделей необходимо отражать однозначное соответствие состояний выходов комбинациям состояний входов, как в данном такте, так и в предыдущих*. Следовательно, одна и та же комбинация входов в данном такте может вызвать переход в разные новые состояния в зависимости от того, каким было предшествующее состояние. Поэтому в модели должны быть отражены не только данный такт, но и предыстория.

В зависимости от сложности объекта используются различные виды моделей. В простейшем случае применяются **циклограммы**, в которых состояния отражают условным изображением включенного или отключенного исполнительного элемента в виде наличия или отсутствия линии. При большом числе состояний применяются **таблицы состояний** и **графы**.

**Циклограмма** представляет собой ряд горизонтальных строк, равных числу командных и исполнительных элементов. Строки условно разбиты на отрезки, число которых равно числу элементарных технологических тактов. Включенное состояние элемента на строке обозначается сплошной линией, отключенное – отсутствием ее. Вертикальными линиями на циклограммах показана "передача управления" – причинно-следственные связи между командными и исполнительными элементами.

Когда элемент включен, совокупность тактов называется периодом включения, а когда отключен – периодом отключения. Такт, предшествующий периоду включения, называется включающим, а периоду отключения – отключающим.

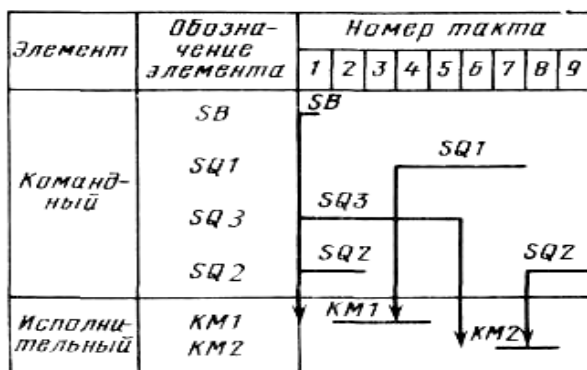


Рис. 20.3. Циклограмма работы грузового подъемника

**Пример.** Рассмотрим циклограмму работы грузового подъемника (рис. 20.3). Грузовой подъемник с тележкой от подачи кратковременной команды кнопкой  $SB$  (пуск) идет вверх [кнопка  $SB$  включает контактор "Вперед"  $KM1$  ( $SB \rightarrow KM1$ ), после чего отключается (такт 1)]. В начале движения отключается нижний конечный выключатель  $SQ2$  (такт 2). После достижения крайнего верхнего положения кабина воздействует на верхний конечный выключатель  $SQ1$ , который дает команду на отключение

$KM1$  (такт 3,  $SQ1 \rightarrow KM1$ ). Контактор  $KM1$  отключается (такт 4). После выката тележки отключается конечный выключатель  $SQ3$  (такт 5) и включается контактор "Назад"  $KM2$ , подъемник идет вниз (такт 6,  $SQ3 \rightarrow KM2$ ), отключается  $SQ1$  (такт 7). После воздействия внизу на нижний конечный выключатель  $SQ2$  отключается  $KM2$  (такт 8,  $SQ2 \rightarrow KM2$ ), кабина останавливается (такт 9).

**Таблица состояний.** Число строк таблицы соответствует числу состояний, число столбцов – числу возможных комбинаций переменных; крайний левый столбец фиксирует номера исходных состояний. В клетках проставляются номера состояний, обусловленных исходным состоянием и возникшей комбинацией управляющих переменных.

Таблица 2.2 Таблица состояний грузоподъемника

Состояние гр. под.	Кнопка ПУСК SB X1	Нижний КВ SQ2 X2	Верхний КВ SQ1 X3	КВ выката SQ3 X4	Контактор вверх KM1 Y1	Контактор вниз KM2 Y2
1.Стоит	0	1	1	1	0	0
2.Вверх	1	1	0	0	1	0
3.Вниз	1	0	1	1	0	1

**Граф-схема.** При представлении цикла в виде графа в вершинах (окружностях) проставляются номера (коды) состояний. Вершины соединяются стрелками, отражающими переходы из одного состояния в другое. Над стрелками записываются комбинации переменных, обуславливающих этот переход.

**Пример.** Подъемник перемещается с одного уровня на другой реверсивным приводом, включаемым исполнительными элементами контакторами  $KM1$  и  $KM2$ . Пуск подъемника осуществляется по команде от этажных кнопок  $SB1, SB2, SB3, SB4$ . Аварийные ситуации предотвращаются реле перегрузки, конечными выключателями  $SQ1, SQ2$  и контролем закрытия дверей шахты  $SQ3, SQ4$ . Возможно состояний три: 1 – движение вверх, 2 – движение вниз, 3 – кабина неподвижна.

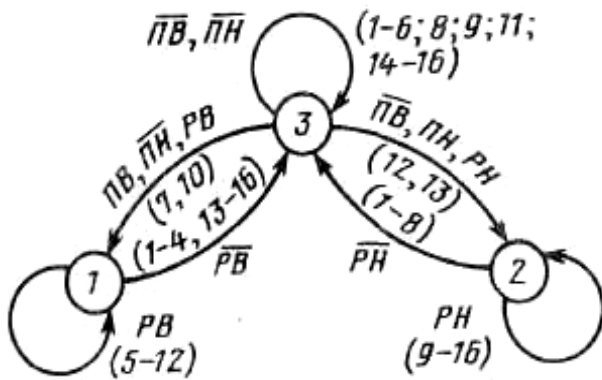


Рис. 20.4. Граф-схема алгоритма функционирования подъемника

Введем в рассмотрение следующие **командные переменные**: пуск вверх  $ПВ$  – нажаты кнопки "Вверх"  $SB1, SB3$  на первом или втором этаже; пуск вниз  $ПН$  – нажаты кнопки "Вниз"  $SB2, SB4$  на первом или втором этаже; есть разрешение двигаться вверх  $РВ$  – закрыты все двери, не нажат конечный выключатель  $SQ1$ , нет перегрузки; есть разрешение двигаться вниз  $РН$  – закрыты все двери, не нажат конечный выключатель  $SQ2$ , нет перегрузки.

На рисунке 20.4 представлен граф, описывающий работу подъемника. Граф имеет три вершины (состояния 1, 2, 3). Пуск из состояния 3 (подъемник неподвижен) в состояние 1 или 2 (движение вверх или вниз) обусловлен наличием требуемой команды ( $ПВ$  или  $ПН$ ), отсутствием противоположной ( $ПН$  или  $ПВ$ ) и наличием разрешений ( $РВ$  или  $РН$ ). Обратный переход обусловлен только отсутствием разрешения ( $РВ$  или  $РН$ ), то есть снятие  $ПВ$  или  $ПН$  остановка не вызывает. Сохранение состояния 1 или 2 обусловлено только наличием  $РВ$  или  $РН$ , а состояние 3 – их отсутствием. В скобках над или под стрелками указаны номера комбинаций переменных (такты), соответствующие данному переходу.

Представление тем или иным образом алгоритма функционирования технического объекта зависит от степени их освоения и приобретенных навыков.

## Синтез дискретных автоматов

### Основные положения двоичной алгебры

Отношения между двумя элементами определяются отношением эквивалентности, обозначаемым знаком равенства, и операциями: сложения (дизъюнкции), обозначаемой знаками  $\langle + \rangle$  или  $\langle \vee \rangle$ , умножения (конъюнкции) –  $\langle \cdot \rangle$  или  $\langle \wedge \rangle$  и отрицания (инверсии) –  $\langle \text{НЕ} \rangle$  или  $\langle \bar{\phantom{x}} \rangle$ . При выполнении всех трех операций отношения эквивалентности определяются следующими выражениями:

$$\begin{aligned} 0 + 0 = 0; & \quad 0 \cdot 0 = 0; & \quad \bar{0} = 1; \\ 1 + 1 = 1; & \quad 1 \cdot 0 = 0; & \quad \bar{1} = 0; \\ 0 + 1 = 1; & \quad 0 \cdot 1 = 0; & \\ 1 + 1 = 1; & \quad 1 \cdot 1 = 1; & \end{aligned} \quad (20.2)$$

Эти постулаты позволяют обосновать законы двоичной алгебры. Законы одинарных элементов:

– универсального множества:  $x + 1 = 1; \quad x \cdot 1 = x;$  (2.3)

– нулевого множества:  $x + 0 = x; \quad x \cdot 0 = 0.$  (2.4)

Законы отрицания (теорема Моргана):

– двойного отрицания:  $\overline{\overline{x}} = x;$  (2.5)

– дополненности:  $x + \bar{x} = 1; \quad x \cdot \bar{x} = 0;$  (2.6)

– двойственности:  $\overline{x_1 + x_2} = \bar{x}_1 \cdot \bar{x}_2; \quad \overline{x_1 \cdot x_2} = \bar{x}_1 + \bar{x}_2.$  (2.7)

Законы комбинационные:

– тавтологии:  $x + x = x; \quad x \cdot x = x;$  (2.8)

– коммутативные или переместительные:  $x_1 + x_2 = x_2 + x_1 \quad x_1 \cdot x_2 = x_2 \cdot x_1$  (2.9)

– ассоциативные или согласовательные:

$(x_1 + x_2) + x_3 = x_1 + (x_2 + x_3); \quad (x_1 \cdot x_2) \cdot x_3 = x_1 \cdot (x_2 \cdot x_3);$  (2.10)

– дистрибутивные или распределительные:

$x_1 \cdot (x_2 + x_3) = x_1 \cdot x_2 + x_1 \cdot x_3; \quad x_1 + (x_2 \cdot x_3) = (x_1 + x_2) \cdot (x_1 + x_3);$  (2.11)

– абсорбции или поглощения:  $x_1 + x_1 \cdot x_2 = x_1; \quad x_1 \cdot (x_1 + x_2) = x_1;$

– склеивания:  $x_1 \cdot x_2 + x_1 \cdot \bar{x}_2 = x_1.$

### Синтез комбинационных автоматов

Для создания сложных комбинационных автоматов используют элементарные комбинационные автоматы (логические элементы), которыми реализуется функционально полный набор двоичных функций. Элементарным комбинационным автоматом будем называть автомат, реализующий логическую функцию двух переменных.

Синтез комбинационных автоматов начинается с содержательного описания функционирования объекта и составления таблицы истинности.

Для получения аналитического выражения, определяющего необходимую структуру автомата, записываются исходные формы переключательной функции в виде совершенной дизъюнктивной нормальной формы (СДНФ) или совершенной конъюнктивной нормальной формы (СКНФ).

Дизъюнктивная нормальная форма представляет собой дизъюнкцию (сумму) минтермов, а конъюнктивная нормальная форма – конъюнкцию (произведение) макстермов.

**Минтермом** или конституэнтной единицы называется логическая функция, принимающая значение 1 только на одном наборе переменных. Образуется как конъюнкция всех входных переменных с отрицанием тех, которые в данном наборе равны нулю. Число минтермов равно числу наборов.

**Макстермом** или конституэнтной нуля называется логическая функция, принимающая значение 0 только на одном наборе переменных. Образуется как дизъюнкция входных переменных, где переменные данного набора, равные 1, взяты с отрицанием. Число макстермов равно числу наборов.

Две эти формы эквивалентны. При минимизации удобно пользоваться СДНФ.

Рассмотрим методы синтеза автоматов с минимальным количеством элементов, которые называются минимальными или оптимальными.

**Пример.** Разработать автомат, реагирующий не менее чем на два сигнала из трех (мажоритарный автомат).

Таблица 20.3 Таблица истинности автомата

$x_1$	0	1	0	1	0	1	0	1
$x_2$	0	0	1	1	0	0	1	1
$x_3$	0	0	0	0	1	1	1	1
$y$	0	0	0	1	0	1	1	1

1. По таблице истинности записывается логическая функция в форме СДНФ. Число слагаемых равно числу наборов, где синтезируемая функция равна 1, а число сомножителей в каждом слагаемом – числу аргументов. Над аргументами, принимающими в данном наборе значение 0, ставится знак отрицания:

$$y = x_1 x_2 \bar{x}_3 + \bar{x}_1 x_2 x_3 + x_1 \bar{x}_2 x_3 + x_1 x_2 x_3 \quad (20.12)$$

2. Минимизация СДНФ. Производится наиболее популярным (табличным) методом Куайн-Мак Ласки (существуют также методы Карно, Квайна [12]).

2.1. Анализируем таблицу истинности:

Таблица 20.4

№ строки	$x_1$	$x_2$	$x_3$	$y$	Минтерм	Кол-во единиц
1	0	0	0	0		
2	1	0	0	0		
3	0	1	0	0		
4	1	1	0	1	$x_1 x_2 \bar{x}_3$	2
5	0	0	1	0		
6	1	0	1	1	$x_1 \bar{x}_2 x_3$	2
7	0	1	1	1	$\bar{x}_1 x_2 x_3$	2
8	1	1	1	1	$x_1 x_2 x_3$	3

2.2. Группируем минтермы по количеству в них единиц:

Таблица 2.5

№ группы	№ строки	$x_1$	$x_2$	$x_3$	Кол-во единиц
1	4	1	1	0	2
1	6	1	0	1	2
1	7	0	1	1	2
2	8	1	1	1	3

2.3. Произведем объединение строк каждой предыдущих и последующих групп. Объединяются строки, в которых какой-либо X различается значением (0 или 1) – *смежные минтермы*. Такой X обозначаем знаком «+»:

Таблица 2.6

№№ строк	$x_1$	$x_2$	$x_3$	Кол-во единиц
4, 8	1	1	+	2
6, 8	1	+	1	2
7, 8	+	1	1	2

2.4. Производится объединение строк с совпадающими позициями «+». В нашем примере таких нет.

2.5. Из двух строк с совпадающими позициями «+» оставляем только одну.

Приведенный выше алгоритм минимизации СДНФ позволяет получить следующую логическую функцию автомата:

$$y = x_1x_2 + x_1x_3 + x_2x_3 \quad (20.13)$$

3. Двукратное инвертирование СДНФ (переход в базис Шеффера):

$$y = \overline{\overline{x_1x_2 + x_1x_3 + x_2x_3}} = \overline{\overline{x_1x_2} \cdot \overline{x_1x_3} \cdot \overline{x_2x_3}} \quad (20.14)$$

4. Построение логической схемы автомата:

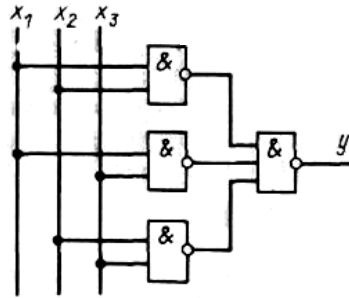


Рисунок 20.5. Логическая схема автомата

### Синтез последовательностных автоматов

Последовательностные автоматы осуществляют преобразование двоичных входных переменных не только с учетом текущей комбинации на входе в данном такте автоматного времени, но и с учетом того, что было в предыдущих тактах.

Создание последовательностного автомата, способного запоминать предшествующие данному такту комбинации сигналов на входе, обеспечивается наличием в комбинационном автомате не только внешних (рабочих) входов  $X$  и выходов  $Y$ , но и вспомогательных (внутренних) переменных  $Z$ , которые должны иметь возможность давать различные комбинации для каждого такта, подлежащего запоминанию, и реализуемых в виде обратных связей. Эти внутренние переменные, подаваемые на вход комбинационной схемы, как бы корректируют результат от воздействия внешних входов, учитывая предыдущие ситуации.

Таким образом, задача внешних входов – задать текущую комбинацию, а внутренних – сохранить и задать на входе комбинацию, однозначно соответствующую сформировавшейся на выходе в предыдущем такте (кодирующую ее). В этом случае выходная комбинация формируется с учетом предыдущего такта. В следующем такте внутренние переменные внесут очередную коррективу (если их комбинация будет отлична от предыдущей) и, следовательно, новые значения выходов уже несут в себе следы двух предыдущих тактов и т.д. Количество таких состояний  $M$  внутренних переменных  $Z$  называется **весом** последовательностного автомата.

Число внутренних переменных  $m_z$ , обеспечивающих возможность кодирования всех  $M$  состояний, называется объемом памяти автомата и равно:

$$m_z = \log_2 M. \quad (20.15)$$

Внешними входными переменными автомата определяется количество его возможных входных комбинаций, а внутренними – через сколько комбинаций внешних начинает повторяться коррекция внутренними. Так, если  $m_x = 3$  и  $m_z = 2$ , то автомат будет иметь восемь состояний с повторяющейся коррекцией через четыре такта.

Следствием этого является характерная особенность последовательностных автоматов: у них вследствие коррекции  $Z$  возможны разные комбинации выходов при одинаковых комбинациях на внешних входах. Технических вариантов реализации даже простых последовательностных автоматов может быть очень много. Достаточно наглядной является схема последовательностных автоматов, приведенная на рисунке 2.6 и представляющая собой наиболее общий случай.

Входы  $X^t$  логикой комбинационной схемы формируют выходы  $Y^t$  в данном такте. В формировании  $Y^t$  при этом участвуют внутренние переменные  $Z^{t-1}$ , комбинация которых сформировалась в предыдущем такте  $t-1$ . Переменные  $Z^{t-1}$  сформировались узлом управления обратными связями (ОС) в цепи обратной связи, который может включать: комбинационную логику, вре-

менные задержки и синхронизирующий вход. Как видно из схемы, именно ОС обеспечивает «запоминание» путем приема на вход выходной комбинации автомата в данном такте и передачу ее с временной задержкой на входы автомата в следующем такте.

Разрыв цепи ОС превращает последовательностные автоматы в комбинационные. Важную роль играет характер задержки: если блок ОС имеет синхронизирующий вход, «отпирающий» и «запирающий» цепь обратной связи в каждом такте, – это синхронный автомат.

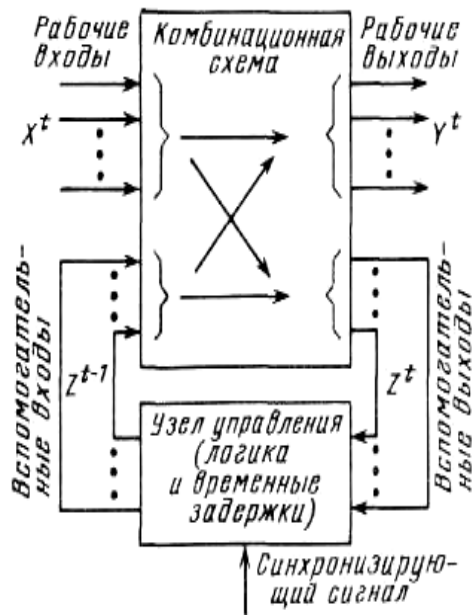


Рис. 20.6. Схема последовательностного автомата

У него равенство  $Z^{t-1} = Z^t$  наступает только после «отпирающего» синхроимпульса цепи обратной связи.

Благодаря обратной связи последовательностные автоматы обладают специфическим свойством, характерным для замкнутых систем управления – они могут быть устойчивыми и неустойчивыми. Устойчивым будем называть состояние, когда комбинация выходов меняется лишь вследствие изменения входа. Неустойчивые автоматы могут после установления комбинации входов несколько раз менять комбинацию на выходе, проходя ряд неустойчивых состояний. Такой процесс может завершиться переходом в устойчивое состояние (затухание колебаний) либо продолжаться неограниченное время (автоколебания).

Рассмотрим этапы более общего синтеза последовательностного автомата, у которого переход из одного состояния в другое зависит не только от прихода очередного тактового импульса, но и от условий, сформированных «внешней» логикой – различными датчиками и внешними управляющими командами.

**Этап 1.** Составим описание цикла функционирования механизма или агрегата, для автоматизации которого разрабатывается автомат. Установим количество состояний (тактов), когда не меняются состояние ни одного дискретного исполнительного элемента, составляющее цикл работы механизма, и причины перехода из одного состояния в другое.

В зависимости от причин может меняться последовательность смены состояний, но не их номенклатура. Под причинами понимается срабатывание какого-либо датчика или командного органа либо изменение какого-либо сочетания их состояний. Например, в описании цикла установлено, что цикл состоит из пяти тактов (состояний  $a_1, \dots, a_5$ ), переходы которых обусловлены различными сочетаниями команд оператора  $x_1 - x_3$  и сигналов датчиков  $z_1 - z_4$  (см. рис. 20.7, а).

**Этап 2.** Составим формальный портрет, отражающий все реально возможные переходы из одного состояния в другое и условия этих переходов. Он может быть составлен в виде тактовой таблицы, структурной схемы алгоритма или в виде графа, вершинами которого являются состояния, а ребра отражают причины перехода из одного состояния в другое. Над ребрами записаны логические условия перехода. На рисунке 2.7, б показан возможный вариант такого графа.

**Этап 3.** Произведем кодирование состояний. Чтобы автомат «знал» какое в данном такте состояние, он должен иметь узел, на выходах которого появляется 1 в зависимости от состояния, а именно: на первом – при  $a_1$ , на втором – при  $a_2$  и т.д. Таким узлом может быть дешифратор с тремя триггерами на входе. При кодировании состояний каждому состоянию ставится в соответствие комбинация состояний выходов триггеров. Чтобы закодировать пять состояний, нужны три триггера с выходами  $Q_0, Q_1, Q_2$ :

Не используются еще три возможных состояния 3-х триггеров ( $2^3 = 8$ )

**Этап 4.** Составим таблицу переходов (см. табл. 2.8), в которой на основании графа отразим переход каждого выхода каждого триггера при всех возможных переходах автомата из одного состояния в другое. В таблице укажем логические условия каждого перехода и состояние входов  $JK$  каждого триггера при известном переходе его выхода, характерные только для выбранного  $JK$ -триггера.

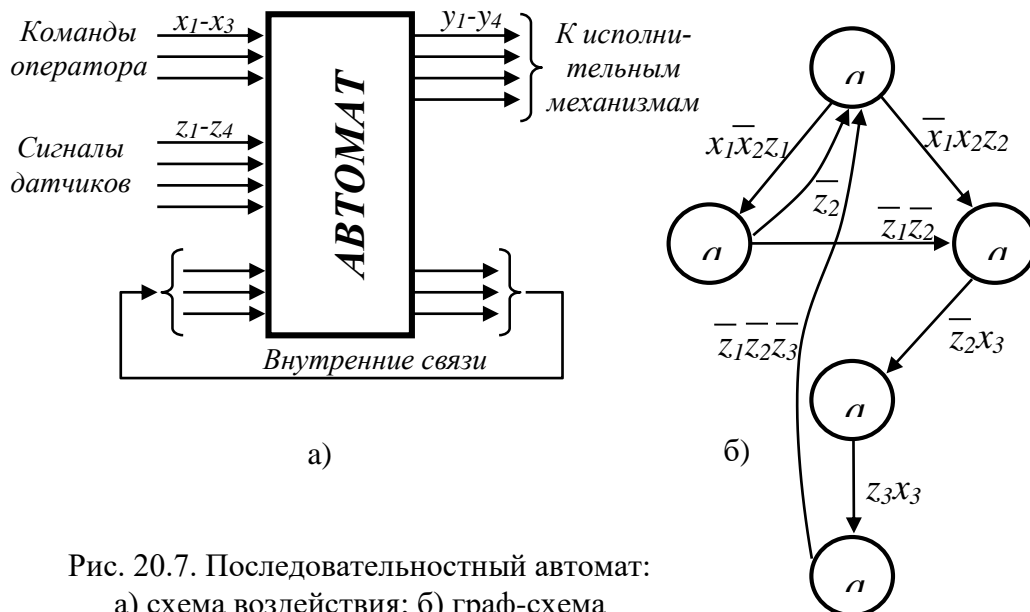


Рис. 20.7. Последовательный автомат:  
а) схема воздействия; б) граф-схема

Таблица 20.7

Со- стояние	$Q_0$	$Q_1$	$Q_2$
$a1$	0	0	1
$a2$	0	1	0
$a3$	0	1	1
$a4$	1	0	0
$a5$	1	0	1

Таблица 20.8

Исходное состояние	Код исходного состояния	Новое состояние	Код нового состояния	Условия перехода	Состояние входов триггеров					
					$J_0$	$K_0$	$J_1$	$K_1$	$J_2$	$K_2$
$a1$	001	$a2$	010	$X_1X_2'Z_1$	~	1	1	~	0	~
		$a3$	011	$X_1'X_2Z_2$	~	0	1	~	0	~
$a2$	010	$a1$	011	$Z_1'$	1	~	~	1	0	~
		$a3$	011	$Z_1'Z_2'$	1	~	~	0	0	~
$a3$	011	$a4$	100	$Z_2'X_3$	~	1	~	1	1	~
$a4$	100	$a5$	101	$Z_3X_3$	1	~	0	~	~	0
$a5$	101	$a1$	001	$Z_1'Z_2'Z_3$	~	0	0	~	~	1

Работа JK-триггера определяется следующей таблицей состояний:

Таблица 2.9

J	K	$Q_{n+1}$	$/Q_{n+1}$
0	0	$Q_n$	$/Q_n$
1	0	1	0
0	1	0	1
1	1	$/Q_n$	$Q_n$

**Этап 5.** По таблице перехода (табл. 2.8) запишем логические уравнения, связывающие входы триггеров с командными входами  $x$  сигналами датчиков  $z$  и выходами триггеров  $Q$  через дешифратор состояний. Запись уравнений производится следующим образом: для каждого входа каждого триггера для переходов, где данный вход равен 1, записывается произведение исходного состояния на логическое условие его перехода в последующее.

Уравнения имеют вид:

$$\begin{aligned}
 J_0 &= a_2(\bar{z}_1 + \bar{z}_1\bar{z}_2) + a_4z_3x_3 = a_2\bar{z}_2 + a_4z_3x_3; \\
 K_0 &= a_1x_1\bar{x}_2z_1 + a_3\bar{z}_2x_3; \\
 J_1 &= a_1z_1\bar{x}_2z_1 + a_1\bar{x}_1x_2z_2; \\
 K_1 &= a_2\bar{z}_1 + a_3\bar{z}_2x_3; \\
 J_2 &= a_3\bar{z}_2x_3; \\
 K_2 &= a_5\bar{z}_1\bar{z}_2\bar{z}_3
 \end{aligned}
 \tag{20.16}$$

Полученный автомат (рис. 20.8) может обеспечить пятитактный цикл управления какими-либо исполнительными элементами. Их не может быть менее трех (чтобы получилось пять разных состояний), но больше может быть (10, 20 и т. п.), то есть каждое состояние будет отличаться комбинацией включенных и отключенных элементов (станков, механизмов).

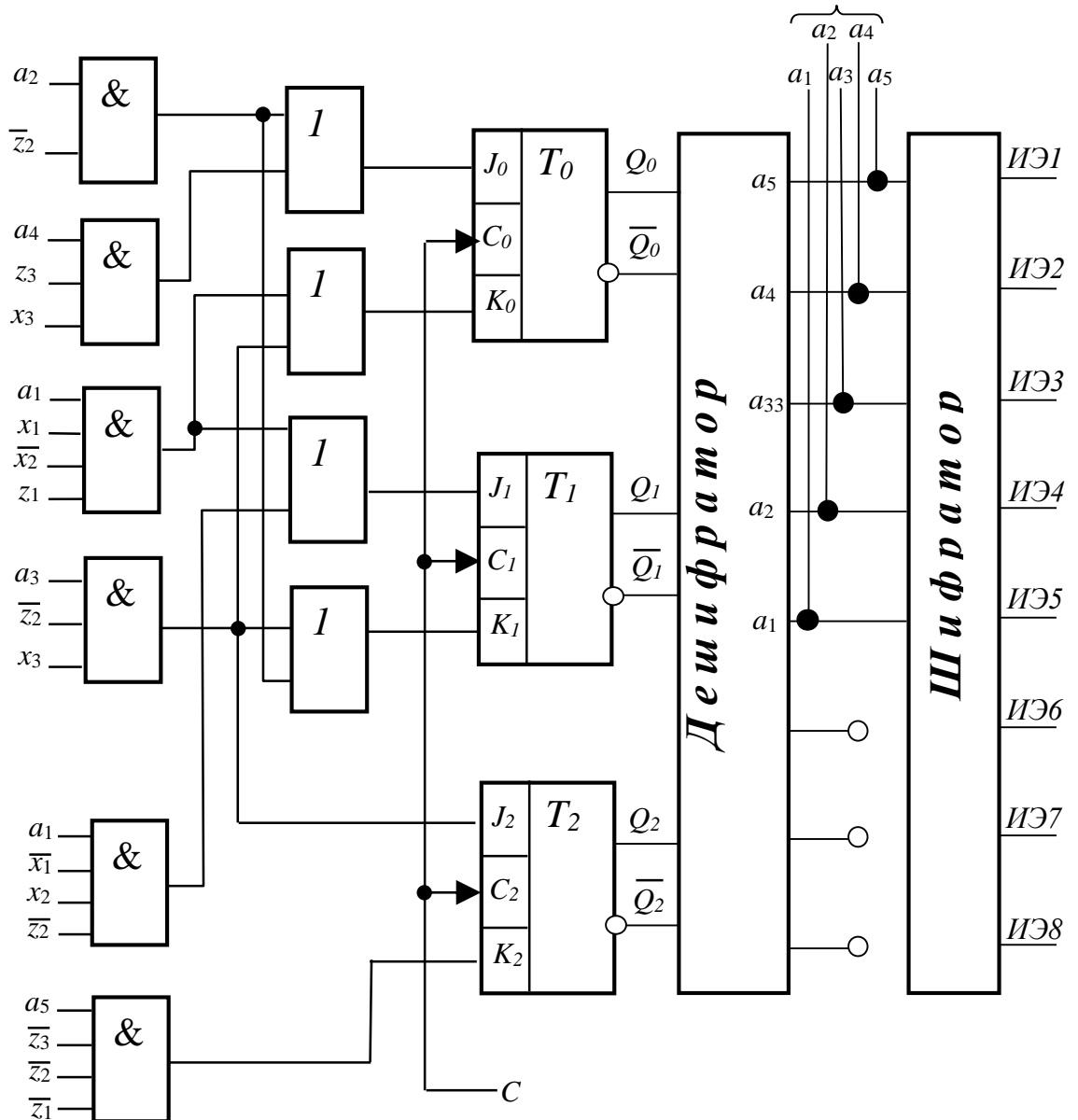


Рис. 2.8. Электрическая схема автомата



## 2.3. Пример логической модели работы ГАУ

### 2.3.1. Общее описание ГАУ

Рассмотрим пример логического моделирования системы управления ГАУ для изготовления деталей типа «Корпус». Планировка участка представлена на рисунке 2.9. Ведомость основного и вспомогательного оборудования приведена в таблице 2.10.

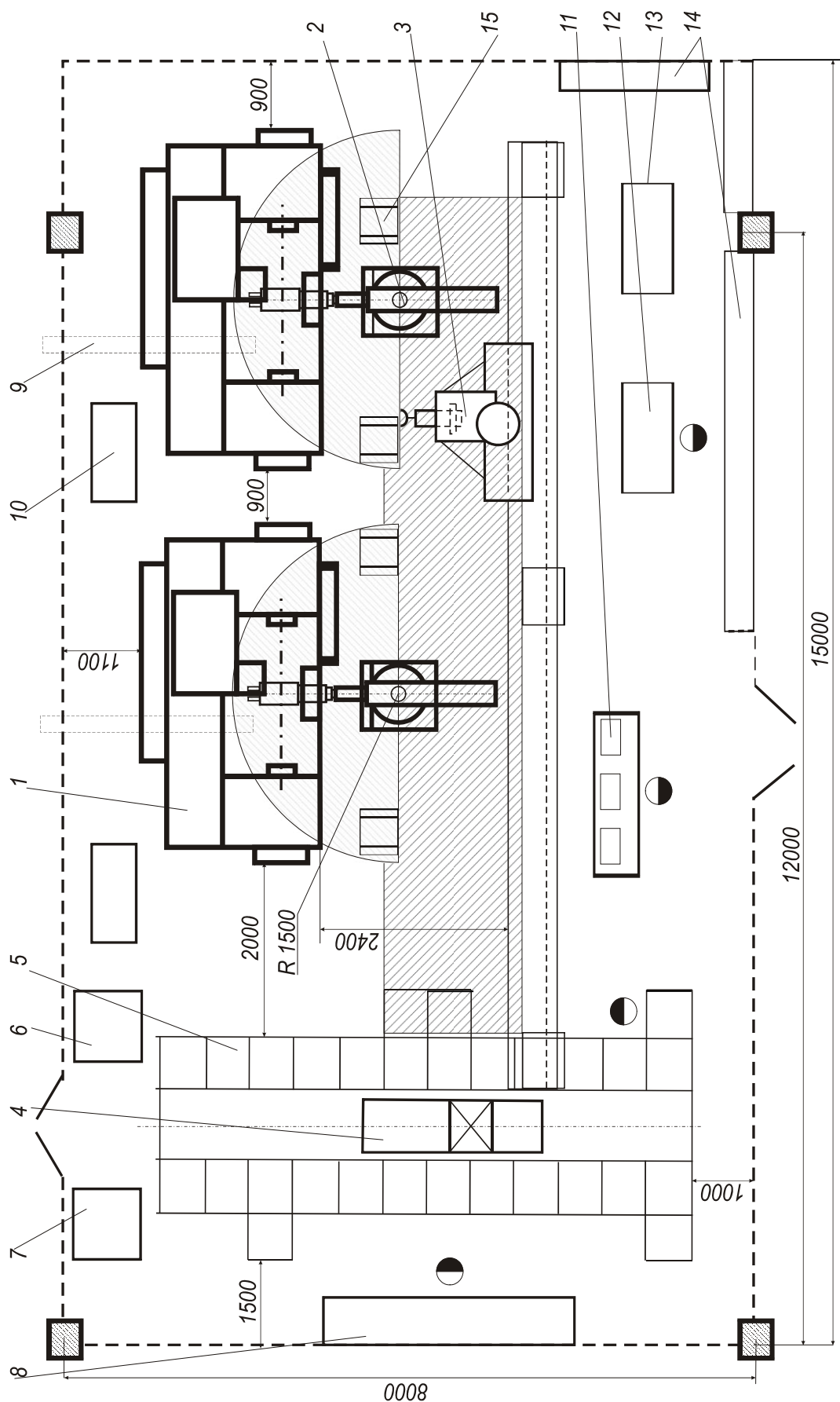


Рис. 2.9. Планировка гибкого автоматизированного участка

Таблица 2.10

Ведомость оборудования производственного участка

№ п.п.	Наименование	Модель	Количество
1	Токарный обрабатывающий центр	INTEGREX 200 III ST	2
2	Робот напольный	M10П.62.01	2
3	Робот порталный	СМ40Ф2.80.01	1
4	Робот-штабелер	РШ-500	1
5	Автоматизированный склад	АС-50	1
6	Склад готовых деталей		1
7	Склад для заготовок		1
8	Стол сборки-разборки тары		1
9	Конвейер для уборки стружки		2
10	Шкаф электроавтоматики		2
11	Пульт управления ГАУ		1
12	Контрольный стол		1
13	Место отдыха		1
14	Щкаф-стеллаж		3
15	Позиция загрузки-выгрузки		4

Заготовки в тарах (см. рис. 2.9), собираемых в поз. 9, поступают на хранение на автоматизированный склад 4 и по требованию выдаются на позицию выдачи склада. Портальный робот 3 забирает тару со склада и загружает одну из двух позиций загрузки многоцелевых станков. Два напольных робота 2, каждый из которых обслуживает загрузку-разгрузку двух многоцелевых токарных станков 1, поштучно забирают заготовки с приемных позиций станков и, по окончании обработки, укладывают готовые детали в тару на разгрузочных позициях 15. При заполнении тары готовыми изделиями порталный робот должен забрать тару с разгрузочных позиций и передать ее на автоматизированный склад.

### 2.3.2. Проектирование материальных потоков ГАУ

В составе ГАУ функционируют шесть исполнительных элементов, выполняющих задачи перемещения и формообразования заготовок и деталей большой номенклатуры. В таблице 20.11 закодированы и описаны локальные задачи каждого элемента, которые задаются в виде управляющих программ при наладке оборудования.

Таблица 20.11 Перечень состояний исполнительных элементов ГАУ

№ п.п.	Код состояния	Обозначение элемента	Описание задачи состояния элемента
<b>Элемент 1 Автоматизированный склад под управлением робота-штабелера</b>			
1.	001000	СРШ1	Бездействие робота-штабелера
2.	001001	СРШ2	Выдать тару заданного кода на выдающую позицию АС
3.	001002	СРШ3	Принять тару заданного кода с приемной позиции АС
4.	001003	СРШ4	Загрузить позицию комплектования тарой заданного кода
5.	001004	СРШ5	Разгрузить позицию комплектования тары заданного кода
6.	001005	СРШ6	Выдать тару деталей со склада
7.	001006	СРШ7	Принять тару заготовок на склад
<b>Элемент 2 Портальный промышленный робот</b>			
8.	002001	СПР1	Бездействие порталного робота
9.	002002	СПР2	Забрать тару с выдающей позиции АС
10.	002003	СПР3	Положить тару на приемную позицию АС
11.	002004	СПР4	Загрузить приемную позицию МС1
12.	002005	СПР5	Загрузить приемную позицию МС2

№ п.п	Код состояния	Обозначение элемента	Описание задачи состояния элемента
13.	002006	<i>СПР6</i>	Разгрузить приемную позицию МС1
14.	002007	<i>СПР7</i>	Разгрузить приемную позицию МС2
15.	002008	<i>СПР8</i>	Загрузить выдающую позицию МС1
16.	002009	<i>СПР9</i>	Загрузить выдающую позицию МС2
17.	002010	<i>СПР10</i>	Разгрузить выдающую позицию МС1
18.	002011	<i>СПР11</i>	Разгрузить выдающую позицию МС2
<b>Элемент 3 Напольный робот станка 1</b>			
19	003001	<i>СНП11</i>	Бездействие робота
20.	003002	<i>СНП12</i>	Загрузить заготовку в левый шпиндель МС1
21.	003003	<i>СНП13</i>	Забрать деталь из правого шпинделя МС1
22.	003004	<i>СНП14</i>	Смена схвата под заданный код детали
<b>Элемент 4 Напольный робот станка 2</b>			
23	004001	<i>СНП21</i>	Бездействие робота
24.	004002	<i>СНП22</i>	Загрузить заготовку в левый шпиндель МС2
25.	004003	<i>СНП23</i>	Забрать деталь из правого шпинделя МС2
26.	004004	<i>СНП24</i>	Смена схвата под заданный код детали
<b>Элемент 5 Токарный обрабатывающий центр 1 (МС1)</b>			
27.	005001	<i>СС11</i>	Бездействие МС1
28.	005002	<i>СС12</i>	Обработка детали заданного кода
29.	005003	<i>СС13</i>	Смена управляющей программы
30.	005004	<i>СС14</i>	Ожидание подналадки техоснастки
31.	005005	<i>СС15</i>	Ожидание разгрузки готовой детали
<b>Элемент 6 Токарный обрабатывающий центр 2 (МС2)</b>			
32.	005001	<i>СС21</i>	Бездействие МС2
33.	005002	<i>СС22</i>	Обработка детали заданного кода
34.	005003	<i>СС23</i>	Смена управляющей программы
35.	005004	<i>СС24</i>	Ожидание подналадки техоснастки
36.	005005	<i>СС25</i>	Ожидание разгрузки готовой детали

### Проектирование информационных потоков ГАУ

Для управления технологическим оборудованием в составе ГАУ необходим комплекс источников информации о реальном состоянии каждого исполнительного элемента, а также общее управляющее устройство, которое принимает решение о запуске того или иного состояния каждого исполнительного элемента. Функции источников информации о состоянии исполнительных элементов выполняют датчики. На участке применяются следующие датчики.

**Контактные датчики положения** (рис. 2.10) отличаются механическим воздействием переключающего упора на чувствительный элемент. Обычно это

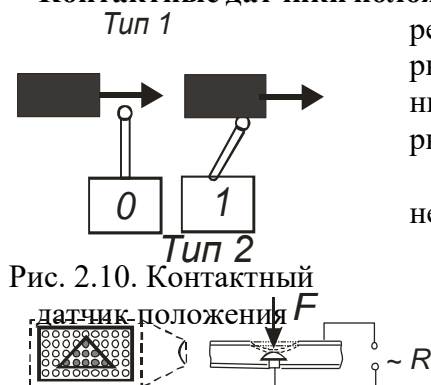


Рис. 2.10. Контактный датчик положения

рычажные выключатели, простые и дешевые, но подверженные механическому износу и не способные работать при быстрых перемещениях объекта.

**Тактильный датчик** (рис. 2.11) контактного давления, не имеющий щупа. Матрица датчиков контактного давления размещается на плоской поверхности и служит для распознавания формы касающегося ее объекта. Датчик контактного давления может быть выполнен в виде матрицы металлических контактов, над которой через эластичный материал с отверстиями уложена фольга. При

Рис. 2.11. Тактильный датчик

нажатии на фольгу контакт замыкается. Объект распознается по расположению замкнутых контактов в матрице.

**Волоконно-оптические датчики** положения (рис. 2.12) реагируют на пересечение или отражение светового потока, передаваемого через оптическое волокно. Современные датчики имеют встроенную регулировку зоны чувствительности. Начало и конец зоны чувствительности устанавливают нажатием кнопки на корпусе датчика при обучении срабатыванию датчика.

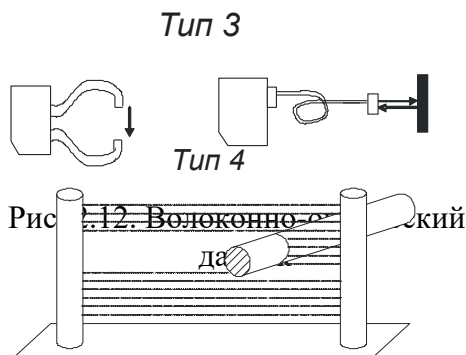


Рис. 2.13. Датчик типа «Световой барьер»  
Тип 5

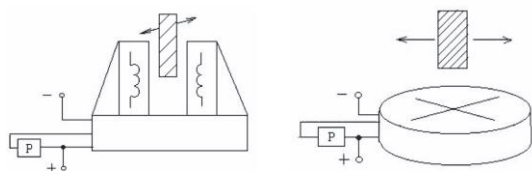


Рис. 2.14. Генераторный датчик положения

автогенератор, детектор, пороговый элемент и выходной усилитель-формирователь. Чувствительный элемент представляет собой катушку индуктивности. Прохождение металлического объекта на расстоянии 1 – 3 мм от торца датчика приводит к изменению индуктивности, срыву генерации и включению реле между проводом питания и третьим проводом датчика. Дифференциал хода составляет 0,15 – 0,60 мм, погрешность положения точки переключения 0,05 – 0,10 мм, частота переключения до 800 Гц, ток нагрузки 200 мА.

Схема расположения датчиков на рассматриваемом участке изготовления деталей типа «Корпус» приведена на рисунке 2.15. Ведомость информационных датчиков ГАУ приведена в таблице 2.12.

**Датчик «Световой барьер»** (рис. 2.13) представляет собой стойки фотоизлучателей и фотоприемников, направленных друг к другу. Он предназначен для бесконтактного ограждения опасных зон. Каждый фотоприемник принимает узконаправленный оптический луч от противоположного фотоизлучателя. При пересечении луча объектом ограждаемый технологический процесс останавливается и подается сигнал тревоги.

**Генераторные датчики** (рис. 2.14) подразделяют на щелевые, плоскостные и торцевые. В щелевом датчике переключающий упор проходит через щель шириной 3 – 100 мм. Дифференциал хода составляет 1,5 – 2,0 мм при ширине щели до 6 и 5 – 15 мм при ширине щели 20 – 100 мм. В плоскостном датчике переключающий упор перемещается на расстоянии 25 – 16 мм от плоской поверхности датчика.

В современной технике автоматизации наиболее распространены генераторные датчики положения торцевого типа. За рубежом их называют индуктивными. Датчик положения подвижных звеньев механизмов выполнен в виде неразборного болта диаметром 8 – 12 мм и длиной 50 – 60 мм. Внутри болта размещены

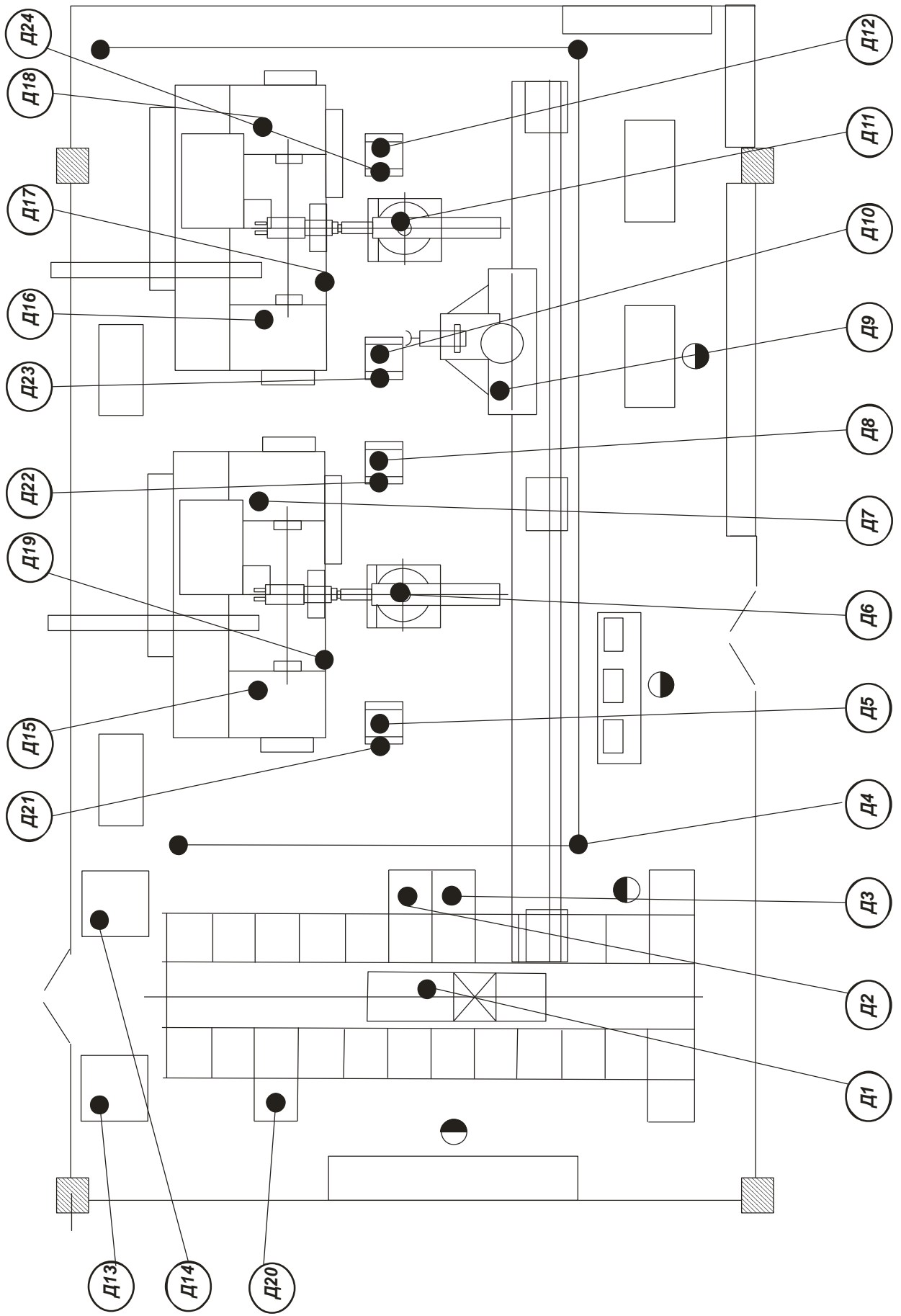


Рис. 2.15. Схема расположения информационных датчиков

Таблица 2.12 Ведомость информационных датчиков ГАУ

№	Тип	Назначение датчика
Д1	Тип 5	Определяет положение робота-штабелера
Д2	Тип 2	Фиксирует наличие тары с определенным типом заготовок
Д3	Тип 2	Фиксирует наличие тары с определенным типом деталей
Д4	Тип 4	Световое ограждение рабочей зоны ГАУ
Д5	Тип 2	Фиксирует наличие тары с определенным типом деталей
Д6	Тип 1	Определяет позицию напольного робота в состоянии ожидания
Д7	Тип 3	Определяет наличие заготовки в правом шпинделе МС1
Д8	Тип 2	Фиксирует наличие тары с определенным типом заготовок
Д9	Тип 1	Определяет позицию портального робота в состоянии ожидания
Д10	Тип 2	Фиксирует наличие тары с определенным типом деталей
Д11	Тип 1	Определяет позицию напольного робота в состоянии ожидания
Д12	Тип 2	Фиксирует наличие тары с определенным типом деталей
Д13	Тип 2	Фиксирует наличие тары с определенным типом деталей
Д14	Тип 2	Фиксирует наличие тары с определенным типом заготовок
Д15	Тип 3	Определяет наличие заготовки в левом шпинделе МС1
Д16	Тип 3	Определяет наличие заготовки в левом шпинделе МС2
Д17	Тип 3	Определяет защиту рабочей зоны МС2
Д18	Тип 3	Определяет наличие заготовки в правом шпинделе МС2
Д19	Тип 3	Определяет защиту рабочей зоны МС2
Д20	Тип 1	Определяет наличие тары в позиции комплектования АС
Д21	Тип 3	Определяет, что тара на загрузке МС1 пустая
Д22	Тип 3	Определяет, что тара на выгрузке МС1 станка наполнилась
Д23	Тип 3	Определяет, что тара на загрузке МС2 станка пустая
Д24	Тип 3	Определяет, что тара на выгрузке МС2 станка наполнилась

#### 2.3.4. Определение вспомогательных функций системы управления

Логические функции представляют собой конъюнкцию командных, внутренних переменных системы. Также в логической функции состояния одного исполнительного элемента могут присутствовать логические функции состояний других исполнительных элементов ГАУ. Все логические функции системы основаны на булевых (логических, бинарных) переменных, используемых в дискретной математике.

Разберем подробно процесс формирования некоторых логических функций исполнительных элементов ГАУ. При этом используются обозначения, приведенные в п. п. 2.3.1, 2.3.2.

##### **Функция загрузки заготовки в станок 1 напольным роботом**

Пусть напольный робот у станка 1 находится на текущем такте в состоянии бездействия (состояние *СНП11*) (см. табл. 2.10 – 2.12). Его переход в состояние загрузки станка (*СНП12*) может быть произведен в случае, если:

- на приемной позиции станка 1 уложена тара с заготовками ( $D5=1$ );
- станок 1 бездействует (состояние  $CC11=1$ );
- диспетчером установлено рабочее состояние ГАУ (кнопка  $SB = 1$ );
- диспетчером установлен код детали, обрабатываемой на ГАУ ( $K3 = 1$ );
- на станке загружена управляющая программа с кодом детали ( $CC13=1$ );
- ограждение рабочей зоны станка убрано ( $D19=0$ );
- в зоне работы роботов не находятся люди ( $D4=0$ ).

Определим название функции как «Разрешение на загрузку станка 1», или сокращенно РЗС1:

$$PZC1 = D5 \cdot CC11 \cdot SB \cdot K3 \cdot CC13 \cdot \overline{D19} \cdot \overline{D4} \quad (2.17)$$

где  $\overline{D19} \cdot \overline{D4}$  – инверсные значения сигналов с датчиков *D19* и *D4*.

В случае, если функция  $P3C1$  принимает значение 1, вызывается локальная управляющая программа у напольного робота на загрузку станка 1.

**Функция, разрешающая портальному роботу установить тару с готовыми деталями на приемную позицию автоматизированного склада**

Пусть портальный робот находится на текущем такте в состоянии бездействия (состояние  $СПР1$ ) (см. табл. 9 и 10). Его переход в состояние загрузки приемной позиции автоматизированного склада ( $СПЗ$ ) может быть произведен в случае, если:

- робот уже взял тару с одной из двух позиций разгрузки станка (состояние  $СПР6=1$  или  $СПР7=1$ );
- диспетчером установлено рабочее состояние ГАУ (кнопка  $SB = 1$ );
- в зоне работы роботов не находятся люди ( $Д4=0$ );
- приемная позиция АС не загружена ( $Д3=0$ ).

Определим название функции как «Разрешение на загрузку приемной позиции АС», или сокращенно  $P3ПАС$ :

$$P3ПАС = (СПР6 + СПР7) \cdot SB \cdot \overline{Д4} \cdot \overline{Д3} \quad (2.18)$$

В случае если функция  $P3ПАС$  принимает значение 1, вызывается локальная управляющая программа у портального робота на загрузку приемной позиции АС. Полный перечень логических функций ГАУ приведен в таблице 2.13.

Таблица 2.13

Вспомогательные логические функции ГАУ

№ п.п	Формула	Описание
1.	$P3C1 = Д5 \cdot CC11 \cdot SB \cdot CC13 \cdot K3 \cdot \overline{Д19} \cdot \overline{Д4}$	Разрешение на загрузку МС1
2.	$P3ПАС = (СПР6 + СПР7) \cdot SB \cdot \overline{Д4} \cdot \overline{Д3}$	Разрешение на загрузку приемной позиции АС
3.	$P3C2 = Д10 \cdot CC21 \cdot SB \cdot K3 \cdot CC23 \cdot \overline{Д17} \cdot \overline{Д4}$	Разрешение на загрузку МС2
4.	$PO1 = Д15 \cdot Д19 \cdot \overline{Д7} \cdot CCHП11 \cdot SB$	Обработка детали на МС1
5.	$PO2 = Д16 \cdot Д17 \cdot \overline{Д18} \cdot CCHП21 \cdot SB$	Обработка детали на МС2
6.	$PC1 = \overline{Д7} \cdot \overline{Д8} \cdot \overline{Д19} \cdot CC11 \cdot SB$	Разгрузить МС1
7.	$PC2 = \overline{Д18} \cdot \overline{Д12} \cdot \overline{Д17} \cdot CC21 \cdot SB$	Разгрузить МС2
8.	$P3ПП1 = Д21 \cdot CCHП11 \cdot Д5 \cdot \overline{Д4}$	Разрешить забрать тару с приемной позиции МС1
9.	$P3ПП2 = Д23 \cdot CCHП21 \cdot Д10 \cdot \overline{Д4}$	Разрешить забрать тару с приемной позиции МС2
10.	$P3ППВ1 = Д22 \cdot Д8 \cdot CCHП11 \cdot \overline{Д4}$	Разрешить забрать тару с позиции выгрузки МС1
11.	$P3ППВ2 = Д24 \cdot Д12 \cdot CCHП21 \cdot \overline{Д4}$	Разрешить забрать тару с позиции выгрузки МС2
12.	$PПАС = (P3ПП1 + P3ПП2) \cdot \overline{Д3} \cdot \overline{Д4} \cdot SB$	Разрешить положить тару на приемную позицию АС
13.	$PПС = Д13 \cdot K3 \cdot SB \cdot Д1$	Принять тару на склад
14.	$PВПС = \overline{Д14} \cdot K3 \cdot SB \cdot Д1$	Выдать тару со склада
15.	$PКП = \overline{Д20} \cdot Д1 \cdot SB$	Комплектовать тару
16.	$PПК = Д20 \cdot Д1 \cdot SB$	Принять тару с комплектования

\*Примечание: Логическая переменная  $K3$  обозначает, что диспетчер на пульте установил кодовое значение детали, производимой в настоящий момент на ГАУ.

### 2.3.5. Сетевая модель управления ГАУ

Представим технологические циклы работы ГАУ в виде сетевых графов (рис. 2.16 – 2.18) в вершинах (окружностях) проставляются номера (коды) состояний. Вершины соединяются стрелками, отражающими переходы из одного состояния в другое. Над стрелками записываются комбинации переменных, обуславливающих этот переход.

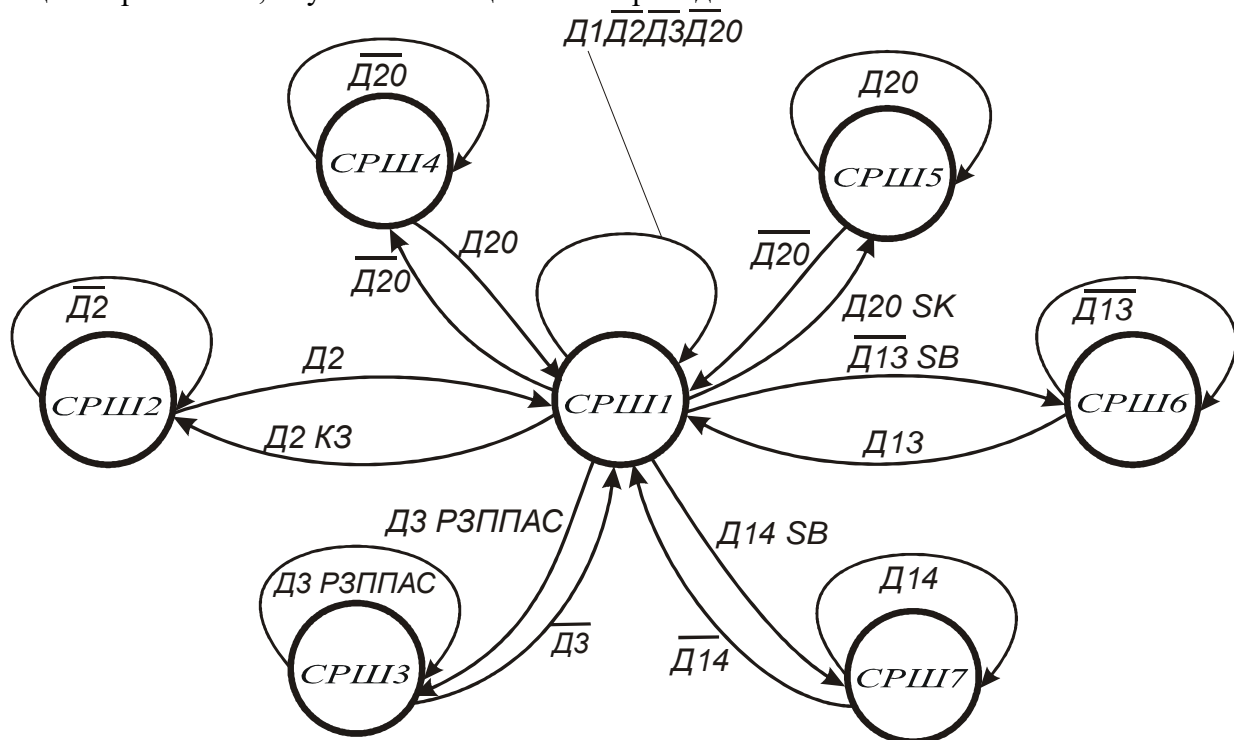


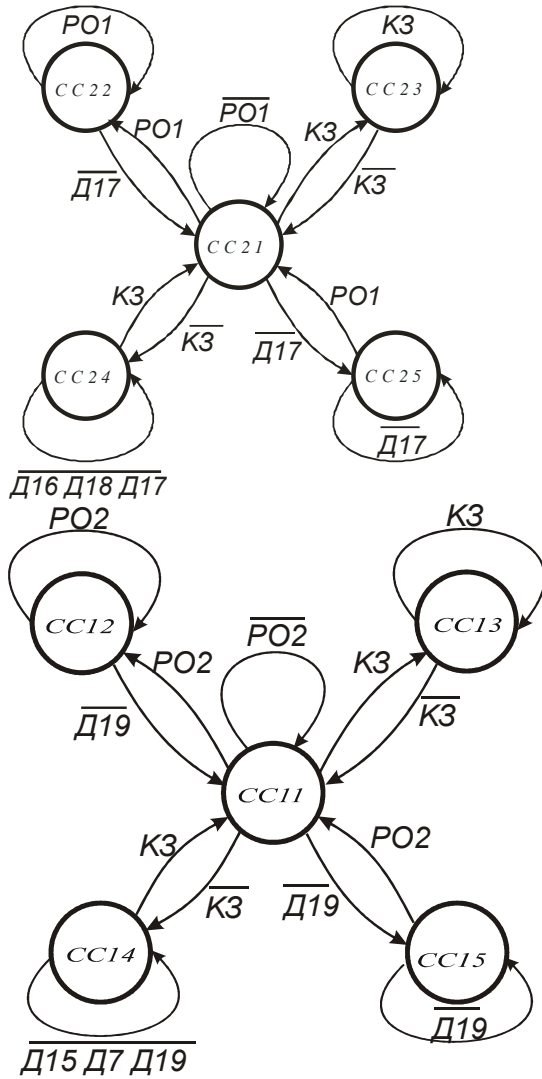
Рис. 2.16. Сетевая модель управления роботом-штабеллером

Логические функции управления роботом-штабеллером:

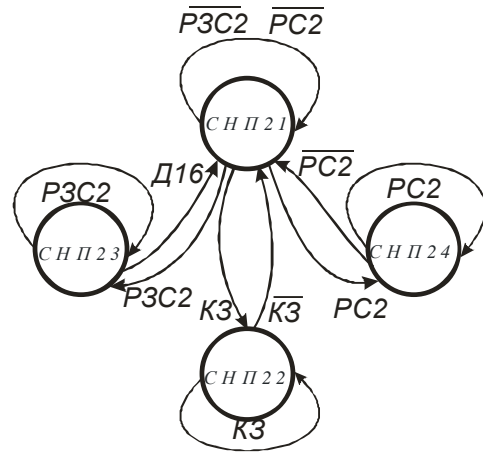
$$\begin{aligned}
 СРШ1 &= СРШ1 \cdot \overline{D1} \cdot \overline{D2} \cdot \overline{D3} \cdot \overline{D20} + СРШ2 \cdot D2 + СРШ3 \cdot \overline{D3} + \\
 &+ СРШ4 \cdot D20 + СРШ5 \cdot \overline{D20} + СРШ6 \cdot D13 + СРШ7 \cdot \overline{D14} \\
 СРШ2 &= СРШ2 \cdot \overline{D2} + СРШ1 \cdot D2 \cdot K3 \\
 СРШ3 &= СРШ3 \cdot \overline{D3} \cdot R3PPAC + СРШ1 \cdot D3 \cdot R3PPAC \\
 СРШ4 &= СРШ4 \cdot \overline{D20} + СРШ1 \cdot D20 \\
 СРШ5 &= СРШ5 \cdot \overline{D20} + СРШ1 \cdot D20 \\
 СРШ6 &= СРШ6 \cdot \overline{D13} + СРШ1 \cdot D13 \cdot SB \\
 СРШ7 &= СРШ7 \cdot \overline{D14} + СРШ1 \cdot D14 \cdot SB
 \end{aligned}
 \tag{2.19}$$



Сетевая модель управления 2-м станком



Сетевая модель управления 1-м напольным роботом



Сетевая модель управления 2-м напольным роботом

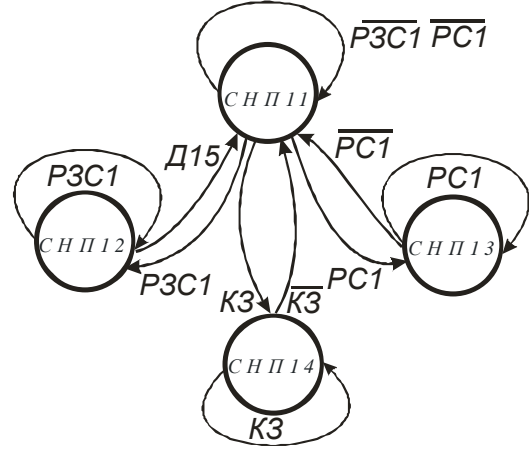


Рис. 2.17. Сетевые модели управления элементами ГАУ

Рассмотрим для примера работу логической функции состояния  $CPШЗ$  робота-штабелера:

$$CPШЗ = CPШЗ \cdot ДЗ \cdot PЗПАС + CPШ1 \cdot ДЗ \cdot PЗПАС$$

Робот-штабелер переходит в состояние  $CPШЗ=1$  (принять тару заданного кода с приемной позиции АС, см. табл. 2.11) если в единичное состояние перейдет одна из комбинаций логических переменных:

- $CPШЗ \cdot ДЗ \cdot PЗПАС$  – робот уже находился в состоянии  $CPШЗ=1$  и датчик  $ДЗ=1$  (есть тара с определенным типом деталей в загрузочной зоне, см. табл. 11) и  $PЗПАС=1$  (Получено разрешение на загрузку приемной позиции АС, см. табл. 2.13);
- $CPШ1 \cdot ДЗ \cdot PЗПАС$  – робот находился в состоянии бездействия  $CPШ1=1$  и датчик  $ДЗ=1$  (есть тара с определенным типом деталей в загрузочной зоне, см. табл. 11) и  $PЗПАС=1$  (Получено разрешение на загрузку приемной позиции АС, см. табл. 2.13).

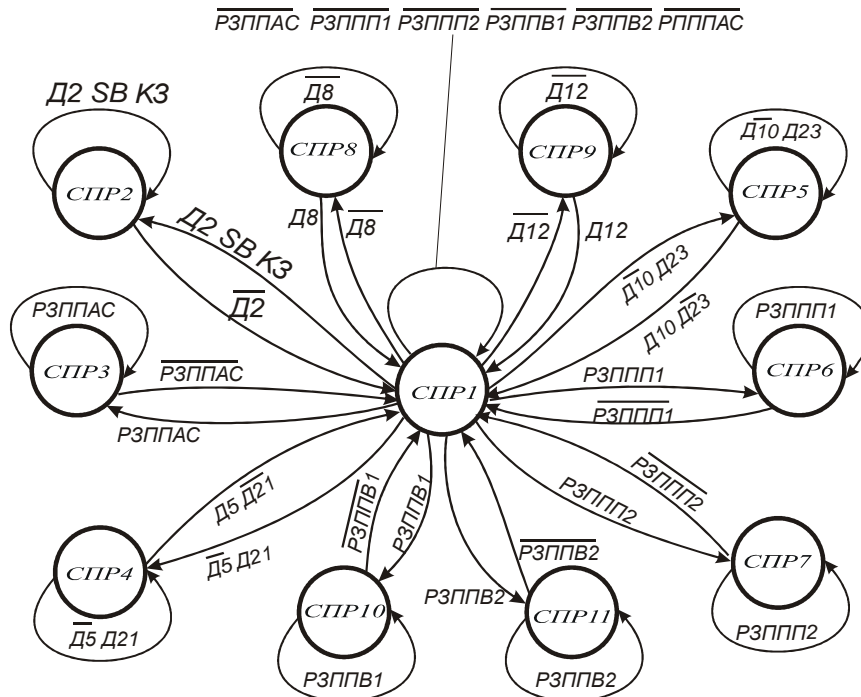


Рис. 2.18. Сетевая модель управления порталным роботом

Полученные сетевые графы могут быть использованы для наглядного отображения правил управления элементами ГАУ, а построенные по ним логические функции применяют при программировании логического контроллера для управления ГАУ.

#### 2.4. Задание на выполнение самостоятельной работы

Для заданного набора автоматизированных элементов ГАУ составить общее описание их работы, определить набор состояний каждого элемента, произвести расстановку информационных датчиков, составить при необходимости вспомогательные логические функции работы системы управления ГАУ, представить работу не менее двух элементов ГАУ в виде сетевых графов и определить для этих элементов набор логических функций управления. При желании студент имеет право включать дополнительные элементы в состав ГАУ. Варианты заданий приведены в таблице 2.14.

Таблица 2.14 Варианты заданий

№№ варианта	Состав автоматизированных элементов ГАУ*
1.	М1С, М2С, НР1, ТС1, ТС2
2.	ОЦ1, ОЦ2, АС, ПР
3.	М1С1, М1С2, ПР, АС
4.	НР, УСС, ОЦ1, ОЦ2, ТС
5.	КИМ, НР, ОЦ, ПЗ, ПВ
6.	НР, УСС, ТС, М1С, ПВ
7.	ПЗ, ТС, ПР, М1С, М2С,
8.	ОЦ1, ОЦ2, НСЛ, ПР
9.	ОЦ1, ОЦ2, АС, НР1, НР2
10.	ПР, УСС, ОЦ, ТС
11.	ОЦ, АС, ПЗ, ПВ, НР
12.	М1С1, М1С2, ПР, АС
13.	М2С, НР, ТС, ПВ
14.	ОЦ1, ОЦ2, НСК1, НСК2, РТ
15.	М1С, М2С, ПЗ, ПВ, НР
16.	НСЛ, РТ, ОЦ1, М2С

№№ варианта	Состав автоматизированных элементов ГАУ*
17.	ТС, НР, М2С1, М2С2
18.	АС, РТ, НР, М2С
19.	ТС1, ТС2, НР, М2С
20.	ОЦ, НСК, РТ, АС

Перечень применяемых сокращений:

Обрабатывающий центр	ОЦ
Многоцелевой токарный одношпиндельный станок	М1С
Многоцелевой токарный двухшпиндельный станок	М2С
Напольный робот	НР
Портальный робот	ПР
Роботизированная тележка	РТ
Позиция загрузки	ПЗ
Позиция выгрузки	ПВ
Тактовый стол	ТС
Автоматизированный склад	АС
Накопитель спутников линейный	НСЛ
Накопитель спутников круговой	НСК
Устройство смены схватов	УСС
Контрольно-измерительная машина	КИМ

## 2.5. Контрольные вопросы

1. Что такое СКНФ, СДНФ?
2. Зачем производится минимизация СДНФ?
3. Отличия последовательностных и комбинационных автоматов?
4. Назначение дешифратора в схеме на рис. 2.8?
5. Определите логическую функцию перехода в состояние *СПР11* на рис. 2.18?
6. Определите логическую функцию перехода в состояние *СНП14* на рис. 17?
7. Дайте текстовое описание функции *РППАС* в табл. 2.13
8. Дайте текстовое описание функции *РЗС2* в табл. 2.13
9. Запишите логические функции управления бункерами Б2 и Б3 из п. 2.2.
10. Перечислите основные этапы формализации технологического цикла и дайте пояснения по каждому этапу.

## ЛЕКЦИЯ 21. Статистическое моделирование процессов и объектов

Принципиальная особенность *случайной величины* (СВ) – невозможность предсказания ее конкретного значения до испытания.

**Дискретные СВ** - целочисленные значения (число двигателей, работающих на объекте в какой либо момент времени). **Непрерывные СВ** – любое значение в некотором интервале (размер поверхности детали).

Основные параметры СВ приведены в таблице.

Название	Генеральные	Выборочные
Объем СВ	$N \rightarrow \infty$ Бесконечное (или очень большое число) проявление СВ	$n$ – некоторое число наблюдаемых значений СВ $X_i, i = 1..n$ $n \leq 5$ – малая выборка $10 < n < 30$ – средняя выборка $50 < n < 200$ – большая выборка
Математическое ожидание СВ	Определяет наиболее вероятную зону нахождения СВ при бесконечно числе ее проявлений $a1 < MX < a2$ Не является конкретным числом, определяется границами доверительного интервала $a1, a2$	Выборочное среднее $\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$
Дисперсия СВ	Определяет наиболее вероятный размер области рассеяния СВ при бесконечно числе ее проявлений $b1 < SX < b2$ Не является конкретным числом, определяется границами доверительного интервала $b1, b2$	Выборочная дисперсия $\sigma = \frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n - 1}$

Как охарактеризовать СВ? Необходимо получить числовые характеристики ее генеральных и (или) выборочных параметров.

**Для получения количественных оценок СВ необходимо провести СТАТИСТИЧЕСКИЙ АНАЛИЗ.**

Основные этапы статистического анализа:

- 1 Получить выборочную совокупность СВ  $X_i, i=1..n$
- 2 Разделить поле рассеяния  $RX = \max(X) - \min(X)$  на  $K$  интервалов получив вектор границ интервалов  $Y$

$$K = 1 + 1.43 * \ln(n)$$

$$Y_1 = \min(X) \quad Y_{j+1} = Y_j + (\max(X) - \min(X)) / K, \quad j = 1..K+1$$

- 3 Для каждого интервала  $[Y_j, Y_{j+1}]$  определить количество попавших в него значений  $X$ , то есть получить вектор частот  $v_j$ .

4. Построить ГИСТОГРАММУ распределения величины  $X$  – столбчатую диаграмму зависимости  $v_j$  от  $Y_j$  (см. рисунок)

- 5 Построить кривую эмпирического распределения

- 6 Сопоставить полученный закон эмпирического распределения с известным теоретическим законом распределения. Это можно сделать и визуально. Но лучше использовать специальные критерии соответствия (например, для нормального распределения Гаусса – критерии Колмогорова или ХИ-квадрат)

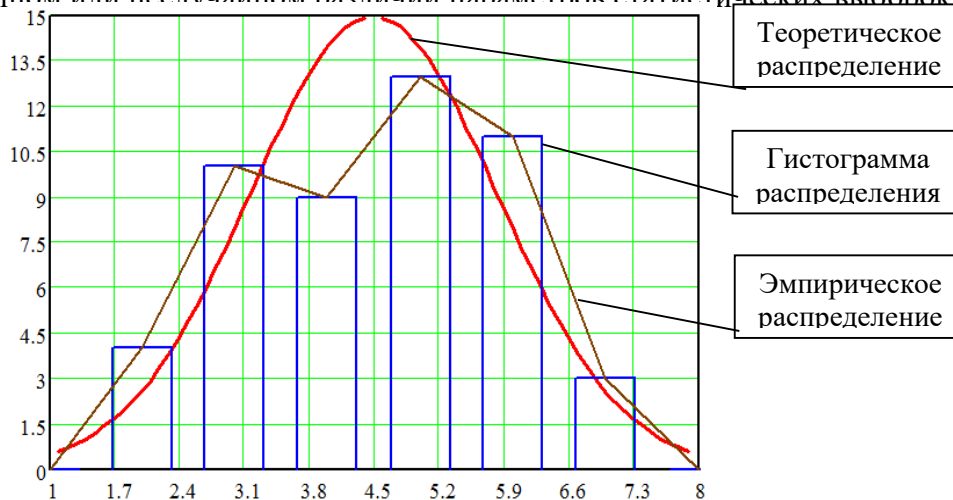
**Как только определен закон распределения СВ, любые ее параметры могут быть вычислены.**

Существует **два типа задач** математической статистики.

1. По выборочным числовым характеристикам СВ получить представление о числовых характеристиках распределения генеральной совокупности.

2. По выборочным числовым характеристикам СВ обоснованно судить о правомерности принятия некоторой статистической гипотезы, например:

- выдвинутой гипотезе распределения генеральной совокупности;
- принадлежности двух или более выборок одной генеральной совокупности;
- случайном или неслучайном различии параметров статистических выборок



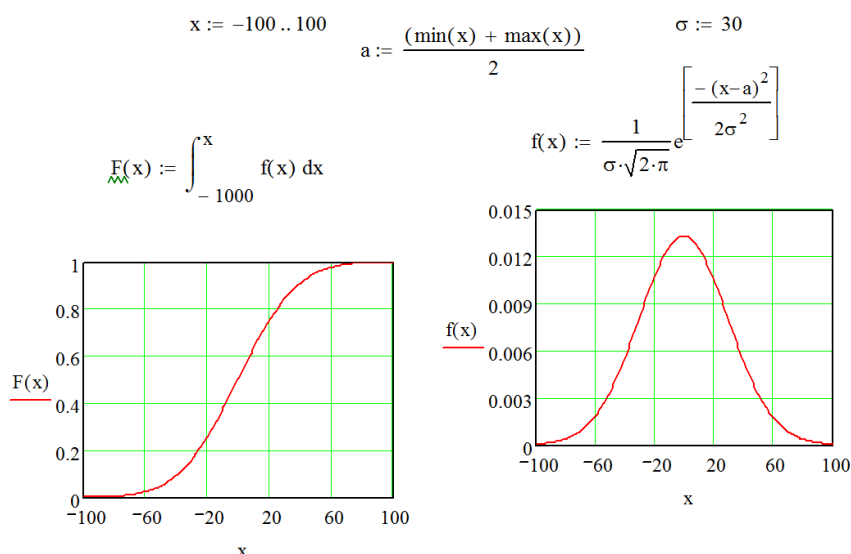
**Список гипотез для задач второго типа НЕОГРАНИЧЕН и зависит от творческой активности и успешности экспериментатора.**

Решение задач основано на использовании законов распределения СВ и их квантилей распределения.

**Закон распределения СВ** - это отношение, устанавливающее связь между возможными значениями СВ и вероятностями, с которыми принимаются эти значения. Закон распределения полностью характеризует СВ.

Пусть  $X$  – некоторая случайная величина, проявляющая себя на числовой оси  $X$ . Рассматривается событие, заключающееся в условии, что  $X$  окажется меньше чем некоторая величина  $X_p$  ( $x < X_p$ ). Вероятность этого события обозначается как  $p(x < X_p)$ . Уровень значимости  $q=1-p$ . **Всегда  $p+q=1$ .**

Предположим, что известна функция закона распределения случайной величины  $X$ , то есть для каждого  $X_p$  имеется соответствующая ему вероятность события  $p(x < X_p)$  – одностороннее распределение или  $p(X_{p1} < x < X_{p2})$  – двухстороннее распределение.



Рассмотрим процесс построения квантилей стандартного нормального распределения. (MathCad)

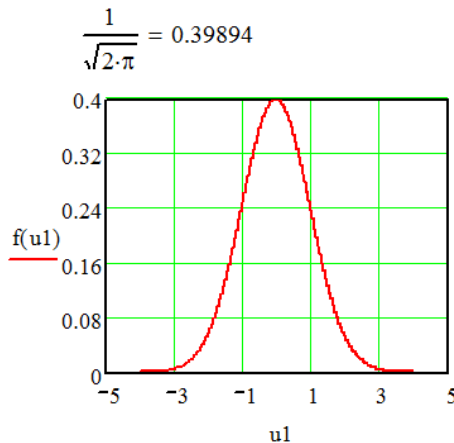
**Квантиль закона распределения**  $X_p$  - значение случайной величины  $X$  с приписанной ей вероятностью  $P$ . Другое определение квантиля - табулированное значение функции закона распределения случайной величины.

Рассмотрим построение закона нормального распределения Гаусса  $F(x)$  и функции плотности нормального распределения  $f(x)$  (MathCad).

Введем новую случайную величину  $u(x) := \frac{x-a}{\sigma}$   $u := -4, -3..5$

$$f(u) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \quad u1 := -4, -3.99..4$$

$$F(u) := \int_{-5}^u f(u) du$$



u =	if(u < 0, F(u), F(u) - F(-u))
-4	3.13846·10 <sup>-5</sup>
-3	0.00135
-2	0.02275
-1	0.15865
0	0
1	0.68269
2	0.9545
3	0.9973
4	0.99994
5	1

$$F(1.96) - F(-1.96) = 0.95$$

**Стандартное нормальное распределение – симметричное двухстороннее распределение**

Наиболее часто применяемый в технике уровень вероятности  $p=0.95$ . Квантиль стандартного нормального распределения составляет  $\pm 1,96$

*Алгоритм решения задач первого типа*

Пусть известен закон распределения случайной величины  $X$  - стандартное нормальное распределение,  $\sigma$  и результат одного наблюдения  $x_0$ . Определить генеральное среднее случайной величины  $X$ .

**ЗАДАЧА ПЕРВОГО ТИПА**, так как необходимо найти параметр генеральной совокупности !!!

- Находим связь заданной в задаче СВ с СВ, квантили которой известны.  
Так как  $X$  подчиняется стандартному нормальному распределению, то (см. рис. выше)

$$x = a + \sigma \cdot u \Rightarrow u = (x - a) / \sigma$$

- Выбрать доверительную вероятность ответа.  $P=0.95$
- Установить квантильные границы и сформировать разрешающее задачу неравенство

$$u_{p=0.95} = \pm 1,96$$

$$-1.96 < u < 1.96$$

$$-1.96 < (x_0 - a) / \sigma < 1.96$$

- Решить неравенство относительно определяемой величины

$$x_0 - 1.96\sigma < a < x_0 + 1.96\sigma$$

*Алгоритм решения задачи второго типа*

В опыте получено одно значение СВ  $u=u_0$ . Проверить гипотезу о том, что  $U$  подчиняется стандартному нормальному распределению.

**ЗАДАЧА ВТОРОГО ТИПА**, так как необходимо проверить статистическую гипотезу!!!

- Находим связь заданной в задаче СВ с СВ, квантили которой известны.

Так как и проверяется на соответствие стандартному нормальному распределению,

$$u = u$$

2. Выбрать доверительную вероятность ответа.  $P=0.95$
3. Установить квантильные границы и сформировать разрешающее задачу неравенство
 
$$u_{p=0.95} = \pm 1,96$$

$$-1.96 < u_0 < 1.96$$
4. Проверить истинность неравенства относительно определяемой величины  
 Если неравенство истинно – принять гипотезу. Если неравенство ложно – отвергнуть гипотезу.  
 При решении статистических задач можно совершить ошибки первого или второго рода.  
 Ошибка 1 рода - отвергнуть хорошую гипотезу.  
 Ошибка 2 рода - не отвергнуть неверную гипотезу.  
 Вероятность совершить ошибку первого рода падает с уменьшением  $q=1-P$ .  
 Вероятность совершить ошибку второго рода уменьшается с ростом  $q=1-P$ .

### Статистические критерии и их применение

Решение статистических задач проводится на основе ряда специальных распределений с известными (табулированными) квантилями. Общий подход состоит в том, что вводится случайная величина, связанная с изучаемой и удобная для решения той или иной конкретной задачи.

#### Распределение Стьюдента (t-критерий).

Распределение было получено Госсетом (псевдоним Стьюдента) в 1908 г. Зависит от объема выборки  $N$  или числа степеней свободы  $f=N-1$ , с которым определена выборочная дисперсия  $S^2$  (среднее квадратическое отклонение (СКО) выборки  $S$ ), и от заданной вероятности ответа, определяемой параметром уровня значимости  $q$ . Формула критерия:

$$t = \frac{\bar{x} - a}{s} \sqrt{N}, \quad (4.1)$$

где  $a$  — генеральное среднее исследуемой случайной величины  $X$ ;

$\bar{x}$  — выборочное математическое ожидание  $X$ .

Распределение симметрично относительно начала координат, т. е.  $t_{q/2} = -t_{1-q/2}$ .

Более полные таблицы квантилей распределения Стьюдента для уровня значимости  $q \neq 0,05$  содержатся в [1].

Таблица 4.1 Квантили распределения Стьюдента при  $q=0,05$

$f=N-1$	1	2	3	4	5	6	7	8	9	10	20	30	$\infty$
$t_{1-q/2}$	12,71	4,30	3,18	2,78	2,57	2,45	2,37	2,31	2,26	2,23	2,09	2,04	1,96

Распределение Стьюдента дает возможность находить генеральное среднее или проверять статистические гипотезы при очень малых выборках.

**Пример 1.** Известны три значения нормально распределенной случайной величины  $X$ :  $x_1 = 10$ ;  $x_2 = 9,5$ ;  $x_3 = 10,2$ . Требуется оценить генеральное среднее с вероятностью  $p = 0,95$  (задача первого типа).

Определим выборочное ( $N = 3$ ) математическое ожидание  $\bar{x}$

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{10 + 9,5 + 10,2}{3} = 9,9.$$

Определим выборочное СКО при  $f = N-1=2$ :

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{(10-9,9)^2 + (9,5-9,9)^2 + (10,2-9,9)^2}{2}} = 0,36.$$

Запишем выражение для  $t$  в соответствии с (4.1):

$$t = \frac{\bar{x} - a}{s} \sqrt{N} = \frac{9,9 - a}{0,36} \sqrt{3}.$$

По табл. 1 определим квантильные границы  $t_{1-q/2} = 4,30$ ,  $t_{q/2} = -4,30$  и запишем неравенство

$$-4,30 < \frac{9,9 - a}{0,36} \sqrt{3} < 4,30.$$

Решив неравенство относительно  $a$ , получим с доверительной вероятностью  $p = 0,95$   $9,0 < a < 10,8$ .

**Пример 2.** Проверить гипотезу, состоящую в том, что нормально распределенная случайная величина  $X$  имеет генеральное математическое ожидание  $a = 10$  на основании результатов двух испытаний:  $x_1=8,1$ ,  $x_2=8,9$  (задача второго типа).

По результатам испытаний определяем  $\bar{x} = 8,5$  и  $s = 0,57$ .

Вычисляем по (4.1) значение критерия Стьюдента  $t$ :

$$t = \frac{8,5 - 10}{0,57} \sqrt{2} = -3,71.$$

Выбираем уровень значимости  $q = 0,05$  и по табл. 4.1 находим для  $f=N-1=1$  границы критических областей гипотезы  $t_{1-q/2} = 12,71$ ,  $t_{q/2} = -12,71$ .

Гипотеза не отвергается на уровне значимости  $q=0,05$ , поскольку  $t = -3,71$  лежит вне критических областей. Отметим, что очень малая информация ( $N = 2$ ) и низкий уровень значимости не дают оснований отвергнуть плохую на взгляд гипотезу. Если бы, например, те же результаты

( $\bar{x} = 8,5$  и  $s = 0,57$ ) были получены при  $N = 4$  ( $f = 3$ ), то, как легко видеть, гипотезу следовало отвергнуть на том же уровне значимости. При выборе более жесткого уровня значимости,  $q = 0,2$  и  $T_{1-q/2} = 3,08$ , т. е. гипотеза также должна быть отвергнута.

### Распределение Пирсона

Распределение Пирсона ( $\chi^2$ -критерий) удобно для оценки генеральной дисперсии  $\sigma^2$  по выборочной  $s^2$ . В 1900 г. Пирсон ввел случайную величину

$$\lambda^2 = \frac{(N-1) \cdot S^2}{\sigma^2} \quad (4.2)$$

и нашел ее распределение, зависящее лишь от  $f = N - 1$ . Оно несимметрично, следовательно,  $\lambda_{q/2}^2 = \lambda_{1-q/2}^2$ . Некоторые квантили приведены в табл.2. Более полные таблицы содержатся в справочниках [1].

Таблица 4.2. Квантили распределения Пирсона  $\lambda_{1-q}^2$

$f$	1	2	3	4	5	6	7	8	9	10	15	20
$\lambda_{1-0,95}^2$	0,0039	0,103	0,352	0,71	1,14	1,63	2,17	2,73	3,32	3,94	7,3	10,9
$\lambda_{1-0,05}^2$	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9	18,3	25,0	31,4

*Критерий Пирсона позволяет решать статистические задачи о дисперсиях.*



**Пример 3.** Найти с вероятностью  $p = 0,9$  доверительный интервал для  $\sigma^2$  нормально распределенной случайной величины, если при  $f = 5$ ,  $S^2 = 1$ . Определив по табл. 2 квантильные границы  $\lambda_{0,05}^2 = 1,14$  и  $\lambda_{0,95}^2 = 11,1$ , имеем  $1,14 < fs^2 / \sigma^2 < 11,1$  или после преобразований  $0,45 < \sigma^2 < 4,39$

**Пример 4.** Проверить гипотезу, состоящую в том, что генеральная дисперсия нормально распределенной случайной величины  $\sigma^2 = 2$ , если обработка данных десяти опытов ( $f = 9$ ) дала  $S^2 = 4$ . В связи с тем, что  $s^2 \geq \sigma^2$ , имеет смысл использовать одностороннюю оценку сверху. Выбрав  $q = 0,05$ , определим по табл. 2  $\lambda_{0,95}^2 = 16,9$ . Обнаруживаем, что  $\lambda^2 = \frac{9 \cdot 4}{2} = 18 > 16,9$ , т. е. гипотеза должна быть отвергнута на уровне значимости 0,05,

поскольку рассчитанная величина  $\lambda^2$  оказалась в критической области.

### Распределение Фишера (F-критерий)

Используется для проверки однородности двух выборочных дисперсий  $s_1^2$  и  $s_2^2$  (обычно принимают  $s_1^2 > s_2^2$  и используют односторонние оценки), найденных соответственно с  $f_1$  и  $f_2$ , с целью установить или отвергнуть их принадлежность одной генеральной совокупности. Фишером введена случайная величина

$$F = \frac{s_1^2}{s_2^2} \quad (4.3)$$

и построено ее распределение, зависящее от  $f_1$  и  $f_2$ . Некоторые квантили распределения Фишера для  $q=0,05$  приведены в табл. 4.3.

Таблица 4.3 Квантили распределения Фишера  $F_{1-0,05}$

$f_2$	$f_1$								
	1	2	3	4	5	6	12	24	$\infty$
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2	18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
$\infty$	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

**Пример 5.** Проверить гипотезу об однородности двух выборочных дисперсий нормально распределенной случайной величины. Результаты опытов:  $s_1^2 = 0,9$  при  $f_3 = 3$  и  $s_2^2 = 0,2$  при  $f_2 = 5$ .

Выбрав  $q=0,05$ , находим по табл. 3  $F_{1-q} = 5,4$ . Из данных опытов имеем  $F = 0,9/0,2 = 4,5$ , следовательно,  $F < F_{1-q}$  и гипотеза об однородности  $s_1^2$  и  $s_2^2$ , т. е. о принадлежности их

одной генеральной совокупности, характеризуемой  $\sigma^2$ , не отвергается на уровне значимости 0,05

### Распределение Кохрена (G-критерий)

Используется для проверки однородности  $k$  выборочных дисперсий, найденных с одинаковыми числами степеней свободы  $f_j = N - 1$ . Кохреном введена случайная величина

$$G = \frac{s_{j,\max}^2}{\sum_{j=1}^k s_j^2}, \quad (4.4)$$

где  $s_{j,\max}^2$  — наибольшая из сравниваемых дисперсий; и построено ее распределение, зависящее от  $f_j$  и  $k$ . В табл. 4.4 приведены некоторые квантили  $G_{1-q}$  для  $q=0,05$ .

Таблица 4.4 Квантили распределения Кохрена  $C_{1-0,05} \times 10^2$

$k$	$f_j$										
	1	2	3	4	5	6	7	8	9	10	$\infty$
2	100	98	94	91	88	85	83	82	80	79	50
3	97	87	80	75	71	68	65	63	62	60	33
4	91	77	68	63	59	56	54	52	50	49	25
5	84	68	60	54	51	48	46	44	42	41	20
6	78	62	53	48	44	42	40	38	37	36	17
7	73	56	48	43	40	37	35	34	33	32	14
8	68	52	44	39	36	34	32	30	29	28	13
9	64	48	40	36	33	31	29	28	27	26	11
10	60	45	37	33	30	28	27	25	24	24	10
20	39	24	22	19	17	16	15	14	14	13	5
$\infty$	0	0	0	0	0	0	0	0	0	0	0

**Пример 6** Проверить гипотезу об однородности выборочных дисперсий  $s_1^2 = 3$ ,  $s_2^2 = 5$ ,  $s_3^2 = 15$ ,  $s_4^2 = 2$  и  $s_5^2 = 4$ , каждая из которых определена с  $f_j = 4$  на уровне значимости  $q = 0,05$ .

По табл. 4.4 находим для односторонней оценки при  $f_j = 4$  и  $k = 5$   $G_{0,95} = 0,54$ . Из опытных данных в соответствии с (4.4) имеем

$$G = \frac{15}{3 + 5 + 15 + 2 + 4} = 0,52 < 0,54$$

т. е. на уровне значимости 0,05 гипотеза об однородности дисперсий не отвергается.

Отметим здесь, что наилучшей оценкой  $k$ -однородных дисперсий служит дисперсия, определяемая как

$$s^2 = \left( \sum_{j=1}^k s_j^2 \right) / k \quad (4.5)$$

с числом степеней свободы  $f = kf_j = k(N - 1)$ . Эта оценка может использоваться для определения доверительного интервала для генеральной дисперсии  $\sigma^2$ .

### $\tau$ -распределение ( $\tau$ -критерий)

Используется для проверки однородности наблюдений, исключения грубых ошибок или выбросов. Квантили распределения случайной величины

$$\tau = \frac{|x_{ip} - \bar{x}|}{s}, \quad (4.6)$$

зависящего лишь от объема выборки  $N$ , по которой определяются  $\bar{x}$  и  $s$ , приведены для вероятности  $p=0,95$  в табл. 5.

Таблица 4.5 Квантили  $\tau$ -распределения при  $q=0,05$

$N$	3	4	5	6	7	8	9	10	15	20
$\tau_{1-q}$	1,41	1,69	1,87	2,00	2,17	2,24	2,29	2,29	2,49	2,62

**Пример 7.** На одном из пяти одинаковых агрегатов (третьем), выполняющих однотипные технологические операции, были внедрены мероприятия по экономии электроэнергии. Оценить их эффективность, если зарегистрированное месячное потребление энергии каждым агрегатом составляет  $x_1 = 10$ ,  $x_2 = 12$ ,  $x_3 = 8$ ,  $x_4 = 9$ ,  $x_5 = 11$ .

Как и предполагалось, на третьем агрегате расход энергии минимален, т. е.  $x_3 = 8$ . Выбран  $q=0,05$ , найдем по табл. 5 что,  $\tau_{0,95} = 1,87$ . По опытным данным определим  $\bar{x} = 10$ ,  $s=1,58$  и вычислим  $\tau = \frac{|8-10|}{1,58} = 1,27$ .

Поскольку  $\tau = 1,27 < 1,87$ , т. е. не попадает в критическую область гипотезы, нет оснований считать, что достигнутый результат неслучаен, т.е. обусловлен проведенными мероприятиями. Он вполне может быть объяснен естественным разбросом в потреблении электроэнергии отдельными агрегатами, и оснований для выплаты премии в этом случае статистика не дает.

### Дисперсионный анализ

Поставим задачу определить факт влияния факторов  $X$  на функцию цели  $Y$  в условиях действия случайных факторов  $W$ .

Задача решается с помощью дисперсионного анализа (ДА), который позволяет выявить, приведет ли изменение  $X$  к значимому изменению  $Y$  то есть фактически является ли  $X$  варьируемым фактором.

Результат опыта  $Y$  можно представить как сумму 2-х слагаемых:

$$Y = Y' + Y'',$$

где  $Y'$  - часть результата, обусловленная влиянием  $X$ ,

$Y''$  - часть результата, обусловленная влиянием  $W$ .

По свойствам дисперсии:

$$DY = DY' + DY'',$$

где  $DY$  – общая дисперсия результатов наблюдений

$DY'$  - дисперсия  $Y$ , обусловленная влиянием  $X$  ;

$DY''$  - дисперсия  $Y$ , обусловленная влиянием  $W$ .

Если оценить однородность общей дисперсии результатов  $DY$  и дисперсии от влияния случайных факторов  $DY''$ , то в случае принятия гипотезы компонента  $DY'$  (дисперсия от влияния  $X$ ) оказывается незначительной, или  $X$  на  $Y$  не влияет.

Таблица 5.1 План проведения дисперсионного анализа

№ опыта	Уровни $X_j$				
	$X_1$	$X_2$	$X_3$	...	$X_M$
1	$Y_{11}$	$Y_{12}$	$Y_{13}$	...	$Y_{1M}$
2	$Y_{21}$	$Y_{22}$	$Y_{23}$	...	$Y_{2M}$
3	...	...	...	...	...
$i = 1 \dots K$	...	...	...	$j = 1 \dots M$	...
$K$	$Y_{K1}$	$Y_{K2}$	$Y_{K3}$	...	$Y_{KM}$

Итак, дисперсионный анализ сводится к определению дисперсии результатов опытов  $DY''$ , обусловленных влиянием случайных факторов  $W$ . Данная задача может быть решена

только путем запланированного повторения  $K$  опытов на  $M$  уровнях. В табл. 6 приведен стандартный план построения исследований.

Для исключения систематических погрешностей данных измерений последовательность проведения  $N = K \cdot M$  опытов определяется с помощью процедуры **рандомизации** (табл. 5.2).

Таблица 5.2 Рандомизация последовательности проведения опытов

№ п.п.	Генератор случайных чисел от 1 до $M \cdot K$							
	1	1	1	1	2	2	...	$K$
Уровень	1	2	3	$M$	2	3	...	$M$
Порядок проведения	5	2	1	3	8	4	...	24

Обработка данных строится по следующему алгоритму.

1. Определяем выборочное математическое ожидание по каждому уровню фактора  $X$ .

$$\bar{y}_j = \frac{\sum_{i=1}^K y_{i,j}}{K}$$

2. Определяем выборочные дисперсии для каждого уровня фактора  $X$ :

$$S y_j^2 = \frac{\sum_{i=1}^K (y_{i,j} - \bar{y}_j)^2}{K - 1}$$

3. Проверим однородность полученных выборочных дисперсий по критерию Кохрена.

$$G = \frac{\max(S y_j^2)}{\sum_{j=1}^M S y_j^2}$$

Если  $G < G_{\text{табл}}$ , гипотеза однородности принимается.

В случае, когда гипотеза однородности отвергается, необходимо прекратить расчет и рекомендовать:

- 1) Увеличить количество опытов;
- 2) Проверить надежность измерительных средств и принять меры к повышению стабильности поведения объекта.
3. Определим дисперсию от влияния случайных факторов.

$$D y'' = \frac{\sum_{j=1}^M S y_j^2}{M}$$

4. Определим математическое ожидание всех наблюдений:

$$\bar{y} = \frac{\sum_{i=1}^K \sum_{j=1}^M y_{i,j}}{M \cdot K}$$

5. Определим общую дисперсию всех наблюдений

$$D y = \frac{\sum_{i=1}^K \sum_{j=1}^M (y_{i,j} - \bar{y})^2}{M \cdot K - 1}$$

6. Проверим дисперсии  $D y''$  и  $D y$  на однородность по критерию Фишера.

$$F_p = \frac{D y}{D y''}$$

Табличное значение критерия Фишера определяем по табл. 4.3 для

$$f_1 = M \cdot (K - 1), \quad f_2 = M \cdot K - 1$$

Если  $F_p > F_{\text{табл}}$ , то гипотеза однородности дисперсий  $D y$  и  $D y''$  отвергается и, следовательно, с вероятностью  $P$  (см. табл. Фишера) управляемый фактор  $X$  оказывает значимое влияние на функцию цели  $Y$ . В противном случае необходимо сделать противоположный вывод.

## ЛЕКЦИЯ 22. Однофакторный регрессионный анализ

Познакомившись с основными целями и приемами дисперсионного анализа, поставим следующую задачу: научиться определять, как влияет фактор  $X$  (пока один) на функцию цели  $Y$  в условиях действия случайных факторов  $W$ .

Заметим, что из двух величин  $X$  и  $Y$ , между которыми устанавливается связь, лишь одна  $Y$  случайная. Если бы обе величины были неслучайными, то мы бы имели дело с обычной функциональной зависимостью, изучаемой в математическом анализе. Если, наоборот, рассматриваются две случайные величины  $X$  и  $Y$ , то для оценки связи между ними можно воспользоваться коэффициентом корреляции [7].

В нашем случае ( $X$ — неслучайная величина,  $Y$ —случайная) зависимость  $Y=f(X)$  называют *регрессией* или *уравнением регрессии*. Эта зависимость в отличие от функциональной отражает связь между *наиболее вероятным значением* случайной величины  $Y$  и величиной  $X$ .

Основой для нахождения регрессии служит экспериментальный материал —  $N$  точек в плоскости  $Y, X$ . Их расположение пока что несущественно: они могут быть получены, как в рассмотренном примере, или могут появиться в результате установки  $X$  на некоторых фиксированных уровнях и проведения на каждом уровне одного или нескольких измерений  $Y$ . Получение регрессии включает три этапа.

**Этап I:** постулируется вид уравнения регрессии, включающего, кроме  $X$  и  $Y$ , несколько неизвестных параметров (коэффициентов), которые находятся по опытным данным.

**Этап II:** определяются неизвестные коэффициенты, входящие в постулированное уравнение регрессии.

**Этап III:** проводится статистический анализ регрессии с целью доказательства ее адекватности экспериментальным данным, которые она представляет.

**Этап I** выполняется обычно на основе априорных соображений (например, нужна линейная модель), предшествующего опыта, интуиции и других не формализуемых приемов.

Обычно (но не обязательно) уравнение регрессии постулируется в виде полинома. Наиболее распространенными и удобными регрессионными моделями можно считать следующие: модель I порядка  $Y=B_0+B_1X$ ; модель II порядка:  $Y=B_0+B_1X+B_{11}X^2$ .

В практике могут использоваться любые другие модели, в том числе и нелинейные относительно коэффициентов (см. табл. 6.1).

В результате замены переменных, все указанные в табл.6.1 функции могут быть преобразованы в линейную модель первого порядка.

Таблица 6.1. Виды уравнений регрессии и их линеаризующие преобразования

№ п/п	Функции	Линеаризующие преобразования			
		Преобразования переменных		Выражения для величин $B_0$ и $B_1$ .	
		$Y'$	$X'$	$B_0'$	$B_1'$
1.	$Y = B_0 + \frac{B_1}{X}$	$Y$	$\frac{1}{X}$	$B_0$	$B_1$
2.	$Y = \frac{1}{B_0 + B_1 \cdot X}$	$\frac{1}{Y}$	$X$	$B_0$	$B_1$
3.	$Y = \frac{X}{B_0 + B_1 \cdot X}$	$\frac{X}{Y}$	$X$	$B_0$	$B_1$
4.	$Y = B_0 \cdot B_1^X$	$\ln(Y)$	$X$	$\ln(B_0)$	$\ln(B_1)$
5.	$Y = B_0 \cdot e^{B_1 \cdot X}$	$\ln(Y)$	$X$	$\ln(B_0)$	$B_1$
6.	$Y = \frac{1}{B_0 + B_1 \cdot e^{-X}}$	$\frac{1}{Y}$	$e^{-X}$	$B_0$	$B_1$
7.	$Y = B_0 \cdot X^{B_1}$	$\ln(Y)$	$\ln(X)$	$\ln(B_0)$	$B_1$

8.	$Y = B_0 + B_1 \cdot \ln(X)$	$Y$	$\ln(X)$	$B_0$	$B_1$
9.	$Y = \frac{B_0}{B_1 + X}$	$\frac{1}{Y}$	$X$	$\frac{B_1}{B_0}$	$\frac{1}{B_0}$
10.	$Y = \frac{B_0 \cdot X}{B_1 + X}$	$\frac{1}{Y}$	$\frac{1}{X}$	$\frac{B_1}{B_0}$	$\frac{1}{B_0}$
11.	$Y = B_0 \cdot e^{\frac{B_1}{X}}$	$\ln(Y)$	$\frac{1}{X}$	$\ln(B_0)$	$B_1$
12.	$Y = B_0 + B_1 \cdot X^n$	$Y$	$X^n$	$B_0$	$B_1$

Отметим, что неудачно выбранная (постулированная) модель может вызвать непроизводительную трату времени, хотя и не приводит к принципиальным ошибкам в результате, поскольку ее адекватность проверяется на последнем этапе и плохая модель не используется. Поэтому на этапе выбора вида регрессии надо использовать всю имеющуюся информацию. Например, если известно, что  $Y$  и  $X$  обратно пропорциональны, можно выбрать в качестве фактора  $X' = 1/X$ . Если априорной или какой-либо другой информации нет, следует начинать с простейшей модели (I порядка) — даже если она и окажется неадекватной, потребует для получения меньших усилий.

**Этап II** — нахождение неизвестных коэффициентов по опытным данным — обычно выполняют на основе метода наименьших квадратов (МНК). Рассмотрим этот широко используемый в науке и технике «метод и научимся его применять».

Пусть постулирована линейная регрессия и требуется найти  $B_0$  и  $B_1$  по  $N$  опытным точкам. В любом  $u$ -м опыте имеем

$$Y_u = \beta_0 + \beta_1 X_u + \varepsilon_u, \quad (22.1)$$

где  $Y_u, X_u$  — значения случайной величины  $Y$  и неслучайного аргумента  $X$  в  $u$ -м опыте;  $\beta_0, \beta_1$  — «истинные» значения коэффициентов регрессии, соответствующие модели  $Y_{ист} = \beta_0 + \beta_1 X$ ;

$\varepsilon_u$  — ошибка, объясняющая отличие наблюдаемого значения от «истинного» и обусловленная воздействием на  $Y$  случайных факторов.

Очевидно, что нам никогда не удастся найти  $\beta_0$  и  $\beta_1$ , а коэффициенты  $B_0$  и  $B_1$ , которые мы должны определить, есть некоторые *наилучшие оценки*  $\beta_0$  и  $\beta_1$ . Метод наименьших квадратов определяет следующее содержание термина «наилучшие»:

$$\sum_u^N (Y_u - Y_{u\text{ уcm}})^2 \Rightarrow \min,$$

т. е. такие, что сумма квадратов отклонений наблюдаемых  $Y$  от «истинных» минимальна.

Для линейной регрессии следует другая запись условия МНК

$$S = \sum_{u=1}^N (Y_u - \beta_0 - \beta_1 X_u)^2 = \sum_{u=1}^N \varepsilon^2 \Rightarrow \min,$$

непосредственно позволяющая найти оценки  $B_0$  и  $B_1$ . Продифференцировав  $S$  по  $\beta_0$  и  $\beta_1$  и приравняв полученные частные производные нулю, имеем следующую систему линейных алгебраических уравнений:

$$\begin{cases} B_0 N + B_1 \sum_{i=1}^N X_i = \sum_{i=1}^N Y_i \\ B_0 \sum_{i=1}^N X_i + B_1 \sum_{i=1}^N X_i^2 = \sum_{i=1}^N (Y_i \cdot X_i) \end{cases}, \quad (22.2)$$

решая которую любым способом, определяем неизвестные коэффициенты уравнения регрессии  $B_0$  и  $B_1$ .

Познакомимся с «инженерным методом» формирования системы линейных уравнений для нахождения неизвестных коэффициентов регрессии. Анализ уравнения 6.1 и первого

уравнения системы 6.2 показывает, что в них есть «нечто общее». Второе уравнение системы 6.2. получается путем «умножения» первого уравнения на сумму  $X$ . Эти наблюдения позволяют сформировать «эвристический алгоритм формирования СЛАУ» заключающийся в следующем:

*Первое уравнение системы соответствует виду выбранного уравнения регрессии с заменами:*

$$B_0 \text{ на } B_0N; \quad X \text{ на сумму } X; \quad Y \text{ на сумму } Y$$

*Каждое последующее уравнение системы получаем из предыдущего путем умножения на сумму  $X$ .*

Проиллюстрируем алгоритм на примере.

Пусть необходимо найти неизвестные коэффициенты регрессионного уравнения вида

$$Y = B_0 + B_1X + B_2X^2$$

В уравнении три неизвестных коэффициента. Следовательно, порядок СЛАУ равен

3. Применяя «эвристический алгоритм», получаем:

$$\begin{cases} B_0N + B_1 \sum_{i=1}^N X_i + B_2 \sum_{i=1}^N X_i^2 = \sum_{i=1}^N Y_i \\ B_0 \sum_{i=1}^N X_i + B_1 \sum_{i=1}^N X_i^2 + B_2 \sum_{i=1}^N X_i^3 = \sum_{i=1}^N (Y_i \cdot X_i) \\ B_0 \sum_{i=1}^N X_i^2 + B_1 \sum_{i=1}^N X_i^3 + B_2 \sum_{i=1}^N X_i^4 = \sum_{i=1}^N (Y_i \cdot X_i^2) \end{cases}$$

Отметим, что изложенная процедура может быть осуществлена в любых случаях — МНК всегда позволит провести линию, соответствующую постулированному выражению и наилучшим образом (в указанном выше смысле) отвечающую экспериментальным точкам.

**Этап III** — анализ адекватности регрессии, в результате которого мы должны выяснить, удачно ли выбрана исходная модель, т. е. адекватно ли отражает полученная регрессия экспериментальный материал, положенный в ее основу.

Проверку адекватности постулированной модели произведем путем сравнения двух дисперсий, одна из которых  $S_Y^2$  характеризует влияние на  $Y$  случайных факторов, а вторая  $S_{ad}^2$  оценивает разброс опытных значений  $Y$  относительно линии регрессии и вычисляется

$$S_{ad}^2 = \frac{\sum_{i=1}^N (Y_i - y(X_i))^2}{N - K}$$

с числом степеней свободы  $f_{ad} = N - K$  (в модели по  $N$  опытам определены  $K$  коэффициентов).

Составим отношение

$$F_p = \frac{S_Y^2}{S_{ad}^2}$$

и сравним его с квантильной границей критерия Фишера  $F_T$  при  $f_1 = N - 1$  и  $f_2 = f_{ad}$ . Если  $F < F_T$  то гипотеза об однородности сравниваемых дисперсий не отвергается на уровне значимости  $q$ , следовательно, наблюдаемый разброс  $Y$  и  $y(X_i)$  может быть объяснен естественным влиянием случайных факторов и модель неадекватно описывает процесс. Если же  $F > F_T$  то  $S_{ad}^2$

значимо меньше  $S_Y^2$ , т. е. постулированная модель статистически значимо описывает процесс и может быть признана адекватной. Заметим, что в случае, когда дисперсия адекватности стремится к нулю, расчетный критерий Фишера стремится к бесконечности. Таким образом, дисперсия адекватности может быть признана основным параметром, оценивающим качество построенной модели.

## ЛЕКЦИЯ 23. Многофакторный регрессионный анализ

Наиболее интересные и практически важные задачи обычно предполагают получение зависимости  $Y$  от *нескольких воздействующих факторов*  $\{X \rightarrow X_1, X_2, \dots, X_i, \dots, X_n\}$ . Такие задачи приводят к многофакторному регрессионному анализу.

Очевидно, что иметь дело с несколькими разнородными независимыми переменными, которые выражаются совершенно различными именованными числами и варьируются в различных пределах и т. п., по крайней мере, неудобно. Для того чтобы избавиться от этого неудобства, перейдем к кодированным факторам, которые будем обозначать строчными буквами  $x$  с соответствующими индексами и которые будут связаны с исходными истинными факторами  $X$  соотношениями

$$x = \frac{X - X_{cp}}{X_{cp} - X_{min}} = \frac{X - X_{cp}}{X_{max} - X_{cp}},$$

где  $X_{min}$ ,  $X_{max}$  — минимальное и максимальное значения фактора, выбранные внутри диапазона его изменения (варьирования);  $X_{cp}$  — среднее значение фактора, определяемое как

$$X_{cp} = \frac{X_{min} + X_{max}}{2}.$$

После выполнения операции кодирования мы имеем дело с безразмерными факторами, варьирующимися около нуля в небольших пределах (часто в активном эксперименте выбираются пределы  $-1 \dots +1$ ).

Постулируем модель следующего вида (коэффициенты при кодированных факторах будем обозначать строчными буквами  $b$ ):

$$\hat{Y} = b_0 + b_1 x_1 + \dots + b_n x_n + b_{12} x_1 x_2 + b_{13} x_1 x_3 + \dots + b_{n-1, n} x_{n-1} x_n + b_{123} x_1 x_2 x_3 + \dots + b_{12 \dots n} x_1 x_2 \dots x_n + b_{11} x_1^2 + \dots + b_{nn} x_n^2.$$

В модель входят нулевой член, члены вида  $b_i x_i$  (линейная регрессия), члены вида  $b_{ijk} x_i x_j x_k$ , отражающие взаимодействия факторов (неполная квадратичная регрессия), и, наконец, члены вида  $b_{ii} x_i^2$  (квадратичная регрессия). Очевидно, что при значительных  $n$  запись вида уравнения становится весьма громоздкой и неудобной. Не изменяя вида модели, запишем ее в более компактной форме

$$\hat{Y} = \sum_{i=0}^m b_i x_i,$$

где  $x_0 \equiv 1$ ;  $x_i$  при  $i > n$  обозначает произведения и квадраты факторов;  $b_i$  при  $i > n$  — коэффициенты при произведениях и квадратах факторов;  $m$  — полное число членов полинома, не считая нулевого (всего  $m+1$  член).

Переходим теперь к определению коэффициентов регрессии  $b_i$  по данным опытов. Как и в случае одномерной регрессии (см. лек. 6), в любом  $u$ -м опыте имеем

$$Y_u = \beta_0 x_{0u} + \beta_1 x_{1u} + \dots + \beta_m x_{mu} + \varepsilon_u,$$

где  $Y_u$  — значение  $Y$  в  $u$ -м опыте  $x_{0u}, x_{1u}, \dots, x_{mu}$  — значения независимых переменных  $x_0, \dots, x_m$ , в  $u$ -м опыте;  $\beta_0, \dots, \beta_m$  — «истинные» значения коэффициентов регрессии, оценки которых  $b_0, \dots, b_m$  входят в модель;  $\varepsilon_u$  — ошибка в  $u$ -м опыте.

Результаты всех  $N$  опытов запишем в виде матричного уравнения

$$Y = X\beta + \varepsilon$$

где



$$Y = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_u \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix}; \quad X = \begin{bmatrix} x_0 & x_{11} \dots x_{m1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot x_{0u} & x_{1u} \dots x_{mu} \\ \cdot \\ \cdot \\ \cdot \\ x_{0N} & x_{1N} \dots x_{mN} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \cdot \\ \cdot \\ \cdot \\ \beta_i \\ \cdot \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_u \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix}.$$

Размер вектора наблюдений, или матрицы-столбца  $Y$  и вектора ошибок, или матрицы-столбца  $\varepsilon$  —  $(N \times 1)$ ; размер вектора коэффициентов  $(m+1 \times 1)$ ; размер матрицы  $X$   $(N \times m+1)$ .

Условие МНК запишем в виде

$$S = (Y - X\beta)^2 \rightarrow \min.$$

Нормальные уравнения получим, продифференцировав по  $\beta$ :

$$\frac{dS}{d\beta} = -2x_i(Y - x\beta)$$

(дифференцирование выполняется по обычным правилам, но перед скобкой записывается  $x_i$  чтобы умножение имело смысл) и приравняв производную нулю:

$$X^T X B = X^T Y$$

Тогда

$$B = (X^T \cdot X)^{-1} X^T Y,$$

что и является матрицей МНК—оценок коэффициентов регрессии.

Обозначив  $X^T X = C$ , получим  $B = C^{-1} X^T Y$ .

Отметим, что в выражение для  $B$  входит обратная матрица  $C^{-1}$ . Исходная матрица  $C$  определена на основе матрицы  $X$ , отражающей значения независимых переменных в опытах. Если мы сами задаем независимые переменные в каждом опыте, т.е. проводим активный эксперимент, мы имеем возможность позаботиться о том, чтобы сконструировать  $C$ , а, следовательно, и  $C^{-1}$  некоторым наилучшим образом.

Познакомимся с «инженерным методом» формирования системы линейных уравнений для нахождения неизвестных коэффициентов многофакторной регрессии.

*Первое уравнение системы соответствует виду выбранного уравнения регрессии с заменами:*

$$B_0 \text{ на } B_{0N}; \quad X_k \text{ на сумму } X_k; \quad k=1..m; \quad Y \text{ на сумму } Y$$

*Каждое последующее уравнение системы получаем путем умножения первого на сумму  $X_k$ .*

Проиллюстрируем алгоритм на примере.

Пусть необходимо найти неизвестные коэффициенты регрессионного уравнения вида

$$Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3$$

В уравнении четыре неизвестных коэффициента. Следовательно, порядок СЛАУ равен 4. Применяя «эвристический алгоритм», получаем:

$$\left\{ \begin{array}{l} B_0 N + B_1 \sum_{i=1}^N X_{i,1} + B_2 \sum_{i=1}^N X_{i,2} + B_3 \sum_{i=1}^N X_{i,3} = \sum_{i=1}^N Y_i \\ B_0 \sum_{i=1}^N X_{i,1} + B_1 \sum_{i=1}^N X_{i,1}^2 + B_2 \sum_{i=1}^N X_{i,2} + B_3 \sum_{i=1}^N X_{i,3} = \sum_{i=1}^N (Y_i X_{i,1}) \\ B_0 \sum_{i=1}^N X_{i,2} + B_1 \sum_{i=1}^N X_{i,1} + B_2 \sum_{i=1}^N X_{i,2}^2 + B_3 \sum_{i=1}^N X_{i,3} = \sum_{i=1}^N (Y_i X_{i,2}) \\ B_0 \sum_{i=1}^N X_{i,3} + B_1 \sum_{i=1}^N X_{i,1} + B_2 \sum_{i=1}^N X_{i,2} + B_3 \sum_{i=1}^N X_{i,3}^2 = \sum_{i=1}^N (Y_i X_{i,3}) \end{array} \right.$$

Перейдем к третьему этапу регрессионного анализа и получим выражения для дисперсии оценок коэффициентов  $\mathbf{B}$ .

Введем в рассмотрение матрицу дисперсии коэффициентов:

$$DB = C^{-1} \cdot DY,$$

где  $DY$  – дисперсия наблюдений

В общем случае  $DB$  — квадратная матрица размером  $(1+m) \times (1+m)$ ; диагональные элементы — дисперсии соответствующих коэффициентов, недиагональные — ковариации, представляющие собою меру неопределенности, возникающей в силу зависимости друг от друга коэффициентов регрессии. Матрица  $DB$  может быть использована для оценки значимости коэффициентов регрессии.

$$t_i = \frac{B_i}{DB_{i,i}}, \quad i = 0, 1..m$$

где  $t_i$  – расчетный критерий Стьюдента в задаче о гипотезе равенства нулю  $i$  – го коэффициента модели.

*Проверка адекватности регрессии выполняется способом, описанным в лекции 6 для однофакторной регрессионной зависимости.*

## Лекция 24. Методика факторного эксперимента

Решение поставленной перед исследователем задачи методом прямого экспериментирования связано с глубоким проникновением в сущность изучаемого процесса.

Подготовка к экспериментальному исследованию начинается с тщательного изучения всех материалов, относящихся к цели и предмету исследования. Например, при исследовании какой-либо технологической операции необходимо выяснить конструктивные особенности обрабатываемой детали, технические требования к ней, характеристики применяемого оборудования, приспособлений и инструментов, режимы резания и др. условия обработки. Затем нужно попытаться дать научное предположение о характере протекающего процесса и вероятное его объяснение, иначе говоря, разработать рабочую гипотезу. Как минимум, рабочая гипотеза устанавливает факторы, обуславливающие развитие исследуемого процесса.

На высшем уровне предполагаемое развитие процесса представляется математическим выражением, описывающим, например, его физико-механические закономерности. Но во всех случаях гипотеза не в состоянии совершенно отобразить реальный процесс и поэтому нуждается в уточнениях, доработках, даже иногда в радикальной замене новой.

### Выбор модели исследуемого процесса

Каждый изучаемый процесс в конечном итоге представляют в виде некоторой модели, например, графиком или математической формулой. Главное свойство модели – это ее способность предсказывать характер протекания процесса при изменении действующих на него факторов.

Рассмотрим следующее выражение, представляющее собой аналитическую зависимость упругой деформации по отверстию детали при дорновании:

$$U = \frac{\sigma_s}{3E} a \frac{3 \frac{b^2}{a^2} + 1}{\frac{b^2}{a^2} - 1} \ln \frac{b^2}{a^2}, \quad (24.1)$$

где  $\sigma_s$  – величина сопротивления материала пластическому деформированию, зависящая также от натяга дорнования;  $2b$ ,  $2a$  – наружный и внутренний диаметры детали;  $E$  – модуль продольной упругости.

Представленная формула – это модель процесса изменения величины упругой деформации  $U$ . Не прибегая к физическим опытам, по ней можно установить искомое значение деформации при изменении натяга дорнования, размеров детали и ее материала.

Модель в данном случае выявляет не только главные факторы рассматриваемого процесса, но и степень влияния каждого из них на процесс в целом. Имея такую модель, исследователь легко осуществит детализацию методики последующего экспериментального исследования, с помощью которого можно установить погрешность формулы и уточнить область ее применения.

В технологии машиностроения многие процессы представляют в виде эмпирических формул, структура которых определяется на основе априорной информации. Например, при исследовании наклепа поверхностного слоя после точения получена следующая зависимость:

$$H_\mu = \frac{138,7 \cdot S^{0,02} t^{0,02} \alpha^{0,01} r^{0,05}}{V^{0,05} (90 + \gamma)^{0,3}} H / \text{мм}^2, \quad (24.2)$$

где  $H_\mu$  – микротвердость на поверхности детали;  $V$  – скорость резания;  $S$  – подача;  $t$  – глубина резания;  $\alpha$  – задний угол;  $\gamma$  – передний угол;  $r$  – радиус округления режущей кромки.

Применение приведенной формулы ограничено материалом детали (сталь 45) и инструмента (Т16К6), скоростью (71...282 м/мин.), подачей (0,05...0,43 мм/об) и т.д. Формулу (8.2.) нельзя использовать непосредственно для расчета величины  $H_\mu$  при обработке, например, закаленной стали эльборовыми резцами. Однако исследователь имеет все основания предположить, что структура модели в новом процессе не изменится и представить ее в виде:

$$H_\mu = \frac{\alpha \cdot S^x t^y \alpha^k r^p}{V^z (90 + \gamma)^f} \quad (24.3)$$

В приведенных выше примерах входящие в формулы факторы, определяющие ход процесса, могут принимать по желанию исследователя любые постоянные значения с заданной точностью и поддерживаться на заданном уровне в течение всего опыта. Такие факторы называются управляемыми, а эксперимент, в котором они участвуют – активным.

Очень важно в период подготовки к эксперименту все факторы, обуславливающие процесс, разделить на основные, оказывающие решающее влияние на развитие процесса, и дополнительные или второстепенные. Дополнительные факторы обычно стараются нейтрализовать в процессе эксперимента путем создания таких условий, при которых их влияние на другие факторы становится незначительным. Однако следует помнить, что деление факторов на основные и дополнительные в общем случае условно и может быть произведено только на основе достоверной априорной информации.

### ***Выбор метода исследования***

Возможность одновременного изучения влияния ряда основных переменных факторов на протекание одного процесса не всегда осуществима. В таких случаях для получения с достаточной степенью надежности четкой закономерности процесса следует выделить в качестве основного фактора лишь один, нейтрализовав все остальные. Затем, последовательно перебирая основные факторы при неизменности остальных, устанавливают общие закономерности развития исследуемого процесса. Такой метод исследования называется однофакторным экспериментом.

Достижение цели с помощью однофакторного эксперимента требует постановки многочисленных опытов, число которых возрастает с ростом количества факторов. Кроме того, возникает неуверенность в оценке искомого параметра, когда входящие в найденную формулу факторы принимают одновременно различные значения, отличающиеся от фиксированных в эксперименте. Преодолеть возникший барьер помогает метод многофакторного планирования экспериментальных исследований. Этот метод позволяет одновременно варьировать всеми факторами в нужном направлении по специальным правилам – алгоритмам. Задачи, для решения которых может использоваться многофакторное планирование эксперимента, чрезвычайно разнообразны. Это – поиск оптимальных условий проведения процесса, построение математических моделей, выбор существенных факторов и др. При этом предпочтение отдается моделям на основе степенных рядов - алгебраическим полиномам различной степени. Например, полином первой степени с факторами  $x_1, x_2, x_3$  и т.д. будет иметь вид:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n. \quad (24.4)$$

Нетрудно убедиться, что к такому виду легко приводится встречавшееся выше выражение (8.3) путем его логарифмирования.

В практике технологии машиностроения встречаются процессы, формирующиеся под воздействием случайных факторов. Так, обрабатываемые детали имеют переменные по величине припуски и различную твердость материала, что вызывает не поддающиеся контролю изменения сил резания и колебания упругих отжатий в системе СПИД. В результате размеры обрабатываемых деталей будут случайным образом отклоняться от настроечного размера, т.е. будут изготовлены с некоторой погрешностью.

Для выявления характера и степени воздействия случайных факторов на процесс прибегают к статистическим моделям. Вид статистической модели часто носит многовариантный характер и может быть установлен экспериментальным путем. Практически статистическую модель выбирают на основе априорной информации. Например, для изучения погрешностей линейных и угловых размеров, шероховатости поверхности, твердости, механических свойств материала, а также масс деталей часто применяют нормальный закон распределения (Гаусса). Изменения эксцентриситета, несоосности и биения, разностенности, непараллельности и неперпендикулярности, конусности изучают на основе закона эксцентриситета (Релея). Нередко прибегают к композиционным законам распределения, представляющим собой совокупность простых распределений двух и более независимых случайных величин.

В технологических экспериментах часто возникает необходимость определения связей между двумя или большим количеством факторов, причем в большинстве случаев эти связи являются не функциональными, а статистическими (вероятностными) и обнаруживаются лишь при

массовом изучении признаков. Наиболее простыми статистическими связями являются линейные, но встречаются и криволинейные (нелинейные). Статистические модели таких процессов выявляют методом корреляционного анализа.

С помощью корреляционного анализа определяют, например, как изменяется поле рассеяния размера детали при изменении припуска или твердости заготовки, изучают взаимосвязи между параметрами черновой и чистовой обработки и т.д.

Как видно из предыдущего, выбор модели процесса и метода исследования часто весьма сложен, а сама модель еще не гарантирует успех. Тем не менее, рассматриваемый этап является совершенно необходимым для определения всех последующих действий экспериментатора, т.к. от этого этапа зависят объем и качество выполненных исследовательских операций и эффективность конечного результата исследований.

На рассмотренном этапе ставится задача принять тот вариант модели и тот метод исследования, который предположительно может обеспечить наиболее быстрое и верное решение задачи. Для выполнения указанного требования в некоторых случаях бывает полезно включить в программу исследований специальные поисковые опыты, предшествующие разработке общей методики.

### ***Разработка методики экспериментальных исследований***

Вне зависимости от сложности экспериментальных исследований проведению экспериментов обязательно должна предшествовать разработка рабочей методики исследований и математической обработки полученных данных.

**Методикой экспериментальных исследований** называется детальный план исследований, устанавливающий совокупность способов, приемов исследования, их последовательность и средства осуществления исследований (оборудование, приспособления, инструмент, измерительную аппаратуру).

Методика, относящаяся ко всему исследованию, называется общей. Отдельно разрабатываемые рабочие методики для проведения, например, поисковых экспериментов или определенных условий эксперимента, носят название частных. Правильно разработанная методика исследования – залог его успешного осуществления. К разработке методики следует относиться самым серьезным образом во избежание получения ошибочных результатов или неподдающегося никакой обработке бесполезного материала.

В рабочей методике должны быть выявлены все элементарные подробности экспериментов. В ней указываются: - цель исследования; - последовательность проведения экспериментов; - подготовка к проведению экспериментов; - правила измерений и порядок регистрации результатов измерений; - характеристики и графики, получаемые в результате математической обработки наблюдений.

### ***Цель и задачи исследований.***

Цель исследования должна быть сформулирована очень четко и допускать количественную оценку. Цель исследования может быть дополнена формулированием отдельных задач, разделяющих содержание исследований на характерные этапы.

Основная цель производственных исследований – это обеспечение качества выпускаемой продукции при минимальных затратах на ее производство (иначе говоря, оптимизация производства). Задачами исследований являются:

- объективный выбор контролируемых параметров (факторов) технологического процесса;
- минимизация расхода материалов на единицу выпускаемой продукции при сохранении ее качества;
- повышение производительности труда за счет внедрения новой техники и технологии при неизменности качества выпускаемой продукции;
- повышение однородности, стабильности качества продукции, ее надежности в эксплуатации;
- увеличение надежности и быстродействия применяемых средств производства, управления и контроля и др.

**Пример 1.** Цель исследования микротвердости поверхности деталей после эльборового точения: выявить условия получения значений  $H_{\mu}$  в установленных чертежом пределах (указываются значения  $H_{\mu}$ ) при удовлетворительной производительности процесса обработки.

Задачи исследования:

- выбор контролируемых параметров (факторов);
- определение степени влияния факторов на процесс;
- обеспечение стабильности достигаемого качества обработки (по  $H_u$ ).

Пример 2. Цель исследования процесса дорнования отверстий трубчатых деталей многозубыми дорнами - определение условий получения отверстий 10 квалитета точности при производительности 1000-1200 деталей в смену.

Задачи исследования:

- выбор контролируемых параметров (факторов);
- оптимизация режимов обработки;
- выбор смазки и др.

Из приведенных примеров видно, что цель и задачи исследования могут быть установлены лишь после тщательного изучения всех материалов, относящихся к рассматриваемой теме.

### ***Объект, факторы процесса и средства исследований.***

На основе предложенной модели процесса необходимо установить:

- вид, форму и размеры объекта исследования;
- количественные значения факторов и уровни их варьирования, а также условия испытаний;
- средства исследования (оборудование, инструмент, измерительные устройства и требования к ним);
- количество подлежащих проведению испытаний.

Факторы исследуемого процесса те же, что и в разработанной модели. По величине факторы должны быть ограничены снизу и сверху в целесообразных границах. Например, скорость резания при точении не может быть равной нулю или бесконечности. При малых скоростях резания низка производительность обработки, при больших – мала стойкость режущего инструмента. Диапазон изменения скорости, следовательно, должен быть целесообразным. Следует иметь в виду, что расширение границ варьирования факторов приводит к существенному повышению затрат на проведение экспериментов, а слишком узкий диапазон – не позволит достигнуть цели исследования.

Условия испытаний определяются на основе выбранного метода исследований. Например, при однофакторном эксперименте следует строго оговорить значения нейтрализованных факторов в каждом опыте. При статистических исследованиях требуется обеспечить взятие выборок, удовлетворяющих условию случайности.

Особое значение приобретает вопрос назначения количества опытов. Например, если предлагаемая модель процесса представляет собой линейную зависимость, то, очевидно, можно ограничиться нахождением лишь двух точек со значительным промежутком между ними. Для построения окружности удовлетворяются тремя точками. При сложных кривых необходимо знать большое количество точек. Соответственно надо будет придавать такое же количество значений каждому исследуемому фактору. Далее, учитывая действие не включенных в исследование факторов, следует весь опыт повторить, т.е. провести серию опытов, количество которых часто принимается не менее 3-х. Это повышает объективность оценки координат каждой определяемой точки путем их усреднения.

### ***Подготовка к проведению исследований.***

В методике раскрываются условия изготовления и монтажа образцов, методы аттестации оборудования и контрольно-измерительных приборов, правила тарировки специальных средств измерений, условия проведения экспериментов, т.е. температура окружающей среды, влажность, технологические средства и способы их применения (например, СОЖ) и др.

В рабочей методике помещаются схемы измерения и их описание (расшифровка), если это требуется по условиям эксперимента. Например, указываются измерительные сечения (осевые и диаметральные) контролируемых после дорнования отверстий, положение и направление измерения площадки износа режущей грани инструмента и т.д. Затем определяется порядок проведения опытов и регистрации их результатов. Для этого в методике должны быть предусмотрены формы специальных таблиц, которые помещаются в журнале исследований.

В последней части методики помещаются указания о методах математической обработки результатов эксперимента, а также формы таблиц, графиков, получаемых в результате этой обработки.

### Методика полного факторного эксперимента (ПФЭ)

Ниже приводится краткое изложение полного факторного эксперимента (ПФЭ) типа  $2^k$ , в котором каждому из включенных в эксперимент факторов  $x^1, x^2, \dots, x^k$  (« $k$ » факторов) устанавливают только два уровня варьирования: верхний и нижний.

Поскольку факторы процесса неоднородны и имеют различную размерность, а числа, выражающие величины факторов, имеют различные порядки, их приводят к единой системе счисления путем перехода от действительных значений факторов к кодированным по формулам:

$$x_{i \text{ осн}} = \frac{x_{i \text{ max}} + x_{i \text{ min}}}{2}; \quad (24.5)$$

$$\Delta x_i = \frac{x_{i \text{ max}} - x_{i \text{ min}}}{2}; \quad (24.6)$$

$$\tilde{x}_i = \frac{x_i - x_{i \text{ осн}}}{\Delta x_i}, \quad (24.7)$$

где  $x_{i \text{ осн}}$  – основной уровень;  $x_{i \text{ max}}$  – верхний уровень;  $x_{i \text{ min}}$  – нижний уровень;  $i$  – номер фактора;  $\Delta x_i$  – интервал варьирования;  $\tilde{x}_i$  – кодированное значение фактора, определяемое для каждого фактора.

Так как число уровней для каждого фактора равно 2, значения  $x_i$  могут быть равны  $x_{i \text{ max}}$  или  $x_{i \text{ min}}$ . Кодовые значения факторов, рассчитанные по формуле (8.5), будут равны «+1» (верхний уровень), «-1» (нижний уровень) и «0» (основной уровень). При построении планов матриц планирования эксперимента цифры (единицы) обычно опускают и пишут только знаки «+» или «-», а вместо  $\tilde{x}_i \rightarrow x_i$ .

Для примера рассмотрим методику планирования ПФЭ для описания зависимости показателя стойкости концевых фрез от заднего угла  $\alpha$  ( $x_1$ ), переднего угла  $\gamma$  ( $x_2$ ) ширины ленточки  $f$  ( $x_3$ ). При трех факторах количество опытов в эксперименте  $N=2^3=8$ . Уровни факторов и интервалы варьирования записывают в таблицу типа 24.1., используя априорную информацию о влиянии факторов.

Таблица 24.1. Уровни факторов и интервалы варьирования

Уровни факторов	Обозначение	$\alpha^\circ$	$\gamma^\circ$	$f$ , мм
		$x_1$	$x_2$	$x_3$
Основной	0	14	15	0,05
Интервал варьирования	$\Delta x_i$	4	6	0,03
Верхний	+	18	21	0,08
Нижний	-	10	9	0,02

Далее строим стандартный план матрицы планирования эксперимента в виде табл. 8.2. В ней представлены все возможные сочетания уровней факторов и их взаимодействий, что достигается следующим образом. В первом столбце фиктивному фактору  $x_0$  придают одинаковые кодовые обозначения «+».

В столбце для переменного фактора  $x_1$  знаки чередуют по одному, для  $x_2$  – по два и для  $x_3$  – по четыре. Аналогично можно построить планы матриц для четырех (чередуют знак по восемь), пяти (по шестнадцати), десяти и т.д. факторов. Кодовые значения произведений факторов определяют путем алгебраического перемножения кодовых значений сомножителей. План эксперимента выделяют рамкой в табл. 24.2., остальные данные используют для расчета коэффициентов модели.

Таблица 24.2. Полный план факторного эксперимента  $2^3$

Номер опыта	Значения факторов в кодовых обозначениях				Комбинация произведений факторов в кодовых обозначениях				Действительное значение показателя стойкости фрез			
	$x_0$	$x_1$	$x_2$	$x_3$	$x_1x_2$	$x_1x_3$	$x_2x_3$	$x_1x_2x_3$	$Y_1$	$Y_2$	$Y_3$	$\bar{Y}$
1	+	-	-	-	+	+	+	-	41,0	34,85	34,85	36,90
2	+	+	-	-	-	-	+	+	43,85	47,45	40,90	44,07
3	+	-	+	-	-	+	-	+	43,30	32,00	29,25	34,83
4	+	+	+	-	+	-	-	-	48,85	58,50	50,50	52,62
5	+	-	-	+	+	-	-	+	26,70	15,38	12,25	18,11
6	+	+	-	+	-	+	-	-	31,52	24,35	36,30	30,72
7	+	-	+	+	-	-	+	-	17,32	30,85	28,70	25,62
8	+	+	+	+	+	+	+	+	30,75	29,50	38,15	32,80

После построения плана матрицы планирования проверяют ее свойства:

– симметричность относительно центра эксперимента: алгебраическая сумма элементов столбца каждого фактора должна быть равна нулю (кроме  $x_0$ ), т.е.

$$\sum_{v=1}^N x_{iv} = 0 \quad (24.8);$$

– нормировку: сумма квадратов элементов каждого столбца равна числу опытов, или

$$\sum_{v=1}^N x_{iv}^2 = N \quad (24.9);$$

– ортогональность: сумма почленных произведений любых двух столбцов матрицы равна нулю, т.е.

$$\sum_{v=1}^N x_{iv} x_{jv} = 0, \quad j \neq i, \quad j = 0, 1, 2 \dots K \quad (24.10)$$

Если разработанный план матрицы отвечает указанным свойствам, можно приступить непосредственно к проведению экспериментов. Порядок проведения опытов в случайной последовательности для каждой серии опытов целесообразно установить с помощью таблиц равномерного распределения случайных чисел. Например, в рассматриваемом случае опыты следует проводить так:

Таблица 24.3. Рандомизация опытов в эксперименте

Номер опыта	1	2	3	4	5	6	7	8	№ серии (повтор)
	Строка плана								
Порядок реализации опытов	1	6	5	8	2	7	4	3	I
	3	8	7	2	1	4	5	6	II
	7	1	4	2	5	8	3	4	III

Эта операция называется **рандомизацией** и применяется с целью исключения систематических ошибок, вызванных внешними условиями (переменной температуры, сырья, оператора и т.д.).

После каждой серии опытов получают  $N$  значений искомого параметра  $Y_1, Y_2, Y_3$  и рассчитывают среднее значение параметра  $\bar{Y}$  для каждой исследуемой точки плана по формуле:

$$\bar{Y}_k = \frac{1}{m} \sum_{j=1}^m Y_{k,j}, \quad (24.11)$$

где  $Y_{j,k}$  – результат  $k, j$  опыта матрицы планирования;  $m$  – число параллельных наблюдений в каждой точке (в примере  $m = 3, j=1..3$ );  $k$  – номер опыта (в примере  $k = 1..8$ ).



Для оценки отклонения показателя параметра от среднего значения следует вычислить дисперсию воспроизводимости по данным «к» параллельных наблюдений плана матрицы планирования в каждой точке по формуле:

$$S_k^2 = \frac{\sum_{j=1}^m (\bar{Y}_k - Y_{k,j})^2}{m-1}. \quad (24.12)$$

Например, в точке 1 плана (строка 1) получим:

$$S_1^2 = \frac{(36,90 - 41,0)^2 + (36,90 - 34,85)^2 + (36,90 - 34,85)^2}{3-1} = 12,60.$$

Аналогично находят значения дисперсий в остальных точках:

$$S_2^2 = 10,76; \quad S_3^2 = 55,44; \quad S_4^2 = 27,41; \quad S_5^2 = 32,36;$$

$$S_6^2 = 36,18; \quad S_7^2 = 52,89; \quad S_8^2 = 21,36$$

Далее определяют критерий Кохрена  $G$ , по которому проверяется гипотеза однородности дисперсий:

$$G = \frac{\max(S^2)}{\sum_{k=1}^N S_k^2}. \quad (24.13)$$

или, для рассматриваемого примера,

$$G = \frac{55,44}{12,60 + 10,76 + 55,44 + 27,41 + 32,36 + 36,18 + 52,89 + 21,36} = 0,22.$$

Полученное значение « $G$ » сравнивают с табличным  $G_{кр}$  (табл.4.4.), определяемым на пересечении столбца  $N_{16\max} = m-1$  и строки  $N_{26} = N$ . В рассматриваемом примере для  $m-1=2$  и  $N=8$  находим  $G_{кр} = 0,5157$ . Если окажется, что расчетное значение « $G$ » меньше табличного  $G_{кр}$ , то гипотеза об однородности дисперсий и воспроизводимости результатов принимается. В противном случае необходимо увеличить число параллельных опытов.

Если дисперсии однородны, их следует усреднить и найти дисперсию исследуемого параметра по формуле:

$$S^2 = \frac{\sum_{k=1}^N S_k^2}{N} = \frac{249}{8} = 31,125, \quad (24.14)$$

где  $S^2$  – средняя арифметическая дисперсий всех различных точек плана матрицы или дисперсия искомого параметра.

Параметры модели исследуемого процесса или коэффициенты регрессии определяют по единой формуле:

$$b_i = \frac{\sum_{v=1}^N x_{iv} \bar{Y}_v}{N}, \quad (24.15)$$

где  $b_i$  – коэффициент регрессии  $i = 1, 2 \dots N$

Произведем расчет коэффициентов регрессии для рассматриваемого примера.

$$\begin{aligned} b_0 &= \frac{x_{01} \bar{Y}_1 + x_{02} \bar{Y}_2 + \dots + x_{08} \bar{Y}_8}{8} = \\ &= \frac{(1 \cdot 36,90 + 1 \cdot 44,07 + 1 \cdot 34,83 + 1 \cdot 52,62 + 1 \cdot 18,11 + 1 \cdot 30,72 + 1 \cdot 25,62 + 1 \cdot 32,80)}{8} = 34,46 \\ b_1 &= \frac{x_{11} \bar{Y}_1 + x_{12} \bar{Y}_2 + \dots + x_{18} \bar{Y}_8}{8} = \end{aligned}$$

$$= \frac{(-1) \cdot 36,90 + 1 \cdot 44,07 + (-1) \cdot 34,83 + 1 \cdot 52,62 + (-1) \cdot 18,11 + 1 \cdot 30,72 + (-1) \cdot 25,62 + 1 \cdot 32,80}{8} = 5,59$$

Таким путем находим все восемь искомым коэффициентов:  
 $b_2 = 243$ ;  $b_3 = -7,64$ ;  $b_{12} = 0,65$ ;  $b_{13} = -0,65$ ;  $b_{23} = 0,39$ ;  $b_{123} = -0,5$ .

Искомое уравнение в преобразованных переменных  $x_i$  будет:

$$Y = 34,46 + 5,59x_1 + 2,43x_2 - 7,64x_3 + 0,65x_1x_2 - 0,65x_1x_3 + 0,39x_2x_3 - 0,5x_1x_2x_3.$$

Далее следует оценить значимость коэффициентов регрессии по  $t$  – критерию Стьюдента. Для этого сначала определяют дисперсию и среднеквадратическое отклонение коэффициентов регрессии по формулам:

$$S_b^2 = \frac{S^2}{Nm} = \frac{31,125}{8 \cdot 3} = 1,296. \quad (24.16)$$

$$S_b = \sqrt{\frac{S^2}{Nm}} = \sqrt{1,296} = 1,13. \quad (24.17)$$

Затем рассчитывают значения  $t_i$  – критерия:

$$t_i = \frac{|b_i|}{S_b}. \quad (24.18)$$

Для рассматриваемого примера получим:

$$t_0 = \frac{34,47}{1,13} = 30,49; \quad t_1 = \frac{5,59}{1,13} = 4,95; \quad t_2 = \frac{2,43}{1,13} = 2,15; \quad t_3 = \frac{7,64}{1,13} = 6,76;$$

$$t_{12} = \frac{0,65}{1,13} = 0,57; \quad t_{13} = \frac{0,65}{1,13} = 0,57; \quad t_{23} = \frac{0,39}{1,13} = 0,34; \quad t_{123} = \frac{0,5}{1,13} = 0,44.$$

Критическое значение  $t_{кр}$  определяют по табл.4.1. в зависимости от числа степеней свободы  $N_{зн} = N(m-1) = 8 \cdot (3-1) = 16$  при заданном уровне значимости  $q = 5\%$ :  $t_{кр} = 2,1190$ .

Если  $t_i > t_{кр}$ , то коэффициент  $b_i$  признается значимым. Согласно расчету, значимыми являются коэффициенты  $b_0$ ,  $b_1$ ,  $b_2$  и  $b_3$ . Остальные коэффициенты  $b_{12}$ ,  $b_{13}$ ,  $b_{23}$  и  $b_{123}$  статически незначимы и могут быть отброшены без пересчета всех остальных.

В математическую модель процесса или объекта включают только значимые коэффициенты. Для нашего примера получим

$$Y = 34,47 + 5,59x_1 + 2,43x_2 - 7,64x_3.$$

Наконец, чтобы получить модель в натуральных переменных, надо в полученное уравнение подставить выражения  $x_i$  из формулы преобразования (8.5).

Принимая во внимание табл. 8.1., для рассматриваемого примера получим:

$$x_1 = \frac{\alpha - 14}{4}; \quad x_2 = \frac{\gamma - 15}{6}; \quad x_3 = \frac{f - 0,5}{0,03},$$

Следовательно

$$Y(\alpha, \gamma, f) = 33,7 + 1,4\alpha + 0,4\gamma - 255f.$$

После вычисления коэффициентов следует проверить пригодность модели или ее адекватность. Для этого вначале находим оценку дисперсии адекватности модели  $S_{ад}^2$  по формуле:

$$S_{ад}^2 = \frac{m}{N-l} \sum_{i=1}^N (\bar{Y} - Y(\alpha_i, \gamma_i, f_i))^2, \quad (24.19)$$

где  $l$  – число значимых коэффициентов (в нашем случае – это  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , т.е.  $l = 4$ );  $Y(\alpha_i, \gamma_i, f_i)$  – математическое ожидание искомого параметра, подсчитанное по уравнению регрессии. Его определяют для каждой строчки плана матрицы, с учетом знака фактора, как алгебраическую сумму коэффициентов уравнения. Так, для первой строки

$$Y(\alpha_1, \gamma_1, f_1) = 34,47 + (-1) \cdot 5,59 + (-1) \cdot 2,43 - (-1) \cdot 7,64 = 34,09.$$

Аналогично получим:

$$Y(\alpha_2, \gamma_2, f_2) = 45,27; \quad Y(\alpha_3, \gamma_3, f_3) = 38,95; \quad Y(\alpha_4, \gamma_4, f_4) = 50,13; \quad Y(\alpha_5, \gamma_5, f_5) = 18,81;$$

$$Y(\alpha_6, \gamma_6, f_6) = 29,99; \quad Y(\alpha_7, \gamma_7, f_7) = 23,67; \quad Y(\alpha_8, \gamma_8, f_8) = 34,85.$$

Тогда

$$S_{ad}^2 = \frac{3}{8-4} \left[ (36,90 - 34,09)^2 + (44,07 - 45,27)^2 + (34,83 - 38,95)^2 + (52,62 - 50,13)^2 + (18,11 - 18,81)^2 + (30,72 - 29,99)^2 + (25,62 - 23,67)^2 + (32,80 - 34,85)^2 \right] = 45,135.$$

Адекватность модели проверяют по формуле:

$$F = \frac{S_{ad}^2}{S^2} = \frac{45,135}{31,125} = 1,45, \quad (24.20)$$

где  $F$  – критерий Фишера, табличное значение которого  $F_{кр}$  определяют по табл. 4.3 в зависимости от числа степеней свободы  $N_{1ad} = N - l$  и  $N_{2ad} = N(m - 1)$ .

Для рассматриваемого примера при  $N_{1ad} = 4$ ,  $N_{2ad} = 16$  и уровне значимости  $q = 5\%$ ,  $F_{кр} = 3,01$ .

Если расчетное значение критерия  $F$  окажется меньше значения  $F_{кр}$  (в примере  $1,45 < 3,01$ ), то гипотеза адекватности модели принимается.

### АНАЛИЗ МОДЕЛИ

Полученная модель может быть использована только в заданных табл. 24.1. интервалах варьирования факторов:  $\alpha = 10 \dots 18^\circ$ ,  $\gamma = 9 \dots 21^\circ$  и  $f = 0,02 \dots 0,08$  мм. В указанных пределах увеличение  $\alpha$  и  $\gamma$  приводит к повышению стойкости инструмента. Влияние угла  $\alpha$  существенно выше, что видно из сопоставления соответствующих коэффициентов. Увеличение ширины ленточки  $f$ , наоборот, понижает стойкость, о чем говорит знак коэффициента. Влияние ширины ленточки – значительное.

После проведения многофакторного эксперимента и определения модели может быть поставлена задача оптимизации факторов на основе метода, например, крутого восхождения.

### Дробный факторный эксперимент (ДФЭ)

В рассмотренном выше примере все эффекты взаимодействия оказались незначимы, благодаря чему была получена линейная модель. Если бы это обстоятельство было известно априори, число опытов можно было бы сократить. Для этого достаточно было бы фактору присвоить вектор-столбец матрицы, принадлежащий взаимодействию факторов  $x_1x_2$  в плане  $N = 2^2$  (табл. 24.4).

Таблица 24.4 Полный факторный эксперимент  $2^2$

№ опытов	$X_0$	$X_1$	$X_2$	$X_3$ ( $x_1x_2$ )	$Y_1$	$Y_2$	$Y_3$	$\bar{Y}$
1	+	-	-	+				
2	+	+	-	-				
3	+	-	+	-				
4	+	+	+	+				

Поставив  $x_3$  на место  $x_1x_2$ , мы сократили количество опытов ПФЭ  $2^3$  с 8 до 4-х, т.е. воспользовались половиной ПФЭ  $2^3$  или полуреplikой. С увеличением количества факторов в эксперименте дробность реплик может возрастать, а количество опытов соответственно уменьшаться по сравнению с ПФЭ (табл. 24.5).

Таблица 24.5. Условные обозначения дробных реплик и количество опытов

Кол-во факторов «К»	Дробная реплика	Условное обозначение $2^{K-C}$	Кол-во опытов для ДФЭ	Кол-во опытов для ПФЭ
3	1/2 от $2^3$	$2^{3-1}$	4	8
4	1/2 от $2^4$	$2^{4-1}$	8	16
5	1/4 от $2^5$	$2^{5-2}$	8	32
6	1/8 от $2^6$	$2^{6-3}$	8	64

7	1/16 от 2 <sup>7</sup>	2 <sup>7-4</sup>	8	128
5	1/2 от 2 <sup>5</sup>	2 <sup>5-1</sup>	16	32
6	1/4 от 2 <sup>6</sup>	2 <sup>6-2</sup>	16	64
7	1/8 от 2 <sup>7</sup>	2 <sup>7-3</sup>	16	128
8	1/16 от 2 <sup>8</sup>	2 <sup>8-4</sup>	16	256
9	1/32 от 2 <sup>9</sup>	2 <sup>9-5</sup>	16	512
10	1/64 от 2 <sup>10</sup>	2 <sup>10-6</sup>	16	1025
11	1/128 от 2 <sup>11</sup>	2 <sup>11-7</sup>	16	2048
12	1/256 от 2 <sup>12</sup>	2 <sup>12-8</sup>	16	4096
13	1/512 от 2 <sup>13</sup>	2 <sup>13-9</sup>	16	8192
14	1/1024 от 2 <sup>14</sup>	2 <sup>14-10</sup>	16	16384
15	1/2048 от 2 <sup>15</sup>	2 <sup>15-11</sup>	16	32768

$C$  – количество линейных эффектов (факторов), приравненных к эффектам взаимодействия.

Рассмотрим методику построения планов ДФЭ для изучения зависимости микротвердости поверхности при эльборовом обтачивании закаленной детали от подачи  $S$ , глубины резания  $t$ , заднего угла  $\alpha$ , радиуса при вершине резца  $r$ , скорости резания  $V$  и переднего угла  $\gamma$ . Для получения модели, сходной по структуре с формулой (24.3), факторы обозначают следующим образом:

$$x_1 = \ln S; \quad x_2 = \ln t; \quad x_3 = \ln \alpha; \quad x_4 = \ln r; \quad x_5 = \ln V; \quad x_6 = \ln(90 + \gamma).$$

После кодирования факторов все их возможные сочетания, в том числе взаимодействия всех порядков, можно представить следующим рядом:

$x_1; x_2; x_3; x_4; x_5; x_6;$   
 $x_1x_2; x_1x_3; x_1x_4; x_1x_5; x_1x_6; x_2x_3; x_2x_4; x_2x_5; x_2x_6;$   
 $x_3x_4; x_3x_5; x_3x_6; x_4x_5; x_4x_6; x_5x_6;$   
 $x_1x_2x_3; x_1x_2x_4; x_1x_2x_5; x_1x_2x_6; x_1x_3x_4; x_1x_3x_5; x_1x_3x_6;$   
 $x_1x_4x_5; x_1x_4x_6; x_1x_5x_6; x_2x_3x_4; x_2x_3x_5; x_2x_3x_6; x_2x_4x_5;$   
 $x_2x_4x_6; x_2x_5x_6; x_3x_4x_5; x_3x_4x_6; x_3x_5x_6; x_4x_5x_6;$   
 $x_1x_2x_3x_4; x_1x_2x_3x_5; x_1x_2x_3x_6; x_1x_2x_4x_5; x_1x_2x_4x_6; x_1x_2x_5x_6;$   
 $x_2x_3x_5x_6; x_2x_4x_5x_6; x_3x_4x_5x_6;$   
 $x_1x_2x_3x_4x_5; x_1x_2x_3x_4x_6; x_1x_2x_3x_5x_6; x_1x_2x_4x_5x_6;$   
 $x_1x_3x_4x_5x_6; x_2x_3x_4x_5x_6;$   
 $x_1x_2x_3x_4x_5x_6.$

Следовательно, полный факторный эксперимент типа 2<sup>6</sup> потребовал бы проведения не менее 64 опытов. Если априори известно, что все взаимодействия, начиная с парных и выше, незначимы, то объем эксперимента резко уменьшится. Действительно, вместо ПФЭ 2<sup>6</sup> можно поставить ДФЭ 2<sup>3</sup>, записав на местах взаимодействий  $x_1x_2$ ,  $x_1x_3$  и  $x_1x_2x_3$  в табл. 24.2 факторы  $x_4$ ,  $x_5$  и  $x_6$ . Мы получим дробный факторный эксперимент типа 2<sup>6-3</sup>, являющийся 1/8 реплики от 2<sup>6</sup>. Весь дальнейший ход эксперимента и математической обработки его результатов не отличается от ПФЭ.