

# **«Теория принятия решений»**

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 1*

### **Основные понятия теории принятия решений, исследования операций и системного анализа**

#### СО Д Е Р Ж А Н И Е

##### ВВЕДЕНИЕ

##### УЧЕБНЫЕ ВОПРОСЫ:

1. Введение в теорию принятия решений.
2. Общая модель и участники процесса принятия решения.
3. Задача оптимизации решений.
4. Классы задач принятия решений.
5. Математические модели и методы принятия решений как основные компоненты исследования операций.

##### ЗАКЛЮЧЕНИЕ

## Литература:

1. Менеджмент / Под ред. Ж.В. Прокофьевой. - М.: Знание, 2000.
2. Моисеев Н.Н. Математические задачи системного анализа. – М.: Наука, 1981.
3. Науман Э. Принять решение, но как? - М.: Мир, 1987.
4. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979.
5. Карминский А.М., Оленев Н.И., Примак А.Г., Фалько С.Г. Контроллинг в бизнесе. Методологические и практические основы построения контроллинга в организациях. - М.: Финансы и статистика, 1998.
6. Хан Д. Планирование и контроль: концепция контроллинга / Пер. с нем. - М.: Финансы и статистика, 1997.
7. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.

## Введение в теорию принятия решений

**Теория принятия решений** — комплексная научная дисциплина, направленная на разработку методов и средств, помогающих одному или нескольким лицам сделать обоснованный выбор наилучшего из имеющихся вариантов.

Слово «решение»:

1. Совокупность рассматриваемых возможностей, которые тем или иным образом выделены человеком, делающим выбор.
2. Процесс поиска наиболее предпочтительных вариантов, включающий в себя обдумывание, изучение какого-либо вопроса или задачи, нахождение правильного ответа.
3. Ответ, например один или несколько выбранных вариантов, результат анализа некоторой проблемы или математической задачи.

Задачи принятия решений часто отождествляются с задачами выбора.

*Основное назначение теории принятия решений состоит в разработке методов и средств, позволяющих одному человеку или группе лиц сформулировать множество возможных вариантов решения проблемы, сравнить их между собой, найти среди них лучшие или допустимые варианты, которые удовлетворяют тем или иным требованиям (ограничениям), и при необходимости объяснить сделанный выбор.*

*Лицо, принимающее решение (ЛПР) или действующее лицо — человек или группа людей, которые фактически осуществляют выбор предпочтительного решения.*

Обычно в роли ЛПР выступает руководитель или группа компетентных в своей области специалистов, обладающих соответствующими знаниями и опытом деятельности, наделенных необходимыми полномочиями для принятия решения и несущих ответственность за реализацию принятого решения.

*Владелец проблемы (ВП)* — человек или группа лиц, имеющих основания и мотивы для постановки проблемы, осознающих необходимость ее решения, инициирующих тем или иным образом принятие и выполнение нужного решения.

*Активные группы (АГ)* — объединяют людей, которые имеют общие интересы по отношению к проблеме, требующей решения, и стремятся оказать влияние на процесс выбора с тем, чтобы добиться нужного им результата. Активные группы — окружение, в котором протекает процесс решения проблемы и действует ЛПР.

Обычно владелец проблемы принадлежит к одной из основных активных групп. В сложных ситуациях выбора на разных этапах могут привлекаться эксперты (Э) и консультанты по принятию решений (К).

Эксперты — компетентные специалисты, профессионально разбирающиеся в решаемой проблеме, обладающие необходимой информацией о проблеме и отдельных ее аспектах, но не несущие ответственности за принятое решение и его реализацию.

Консультанты по принятию решений оказывают помощь ЛПР и владельцу проблемы в организации процесса ее решения, в правильной постановке задачи принятия решения, обеспечивают сбор необходимой информации, разрабатывают модель проблемы, процедуры и методы принятия решения.



*Жизненный цикл решения проблемы - общая модель принятия решения.*

## Введение в системный анализ и исследование операций

*«Системный анализ — это дисциплина, занимающаяся проблемами принятия решений в условиях, когда выбор альтернативы требует анализа сложной информации различной физической природы» (Н.Н.Моисеев) [1].*

Общим в определении понятия «система» является то, что о системе говорят как о множестве, между элементами которого имеются связи и которое удовлетворяет основному ее свойству — *целостности*.

*Исследование операций* — как отрасль знаний, объединяющая теории математических моделей и методов принятия решений предполагает наличие следующих элементов.:

1. Некоторого **процесса (модели операции, как системы действий, объединенных единым замыслом и направленных на достижение определенной цели)**
2. **Цели** проведения операции
3. **Множества решений (управляющих воздействий)**
4. **Метода** отыскания лучшего (оптимального) решения, при котором достигается цель

## **Традиционная схема исследования операции:**

1. Строится математическая модель явления или процесса.
2. Полученный математический объект изучается чисто математическими методами.
3. Математические результаты интерпретируются на реальные практические проблемы.

## **Классификация задач исследования операций**

### ***По зависимости параметров задачи от времени***

1. **Статическая задача.** Принятие решения происходит при условии, что все параметры известны и не изменяются во времени. Процедура принятия решения осуществляется один раз.
2. **Динамическая задача.** В процессе принятия решения параметры задачи изменяются во времени. Процедура принятия решения осуществляется поэтапно и может быть представлена в виде процесса, зависящего от времени, в том числе непрерывно. Пример — навигационная задача.



## *В зависимости от достоверности информации о задаче*

1. **Детерминированная задача.** Все параметры задачи заранее известны. Для решения детерминированных задач в основном применяются методы математического программирования.

2. **Недетерминированная задача.** Не все параметры задачи заранее известны. Оптимальное решение недетерминированной задачи отыскать практически невозможно. Но, некоторое «разумное» решение отыскать можно.

2.1. **Стохастическая задача.** Не все параметры задачи заранее известны, но имеются статистические данные о неизвестных параметрах, (вероятности, функции распределения, математические ожидания и т. д.).

Для отыскания оптимального решения стохастической задачи применяется один из следующих приемов:

- искусственное сведение к детерминированной задаче (неизвестные значения параметров заменяются их средними значениями),
- «оптимизация в среднем» (вводится и оптимизируется некоторый статистический критерий).

2.2. **Задача в условиях полной неопределенности.** Статистические данные о неизвестных параметрах отсутствуют. Такие задачи в основном изучаются в теории игр.

## ***По виду критерия оптимальности***

Критерий оптимальности может иметь любой вид, в том числе не формализуемый.

*Наиболее распространенные формализуемые критерии оптимальности заключаются в оптимизации (минимизации или максимизации) одной либо несколько скалярных целевых функций.*

Функция называется скалярной, если ее значением является некоторое число.

Задача оптимизации скалярной функции на заданном множестве допустимых числовых решений называется задачей математического программирования.

## **Однокритериальные задачи (задачи с одной целевой функцией)**

- *Задача линейного программирования.* Целевая функция — линейная, множество допустимых решений — плоский или выпуклый многогранник.
- *Задача квадратичного программирования.* Целевая функция — квадратичная

$$\sum_{i=1}^n \sum_{j=1}^m c_{i,j} x_i x_j, \text{ множество допустимых решений — выпуклый многогранник.}$$

- *Задача стохастического программирования.* Это подобные линейным задачи с неизвестными числовыми параметрами, о которых имеются статистические данные.
- *Задача дискретного программирования.* Множество допустимых решений — дискретное множество.
- *Задача целочисленного программирования.* Множество допустимых решений — точки целочисленной решетки.
- *Задача булева программирования.* Множество допустимых решений — матрицы из 0 и 1.

## Задача оптимизации решений

### *Задача принятия решения*

Пара  $(\Omega, P)$ ,

где  $\Omega$  – множество вариантов,  
 $P$  – принцип оптимальности.

*Решением задачи* является множество  $\Omega_P \in \Omega$ , получающееся в соответствии с принципом оптимальности  $P$ .

*Основная задача теории оптимальных решений состоит в представлении обоснованных количественных данных и рекомендаций для принятия оптимальных решений. Отсутствие хотя бы одного из элементов  $(\Omega, P)$  лишает задачу смысла.*

Математическим выражением принципа оптимальности  $P$  служит функция выбора  $C_p$ , которая со всеми подмножествами  $X \in \Omega$  составляет его часть  $C_p(X)$ , из которых и находится оптимальное решение исходной задачи  $C_{p, \text{оптим}}(\Omega)$ .

## Классы задач принятия решений

### *По полноте исходных данных*

Задачи принятия решения различаются в зависимости от информации о множестве  $\Omega$  и принципе оптимальности  $P$ :

1. *Общая задача принятия решения*:  $\Omega$ ,  $P$  – неизвестны, необходимо получить  $\Omega_P$  в процессе самого решения.
2. Задача с известным, но ограниченным множеством  $\Omega$  называется *задачей выбора решения*.
3. Задача, в которой  $\Omega$ ,  $P$  – известны называется *задачей оптимизации решений*.

## *По времени и срокам*

1. *Долгосрочное стратегическое планирования:*  
задачи размещения объектов производства в телекоммуникациях — узлов доступа разного уровня, развитие региональной телекоммуникационной сети.
2. *Среднесрочное планирование:*  
задачи маршрутизации трафика, задачи планирования инвестиционного строительства (с ограниченными ресурсами), задачи планирования текущего бюджета.
3. *Оперативное управление:*  
задачи текущего ремонта и устранения повреждений на сети, задачи распределения материалов на текущие нужды — установки и ремонт, транспортные задачи.

## ***По типу постановки***

### Распределительная задача

Имеем

$n$  — число предприятий;

$Y$  — количество единиц некоторого ресурса;

$f_k(x)$  — количество продукции, которое будет произведено на  $k$ -м предприятии, если в него будет вложено  $x$  единиц ресурса (монотонно неубывающая функция).

Требуется: оптимизировать число единиц ресурса  $x_i$ , выделяемого каждому предприятию, с целью достижения максимума объема выпускаемой ими продукции

$$f_1(x_1) + \dots + f_n(x_n) \rightarrow \max \quad (1)$$

$$x_1 + \dots + x_n \leq Y \quad (2)$$

$$x_i \geq 0, \text{ целые, } i = 1, \dots, n. \quad (3)$$

### *Оптимальное использование ресурсов*

Предприятие выпускает  $n$  видов изделий. Для их производства используются  $m$  видов ресурсов (разное сырье, людские ресурсы, финансовые ресурсы и т. п.). Эти ресурсы ограничены, их запасы составляют в планируемый период  $b_1, b_2, \dots, b_m$  условных единиц. Известны также технологические коэффициенты:  $a_{ij}$  – сколько единиц  $i$ -го ресурса требуется в производстве единицы  $j$ -го вида продукции.  $c_j$  – прибыль от реализации единицы  $j$ -го вида продукции. Требуется составить такой план выпуска продукции ( $x_1^*$  – единицы изделий первого вида,  $x_2^*$  – единицы изделий второго вида и т. д.), при котором прибыль предприятия была бы наибольшей.

Пусть будем выпускать план ( ), тогда  $a_{12}x_2$  – единиц 1-го ресурса пойдёт на всю программу выпуска 2-го вида продукции. Аналогично с другими видами продукции, поэтому общий расход  $a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n$  1-го ресурса не должен превышать его запас. Записав это для всех ресурсов, имеем ( ). Суммарная прибыль –  $\vec{c}^T \vec{x}$ , в результате получаем задачу линейного программирования ( ).



$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\
 &\dots\dots\dots \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m
 \end{aligned}$$

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_m \end{pmatrix}$$

$$\max F(x) = \sum_{j=1}^n c_j x_j = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$



*Определение:* в двойственной задаче ищется противоположный оптимум ( $\min \Leftrightarrow \max$ ) целевой функции. Коэффициентами целевой функции являются свободные члены системы ограничений прямой задачи. А свободными членами системы ограничений являются бывшие коэффициенты целевой функции. Матрица коэффициентов меняется на транспонированную, неравенства в системе ограничений – на противоположные.

Таким образом, прямая задача есть двойственная задача для своей двойственной задачи, т. е. они взаимно двойственные.

### *Теорема*

Прямая и двойственная задачи либо обе имеют оптимальные точки  $\vec{x}^*, \vec{y}^*$ , причём  $f_{\max} = \vec{c}^T \vec{x}^* = \vec{b}^T \vec{y}^* = \varphi_{\min}$ , либо обе не имеют решения.

Для оптимальных точек выполняются следующие равенства, называемые условиями *дополняющей не жёсткости*:

$$\begin{cases} x_j^* (A^T \vec{y}^* - \vec{c})_j = 0, j = \overline{1, n} \\ y_i^* (A \vec{x}^* - \vec{b})_i = 0, i = \overline{1, m} \end{cases} .$$

## Задача о рюкзаке

### Общая модель динамической оптимизации размещения капитала

Дано:  $P_1, P_2, \dots, P_N$  — проекты;

$T$  — горизонт планирования (длина наиболее продолжительного проекта);

$s_{tk}$  — доход от проекта  $P_k$  к концу года  $t$ ;

$y_{tk}$  — инвестиции в проект  $P_k$  в начале года  $t$ ;  $s_{0k} = y_{T+1k} = 0$ ;

$r$  — коэффициент дисконтирования затрат

$b_k = \sum_{t=0}^T (s_{tk} - y_{t+1,k}) / (1+r)^t$  — суммарная прибыль от проекта  $P_k$ ;

$C = (c_1, \dots, c_T)$  — доступный капитал для развития проектов

$A_k = (a_{1k}, \dots, a_{Tk})$  — вектор затрат на реализацию проекта  $P_k$  (целые);

Если доход нельзя реинвестировать, то  $a_{tk} = y_{tk}$ , иначе  $a_{tk} = y_{tk} - s_{t-1k}$ .

*Найти* подмножество проектов, которые можно реализовать на капитал  $C$  и которые в сумме дают максимальную прибыль, то есть

$$\max \sum_{k=1}^N b_k x_k$$

при ограничениях

$$\sum_{k=1}^N a_{tk} x_k \leq c_t, \quad t = 1, \dots, T$$

$$x_k \in \{0, 1\}, \quad k = 1, \dots, N$$

*Замечание 1.* При  $T = 1$  получаем линейную распределительную задачу с 0-1 переменными — задачу о рюкзаке.

## Задача об отправке грузов

$I = \{1, \dots, n\}$  — авиалайнеры,  $J = \{1, \dots, m\}$  — контейнеры,

$p_{ij}$  — доход от доставки авиалайнером  $i$  контейнера  $j$ ,

$w_j$  — вес контейнера  $j$ ,

$c_i$  — вместимость авиалайнера  $i$ ,

$x_{i,j} = 1$ , если отправить контейнер  $j$  авиалайнером  $i$ , иначе 0

Модель

$$\max \sum_{i \in I} \sum_{j \in J} p_{ij} x_{ij}$$

при ограничениях:

$$\sum_{i \in I} x_{ij} \leq 1, \quad j \in J,$$

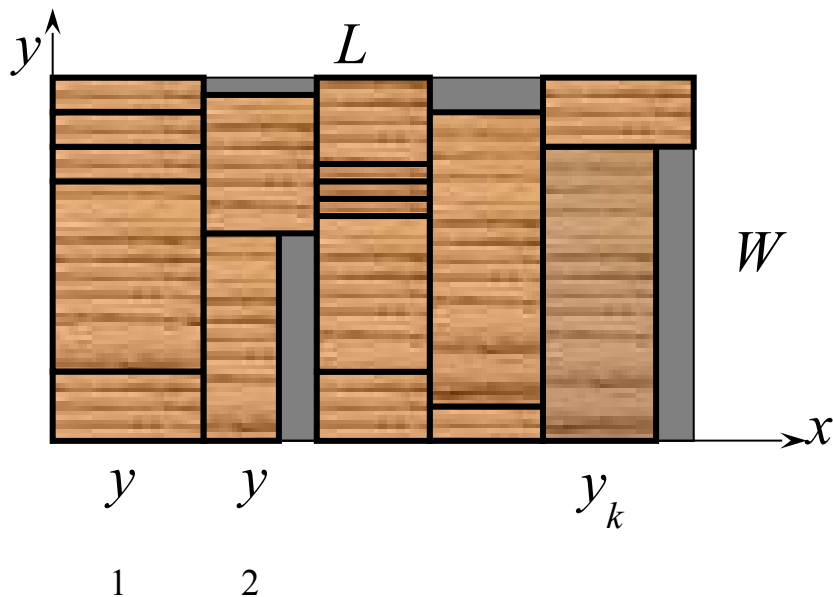
$$\sum_{j \in J} w_j x_{ij} \leq c_i, \quad i \in I,$$

$$x_{ij} \in \{0, 1\}, \quad i \in I, \quad j \in J.$$

Гильотинный раскрой материала

Дано: лист размера  $L \times W$  и  $n$ -типов прямоугольников  $l_j \times w_j$ ,  $j=1, \dots, n$   
 $p_j > 0$  — доход от прямоугольника  $j$ , повороты запрещены, разрезы параллельно осям координат от кромки до кромки. Двухстадийная обработка: сначала режем лист параллельно оси  $y$ , затем параллельно оси  $x$ .

Найти: раскрой листа с максимальным доходом



Пусть

$k$  — число параллельных полос  $k = \lfloor L / l_{\min} \rfloor$

$y_i$  — ширина полосы  $i$ ,  $1 \leq i \leq k$ ,

$x_{ij}$  — число  $j$ -х прямоугольников в полосе  $i$ ,

$$x_{i,j} = \begin{cases} 1, & \text{если } x_{ij} > 0 \\ 0, & \text{иначе} \end{cases}$$

$m_j = \lfloor W / w_j \rfloor$  — максимально возможное число  $j$ -х прямоугольников в полосе.

Модель:

$$\max \sum_{i=1}^k \sum_{j=1}^n p_j x_{ij}$$

при ограничениях

$$\sum_{j=1}^n w_j x_{ij} \leq W, \quad i=1, \dots, k,$$

$$\sum_{i=1}^k y_i \leq L,$$

$$l_j x'_{ij} \leq y_i, \quad i=1, \dots, k, \quad j=1, \dots, n,$$

$$m_j x'_{ij} \geq x_{ij}, \quad i=1, \dots, k, \quad j=1, \dots, n,$$

$$x'_{ij} \in \{0, 1\}, \quad x_{ij} \in \{0, \dots, m_j\}, \quad y_i \geq 0.$$



## Математические модели и методы принятия решений как основные компоненты исследования операций

*Математическая модель* - объективная формализация основных аспектов решаемой задачи или ее описание в математических терминах.

Математическая модель описывает исследуемую систему и позволяет выразить степень качества ее характеристик (эффективность функционирования) в виде *целевой функции*

$$W = f(X, Y),$$

где  $X = (x_1, \dots, x_n)$  — управляемые переменные,

$Y = (y_1, \dots, y_m)$  — неуправляемые переменные (исходные данные).

Связь между переменными  $X$  и исходными данными  $Y$  выражается с помощью ограничений

$$\varphi(X, Y) \leq 0.$$

При принятии решений применяют весь арсенал *методов современной прикладной математики*. Они используются для оценки ситуации и прогнозирования при выборе целей, для генерирования множества возможных вариантов решений и выбора из них лучшего.

В случае многокритериальности задачи используются

- методы замены, свертки, ранжирования критериев,
- имитационное моделирование,
- метод статистических испытаний (Монте-Карло),
- модели надежности и массового обслуживания,
- статистические (эконометрические) методы (методы выборочных обследований).
- вероятностно-статистические модели,
- методы анализа данных.

Особого внимания заслуживают проблемы неопределенности и риска, связанные как с природой, так и с поведением людей.

Различные способы описания неопределенностей:

- вероятностные модели,
- теория нечеткости,
- интервальная математика.

Для описания конфликтов (конкуренции) полезна теория игр.

Для структуризации рисков используют деревья причин и последствий.

В последние годы все большую популярность получает *контролинг* — современная концепция системного управления организацией, в основе которой лежит стремление обеспечить ее долгосрочное эффективное существование.

## **Заключение**

Овладение методами постановки задач принятия решений позволяет оптимизировать подход к решению задач обоснования оперативно-технических требований, разработки структуры системы связи, системы управления ею и распределения ресурсов системы между ее элементами.

# **«Теория принятия решений»**

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 2*

### **Постановка и содержание задачи теории принятия решений**

#### СОДЕРЖАНИЕ

##### ВВЕДЕНИЕ

##### УЧЕБНЫЕ ВОПРОСЫ:

1. Постановка задачи принятия решения. Основные определения.
2. Свойства, качества объекта и процесса принятия решения. Показатели качества и требования к ним.
3. Целевая функция (функция потерь), риски, критерий оптимальности и оценки качества решения.
4. Множество вариантов решения, ресурсы, алгоритмы принятия решений, неопределенности.

##### ЗАКЛЮЧЕНИЕ

### **Литература:**

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.
4. Терентьев В.М., Санин Ю.В. Анализ эффективности функционирования автоматизированных сетей многоканальной радиосвязи. - С-Петербург: ВАС, 1992.

### ***Постановка задачи принятия решения.***

Формально задачу принятия решения  $D$  можно записать в следующем обобщенном виде:

$$D = (F, A, X, G, P), \text{ где}$$

$F$  — *формулировка задачи принятия решения*, которая включает в себя содержательное описание проблемы и при необходимости ее модельное представление, *определение цели или целей*, которые должны быть достигнуты, а также *требования к виду окончательного результата*.

$A$  — *совокупность возможных вариантов (альтернатив)*, из которых производится выбор. Это только реально существующие варианты, в качестве которых в зависимости от задачи выступают объекты, способы достижения цели, действия, решения и т. п., либо гипотетическое множество всех теоретически возможных вариантов, которое может быть даже бесконечным. Выбор возникает только тогда, когда имеется не менее двух возможных вариантов решения проблемы.

$X$  — *совокупность свойств (признаков, атрибутов, параметров)*, *описывающих*

*варианты* и их отличительные особенности.

Во-первых, объективные показатели, которые характеризуют те или иные свойства, присущие вариантам, и *которые, как правило, можно измерить*;

во-вторых, субъективные количественные или качественные оценки, которые обычно даются по специально отобраннным или сконструированным критериям, отражающим важные для выбора свойства вариантов.

G — *совокупность условий, ограничивающих область допустимых вариантов решения задачи. Например, это могут быть ограничения на значения какого-либо признака или различная степень характерности (выраженности) признака для тех или иных вариантов, или невозможность одновременного сочетания определенных значений признаков для реально существующих вариантов.*

P — *предпочтения* одного или нескольких ЛПР, которые служат основой для оценки и сравнения возможных вариантов решения проблемы, отбора допустимых вариантов и поиска наилучшего или приемлемого варианта. Достаточно часто для упрощения - *предпочтения ЛПР, превращаются в ограничения.*



## ***Терминология системного анализа.***

*Элемент* – это минимальный неделимый объект. Элемент можно использовать только как целое, поэтому недопустимо говорить о половине или четверти элемента. Изменение постановки вопросов может потребовать разложения элементов на составные части или объединения нескольких элементов в один.

*Система* – это совокупность связанных элементов, объединенных в одно целое для достижения определенной цели. Наличие цели и заставляет связывать элементы в систему. *Целостность* — наиболее важное свойство системы. Искусственные (инженерные) системы описывают путем определения их функций и структур.

*Функция системы* – это правило получения результатов, предписанных целью (назначением) системы. *Функционировать* – значит реализовать функцию, т.е. получать результаты, предписанные назначением системы.

*Обратная связь* – воздействие результатов функционирования системы на характер этого функционирования. Различают положительную и отрицательную обратную связь.

*Структура системы* – это фиксированная совокупность элементов и связей между ними. Этот смысл отражен в данном определении структуры. Наиболее часто структура системы изображается в форме графа: элементы системы представляются вершинами графа, а связи – дугами (ребрами) графа. Граф – это математическая форма отображения структур. Инженерной формой изображения структур систем являются схемы. Схема и граф – понятия адекватные по содержанию, но различные по форме. В схемах элементы и связи обозначаются любыми фигурами, удобными для инженерных (производственных) применений.

*Организация* – это способ реализации определенных функций в системах, состоящих из большого числа элементов. Конкретная система представляет собой лишь пример реализации некоторого способа организации. Например, большинство современных ЭВМ строится на основе одного принципа организации – принципа программного управления реализацией - алгоритма на основе команд, имеющих операционно-адресную структуру.

***Свойства, качества объекта и процесса принятия решения. Показатели качества и требования к ним.***

Задание на постановку задачи:

*Глобальная система показателей качества функционирования информационной телекоммуникационной системы - ИТКС.*

*Свойство*  $x(t)$  - это объективная особенность объекта, зависящая от его физической сущности, характеризующая отдельную его сторону и позволяющая отличить один объект от другого.

*Качество*  $x(t)$  - это свойство или совокупность свойств объекта  $\vec{x}(t)$ , характеризующих его пригодность для использования по назначению.

*Элементы ИТКС:*

- система информационного обмена (СИО)
- система управления (СУ).

Каждому из свойств объекта поставлен в соответствие

$Y(\vec{x}(t))$  – *частный показатель качества (ПК)*, значение которого характеризует меру (количественную или качественную) этого свойства.

*Система показателей качества (СПК)* объекта - это вектор

$\vec{Y}(\vec{x}(t))$ , *компоненты которого* это показатели его отдельных свойств, представляющие собой *частные показатели качества* объекта.

Постановка задачи:

Процессы в системах связи

- информационного обмена  $\vec{x}_{uo}(t)$  и управления ими  $\vec{x}_y(t)$  .

Взаимосвязанные объекты - система информационного обмена (СИО)  $\vec{x}_{cuo}(t)$  ,  
включающая узлы и линии связи, и автоматизированная система управления связью  
(АСУС)  $\vec{x}_{cy}(t)$  .

Состояние системы связи в любой момент времени

$\vec{X}(t) = [x_1(t), \dots, x_m(t), \dots, x_M(t)]$  , вектора переменных состояния

а качество ее функционирования - вектор показателей качества

$\vec{Y}(\vec{x}(t)) = [Y_1(\vec{x}(t)), \dots, Y_i(\vec{x}(t)), \dots, Y_I(\vec{x}(t))]$  , где

$Y_i(\vec{x}(t))$  - компоненты векторного показателя качества, характеризующие свойства  
процесса функционирования и состояния элементов системы (СИО и АСУ).

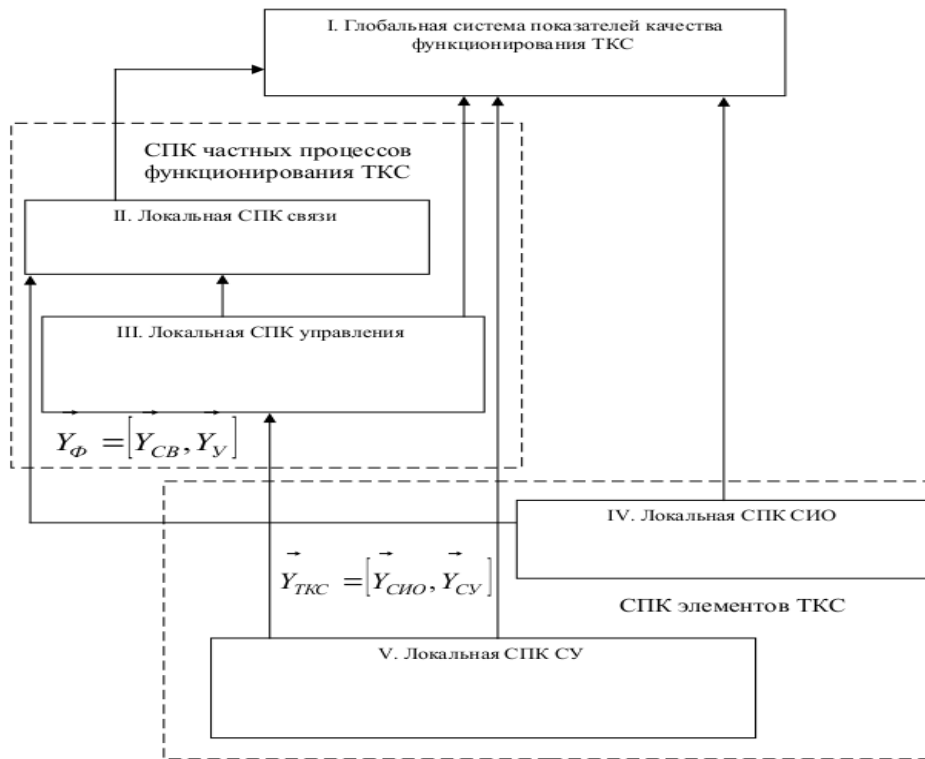


Рисунок 1. Система показателей качества ИТКС

Основные свойства ИТКС - качество информационных услуг потребителям. Взаимосвязь этих свойств ИТКС и соответствующих им ПК функционирования ИТКС представлена на рисунке 1.

В состав ГСПК  $Y_{\phi}(k)=[t_{\partial c}(k); t_{\partial \phi}(k); \vec{Z}_{\phi}(k)]$  входят:

- общее время доставки сообщения на  $k$ -ом шаге функционирования сети,
- общее время доступности к процессу функционирования на  $k$ -ом шаге,
- вектор обобщенных затрат ресурсов на процесс функционирования ИТКС на  $k$ -ом шаге.

Для локальной СПК ведущей выступает система ПК процесса информационного обмена (связи)  $\vec{Y}_{CB}(k)=[t_{DC}(k); t_{DCB}(k); Z_{CB}(k)]^T$ , куда входят

- время доставки *сообщения* на  $k$ -ом шаге функционирования сети в нормальных условиях функционирования,
- время доступности к передаваемому *сообщению* на  $k$ -ом шаге,
- вектор затрат ресурсов на *доставку сообщения* на  $k$ -ом шаге.

Показатели качества связи для сложных маршрутов сообщений по СИО ИТКС:

$$\vec{Y}_{CBi}(k) = \prod (t_1(k); t_2(k)) [M_m(k) M_{Y_{CB}}(k)]^T$$

где

- $\prod (t_1(k); t_2(k))$  - селектор временного интервала, принимающий значение 1 в интервале передачи сообщения  $(t_1(k); t_2(k))$  и 0 за его пределами;
- $M_m(k) = (\varepsilon_{jj'}^m)_{jj'=1, J_s}$ ,  $(m=1, M)$  - матрица маршрута прохождения сообщения на  $k$ -ом шаге функционирования сети с единичными элементами  $\varepsilon_{jj'}=1$  на пути его передачи и нулевыми  $\varepsilon_{jj'}=0$  в противном случае;
- $M_{Y_{CB}}(k)$  - матрица значений  $i$ -х ПК, определяемая для  $jj'$ -х направлений связи на  $k$ -ом шаге функционирования сети.



Локальная СПК управления имеет более низкий уровень иерархии и включает:

$$\vec{Y}_y(k) = [T_{\text{цп}}(k); \Delta Y_{\text{СВ}}^{\vec{Y}}(k); t_{\text{ДУ}}(k); \vec{Z}_y(k)]^T$$

- длительность цикла управления параметрами направления ИТКС при нарушении нормальных условий функционирования на k-ом шаге функционирования ИТКС;
- вектор приращений значений ПК связи на k-ом шаге функционирования сети, обусловленных ошибками в контуре управления;
- время доступности к сигналам управления на k-ом шаге
- вектор затрат ресурсов управления на k-ом шаге функционирования.

Два основных элемента системы и их ЛСПК СИО и СУ содержат:

$$\begin{aligned} \vec{Y}_{СИО}(k) &= [M_{X_Y}(k); M_V(k); M_{X_{pp}}(k); M_{\vec{z}_{СИО}}(k)]^T \\ \vec{Y}_{СУ}(k) &= [M_{X_{Y,СУ}}(k); M_{B_{СУ}}(k); M_{X_{pp,СУ}}(k); M_{\vec{z}_{СУ}}(k)]^T, \end{aligned} \quad \text{где}$$

$M_{\vec{X}_Y}$  и  $M_{\vec{X}_{Y_{cr}}}(k)$  - матрицы параметров устойчивости (живучести, помехоустойчивости и технической надежности) элементов СИО и СУ соответственно на  $k$ -ом шаге функционирования ИТКС;

$M_V(k), M_{B_{СУ}}(k)$  - матрицы пропускных способностей (скоростей передачи информации) по направлениям связи СИО и производительности (быстродействия ЭВМ) элементов СУ на  $k$ -ом шаге функционирования сети;

$M_{X_{pp}}(k), M_{X_{pp,СУ}}(k)$  - матрицы значений параметров защищенности средств связи и управления соответственно на  $k$ -ом шаге;

$M_{\vec{z}_{СИО}}(k), M_{\vec{z}_{СУ}}(k)$  - матрицы затрат ресурсов на ремонт и эксплуатацию средств связи и управления на  $k$ -ом шаге функционирования ИТКС.

***Целевая функция (функция потерь), риски, критерий оптимальности и оценки качества решения.***

Для оценки (меры) качества формируют целевую функцию ЦФ (функцию потерь).

Основные виды целевых функций:

-простая  $L = (Y(\vec{x}(t)) - Y_{mp})$  ;

-модульная  $L = |Y(\vec{x}(t)) - Y_{mp}|$  ;

-квадратичная  $L = [Y(\vec{x}(t)) - Y_{mp}]^2$  .

Значения целевой функции для различных вариантов решения формируют *риск* принятия решений, физический смысл которого определяется физическим смыслом разности между значением ПК и требованиями к нему. Обычно риск служит мерой близости рассматриваемого варианта к оптимальному, удовлетворяющему экстремуму ЦФ.

*Критерием выбора (оптимизации)* называется детерминированное либо статистическое правило выбора (оптимизации) решения, формируемое на основе целевой функции и приписываемой ей цели (направления изменения):  $>$  ;  $<$  ;  $\min$ ;  $\max$ ;  $\min \max$ ;  $\max \min$  и т.д.

Например, *выбор лучшего из возможных*  $i=1, \dots, I$  маршрутов прохождения пакета в сети связи может быть осуществлен по критерию минимума среднего времени задержки пакета на  $i$ -м маршруте, т.е.

$$M = \min [t_{zn}(i)], \quad i=1, \dots, I .$$

*Оптимизация* алгоритма оценивания  $d$  *непрерывной стохастической переменной* состояния объекта  $\vec{x}(t)$  часто осуществляется по критерию минимума среднеквадратического отклонения (дисперсии) значения оценки переменной  $\vec{\hat{x}}(t)$  от ее истинного значения, т.е.

$$M [\vec{\hat{x}}(t, d) - \vec{x}(t)]^2 .$$

***Множество вариантов решения, ресурсы, алгоритмы принятия решений, неопределенности.***

Одним из важных компонентов задачи принятия решения является *множество вариантов решений*.

Например, задачи выбора и оптимизации решений на множестве конечного числа маршрутов прохождения пакетов  $i=1, \dots, I$  и множестве алгоритмов оценивания параметров объекта  $d \in D$ .

В зависимости от типа множества вариантов различают: *дискретные и непрерывные задачи оптимизации решений*.

Поиск рациональных (оптимальных) вариантов решений из множества допустимых проводится на основе *алгоритмов принятия решений - последовательности действий, приводящих к достижению цели.*

*Виды алгоритмов принятия решений:*

- *эвристические*, определяемые предыдущей практикой (опытом) принятия решений;
- *оптимальные* – удовлетворяющие экстремуму критерия выбора (оптимальности);
- *компромиссные* – являющиеся результатом теории игр.

Заключение.

Введенные определения, постановка и содержание задач принятия решений позволяют формулировать задачи с позиций теории сложных систем, что гарантирует им математическую строгость постановки и сходимость получаемых решений к оптимальным.

# **«Теория принятия решений»**

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## ***Лекция 3***

**Методы теории вероятности, случайных процессов и математической статистики в задачах принятия решений.**

### **СОДЕРЖАНИЕ**

**ВВЕДЕНИЕ**

**УЧЕБНЫЕ ВОПРОСЫ:**

- 1. Случайные факторы, определяющие условия функционирования ИТКС и их моделирование.**
- 2. Виды распределения и параметры случайных величин и случайных процессов.**

**ЗАКЛЮЧЕНИЕ**

## **Литература:**

1. Вентцель Е.С. Теория вероятностей.- М.: Физматгиз,1969.-564с.
2. Терентьев В.М. , Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. Вентцель Е.С –Пб.: ВАС 1995 .
3. Бураченко Д.Л. и др. Общая теория связи.С=Пб.: ВАС, 1975г.



## *Случайные факторы, определяющие условия функционирования ИТКС и их моделирование.*

В ходе изучения любого явления природы или технического эксперимента всегда производится одно из двух действий — либо опыт (работа, эксперимент), либо наблюдение — определим это как испытание — воспроизведение какого-либо комплекса условий большое число раз, то есть новое испытание — это есть повторение прежнего, в одних и тех же условиях. Результатом такого изучения или испытания всегда является *событие*.

События которые происходят неизбежно в результате каждого испытания называются *достоверными*. Некоторые события вовсе не могут произойти и такие события называют *невозможными*. Если появление одного события исключает появление другого, то такие события называют *несовместимыми (несовместными)*. События называют *равновозможными*, если есть основание считать, что одно из них не более возможно, чем другое.

Любое множество событий в рамках одного или каждого испытания можно свести в общее множество и назвать это полем событий, а сами события этого поля определить как *случайные*.

Все случайные факторы или события, и случайные изменения параметров сигналов, нагрузки от пользователей телекоммуникационной сети, помех на входе приемных или линейных устройств, а также технических отказов в аппаратуре, проявляются в виде одного из следующих типов: *случайные величины*; *случайные процессы*; *случайные поля*.

***Случайные величины*** это величины, меняющиеся случайно (непредсказуемо) от одной реализации к другой, но постоянные в каждой конкретной реализации явления.

***Случайные процессы*** это случайные величины, изменяющиеся не только от реализации к реализации, но и во времени.

***Случайные поля*** это многопараметрические взаимно обусловленные случайные процессы, описывающие, как правило, распределенные в пространстве и во времени объекты или явления.

## ***Примеры***

*Примерами случайных величин* являются число ошибок в знаках при приеме телеграмм или число ошибочно принятых кодовых комбинаций в системах передачи, величина ошибки в юстировке антенн в ходе развертывания линии связи, число заявок на обслуживание от пользователей АТС в единицу времени.

*Примерами случайных процессов* могут служить процесс изменения амплитуды, фазы, частоты и угла прихода сигнала на выходе канала связи; почасовое изменение нагрузки на сеть связи; изменение числа отказов аппаратуры в зависимости от срока ее службы.

*Примерами случайных полей* являются изменение яркости изображения принимаемого телевизионного сигнала в двух координатах, определяющих точку на плоскости экрана, турбулентность атмосферы или диэлектрическая проницаемость тропосферы, в пределах границ распространения радиосигнала.

# ***Виды распределения и параметры случайных величин, случайных процессов и полей.***

## **1. Случайные величины.**

Значения случайной величины  $x_i$ ,  $i=1\dots N$ , которые она принимает в отдельных опытах, называются *реализациями* случайной величины.

Случайные величины бывают *скалярными* и *векторными*. Полной характеристикой случайной скалярной величины  $x$  является плотность распределения вероятностей ее значений  $w(x)$ , интеграл от которой называют интегральной функцией распределения

$$F(x) = P(x \in X) = \int_{-\infty}^x w(x) dx .$$

со следующими свойствами:

1. значения функции  $\in [0; 1]$  , иначе  $0 \leq F(x) \leq 1$  , или  $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$
2. функция  $F(x)$  - неубывающая  $F(x_2) \geq F(x_1)$  , если  $x_2 \geq x_1$

*Плотность скалярной величины обладает следующими свойствами:*

$$\begin{aligned} w(x) &\geq 0; \\ w(-\infty) &= w(\infty) = 0; \\ \int_{-\infty}^{\infty} w(x) dx &= 1. \end{aligned} \quad (1)$$

Статистические свойства векторной случайной величины полностью описываются совместной *плотностью распределения вероятностей*:

$$w(\vec{x}) = w(x_1, x_2, \dots, x_N),$$

а ее *интегральная функция распределения* имеет вид:

$$F(\vec{x} \subseteq \vec{X}) = \iiint_{\vec{X}} w(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N \quad (2)$$

*Математическим ожиданием* (средним) случайной величины называют начальный момент первого порядка:

$$M[x] = m_x = \int_{-\infty}^{\infty} xw(x) dx. \quad (3)$$

Средневзвешенное отклонение централизованной случайной величины может быть определено следующим образом:

$$m_{\Delta x} = \int_{-\infty}^{\infty} (x - m_x) w(x) dx . \quad (4)$$

момент второго порядка:

$$M(\Delta x)^2 = D_x = \sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 w(x) dx \quad (5)$$

называется *дисперсией* случайной величины, характеризующей степень разброса значения случайной величины относительно своего среднего значения. Наряду с дисперсией мерой рассеяния случайной величины является ее *среднеквадратическое отклонение*

$$\sigma_x = \sqrt{D_x} .$$

Мерой статистической взаимосвязи двух скалярных центрированных случайных величин может служить *взаимный корреляционный момент*:

$$K_{xy} = M[\Delta x \Delta y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) w(x, y) dx dy . \quad (6)$$

Для отражения статистической связи значений одной и той же центрированной случайной величины служит *момент автокорреляции*.

Нормированный корреляционный момент – *коэффициент корреляции*:

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}, \sigma_x \geq 0, \sigma_y \geq 0 . \quad (7)$$

При этом,

если две СВ  $x$  и  $y$  связаны линейно, то  $r_{xy} = \mathbf{1}$  (либо  $-1$ );

если эти величины оказываются *некоррелированными*, то  $r_{xy} = \mathbf{0}$ .

**Распределения вероятностей случайных  
дискретных и непрерывных величин:**

**Биномиальное распределение** — дискретное распределение когда случайная величина имеет всего два значения 0 или 1 — событие наступило либо не наступило (бернуллиевская схема).

$$P(k) = C_n^k p^k q^{n-k}, k=0, \dots, n; q=1-p; p \geq 0, M(k) = np, D(k) = npq. \quad (8)$$

Здесь  $n$  - число проведенных испытаний (опытов),  $p$  - вероятность наступления события в одном опыте,  $k$  - число наступивших событий в  $n$  опытах.

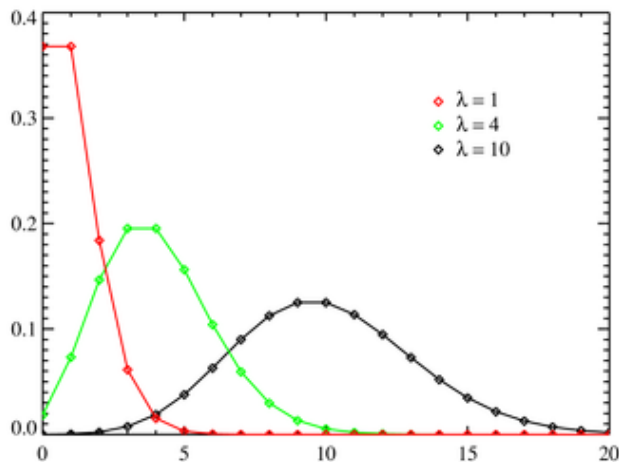
**Распределение Пуассона** — вероятностное распределение дискретного типа, показывает случайную величину, представляющую число событий произошедших за фиксированное время — играет ключевую роль в теории массового обслуживания.

$$P(k) = \frac{a^k}{k!} \exp(-a), a > 0, k=0, 1, \dots, M(k) = a, D(k) = a. \quad (9)$$

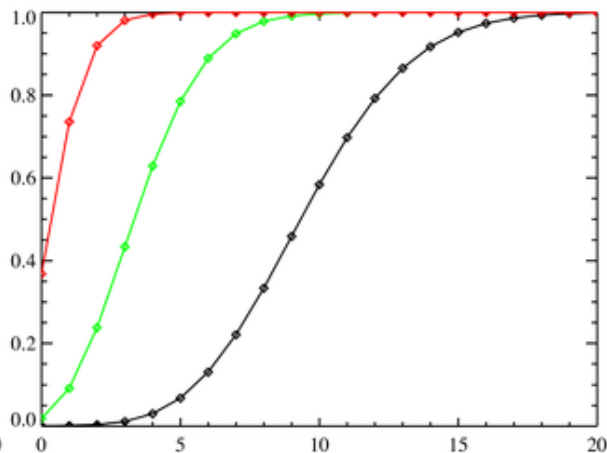


Здесь  $a = \lambda t$  - среднее число событий, происходящих с интенсивностью  $\lambda$  за заданный временной интервал  $t$ ,  $k$ -число событий, происходящих за время  $t$ .

Функция плотность вероятности



Функция распределения



**Равномерное распределение на интервале.**

$$P(k) = \frac{1}{n}, k=0,1,\dots,n>0, M(k) = (n+1)/2, D(k) = (n^2+1)/12. \quad (10)$$

Здесь  $n$ - число возможных состояний ДСВ,  $k$ - номер состояния ДСВ.

**Показательный закон распределения непрерывной СВ.**

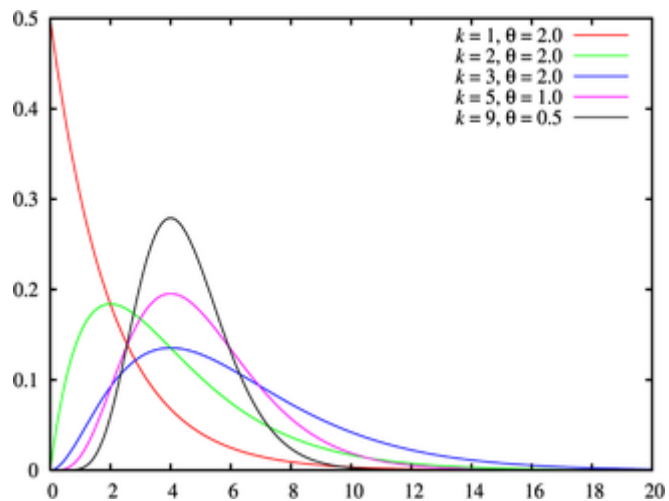
$$w(x) = \lambda \exp(-\lambda x), \lambda > 0, M(x) = 1/\lambda, D(X) = 1/\lambda^2. \quad (11)$$

Здесь  $\lambda$  - параметр распределения.

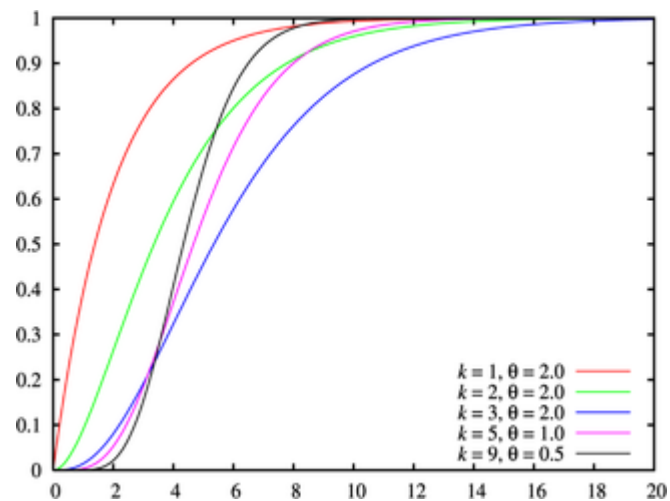
**Гамма распределение — распределение Эрланга.**

$$w(x) = \frac{\lambda(\lambda x)^{h-1} \exp(-\lambda x)}{\Gamma(h)}, x \geq 0, \lambda > 0, h > 0, \Gamma(h) > 0 \quad (12)$$
$$\Gamma(h) = \int_0^{\infty} e^{-\theta} \theta^{h-1} d\theta - \text{гамма-функция Эйлера}; M(x) = h/\lambda, D(x) = h/\lambda^2.$$

Плотность вероятности



Функция распределения



## Нормальный закон распределения — Гаусса или Гаусса-Лапласа.

Нормальное и усеченное нормальное распределения следующего вида:

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right); \quad (13)$$

$$w(x) = \frac{C_0}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right), 0 < x \leq \infty, \quad (14)$$

где  $C_0 = \frac{1}{1 + \Phi\left(\frac{m_x}{\sqrt{\sigma_x^2}}\right)}$  - коэффициент нормирования, определяемый на основе

табулированного интеграла вероятности.

При этом вероятность попадания реализации СВ  $x$  в интервал  $[\alpha, \beta]$  равна:

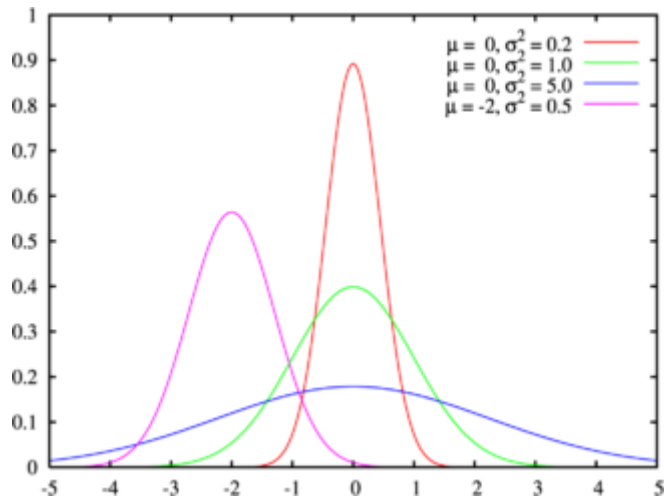
$$P(\alpha < x \leq \beta) = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right) dx .$$

При замене переменной  $t = \frac{x - m_x}{\sqrt{\sigma_x^2}}$  вычисление интеграла (12) сводится к

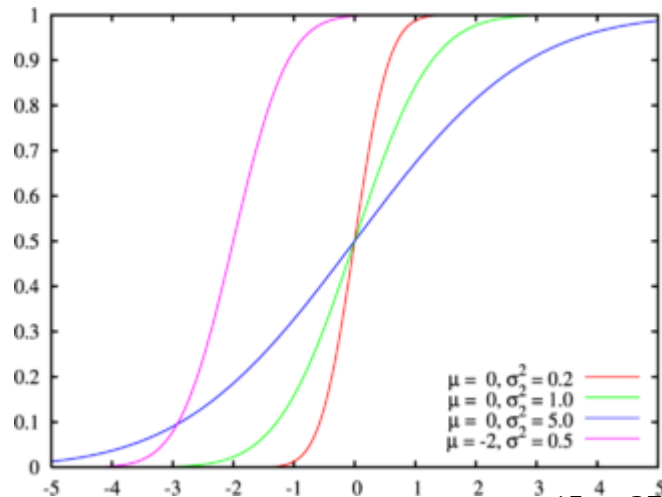
вычислению табулированного интеграла вероятности:

$$\begin{aligned}
 P(\alpha < x \leq \beta) &= [\Phi(B) - \Phi(A)], \\
 A &= \frac{\alpha - m_x}{\sqrt{\sigma_x^2}}; B = \frac{\beta - m_x}{\sqrt{\sigma_x^2}}; \\
 \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \Phi(x) = -\Phi(-x).
 \end{aligned}
 \tag{15}$$

Плотность вероятности



Функция распределения



Для случая линейного преобразования случайной величины:

$y = Ax + b$ , соотношения для числовых характеристик векторной случайной величины

$\vec{x}$  имеют вид:

$$\begin{aligned} \vec{m}_y &= A \vec{m}_x + \vec{b}, \\ A &= \{a_{ij}\}, \vec{b} = \{b_i\}, i=1 \dots N, j=1 \dots N; \\ K_y &= [AK]_x A^T; \\ \sigma_{y_i}^2 &= \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2 K_{x_{ij}}. \end{aligned} \tag{16}$$

Совместная плотность  $w(\vec{x}) = w(y, z)$  устанавливает статистическую взаимосвязь между всеми составляющими вектора  $\vec{x}$  и может быть определена через условную плотность распределения вероятностей  $w(y|z)$  и  $w(z|y)$ , если вектор  $\vec{x} = (y; z)$  имеет своими компонентами случайные величины  $y$  и  $z$ . Из формулы полной вероятности имеем:

$$w(\vec{x}) = w(y, z) = w(y|z)w(z) = w(z|Y)w(y), \tag{17}$$

Для условных плотностей распределения справедлива формула Байеса— формула позволяющая переоценить вероятность гипотез (предположений, условий) после того, когда стали известны результаты испытаний или экспериментов:

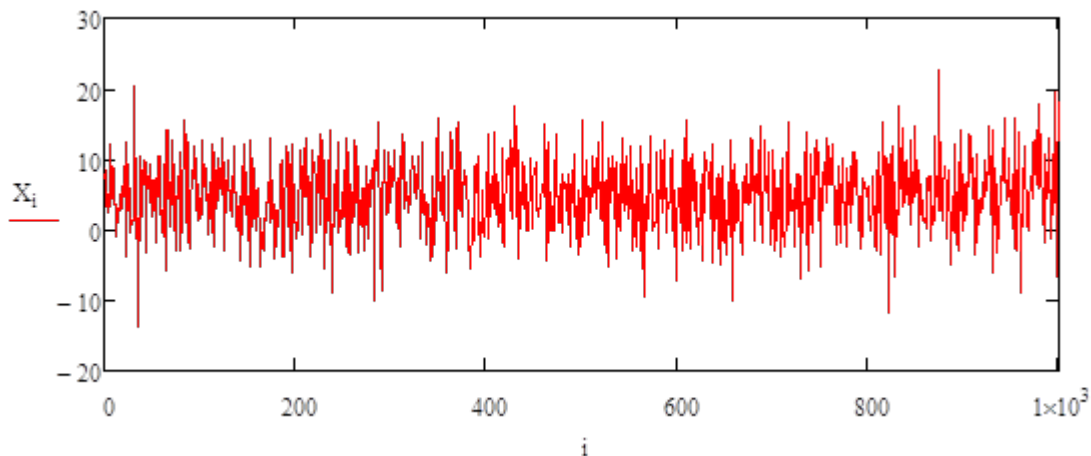
$$\begin{aligned}
 w(y|Z) &= \frac{w(\vec{x})}{\int_{-\infty}^{\infty} w(z|Y)w(y)dy} = \frac{w(y)w(z|Y)}{\int_{-\infty}^{\infty} w(z|Y)w(y)dy}; \\
 w(z|Y) &= \frac{w(\vec{x})}{\int_{-\infty}^{\infty} w(y|Z)w(z)dz} = \frac{w(z)w(y|Z)}{\int_{-\infty}^{\infty} w(y|Z)w(z)dz}.
 \end{aligned}
 \tag{18}$$

Если вектор  $\vec{x}$  состоит из  $N$  независимых случайных величин, то

$$w(\vec{x}) = \prod_{i=1}^N w(x_i).
 \tag{19}$$

## 2 . Случайные процессы.

Процесс изменения во времени некоторой случайной величины называется случайным процессом. Примером случайного процесса может служить изменение амплитуды, фазы и частоты сигнала на выходе канала связи. Одна из реализаций этого процесса изменения амплитуды сигнала представлена на рис.1.





Для удобства статистического описания случайного процесса его можно рассматривать как множество случайных величин, соответствующих различным моментам времени на рассматриваемом интервале:

$$x(t)=[x(t_1), \dots, x(t_k), \dots, x(t_N)], [t_1, \dots, t_N].$$

И полным статистическим описанием случайного процесса будет его многомерная плотность распределения вероятностей значений процесса во всех временных сечениях внутри заданной временной области:

$$w(x(t), t) = w(x_1, t_1; \dots; x_k, t_k; \dots; x_N, t_N). \quad (20)$$

- Наиболее распространенными моделями случайных процессов являются:
- процессы с независимыми приращениями;
- нормальные случайные процессы;
- марковские процессы.

Случайный процесс называют *процессом с независимыми приращениями*, если для любых сечений времени, внутри заданного временного интервала, приращения процесса являются независимыми случайными величинами.

Случайный процесс, у которого  $n$ -мерная плотность распределения вероятностей является *гауссовской*, называется *нормальным*.

Случайный процесс относится к классу *марковских*, если для его  $n$ -мерной плотности распределения выполняется условие:

$$w(x_k, t_k | X_1, t_1; \dots; X_{k-1}, t_{k-1}) = w(x_k, t_k | X_{k-1}, t_{k-1}) \quad (21)$$

Для стационарных в широком смысле процессов выполняется постоянство во времени его моментов: среднего и дисперсии, а также инвариантность к временному сдвигу корреляционной функции:

$$\begin{aligned} m_x(t) &= \int_{-\infty}^{\infty} x(t) w(x(t), t) dx = m_x(t - \tau) = m_x; \\ \sigma_x^2(t) &= \int_{-\infty}^{\infty} (x(t) - m_x(t))^2 w(x(t), t) dx = \sigma_x^2(t - \tau) = \sigma_x^2; \\ K(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t_1) - m_x(t_1)][x(t_2) - m_x(t_2)] \times \\ & w(x(t_1), t_1; x(t_2), t_2) dx_1 dx_2 = K(t_1 - \tau, t_2 - \tau) = K(t_2 - t_1) \end{aligned} \quad (22)$$

При  $t_2 - t_1 = 0$  корреляционная функция случайного процесса равна его дисперсии  $K(0) = D_x = \text{const}$ . Интервал времени  $\tau = \tau_{кор}$ , за пределами которого корреляционная функция  $K_x(\tau)$  не превосходит некоторую малую  $(0,05-0,1)D_x$  величину, называют интервалом корреляции процесса  $\tau_{кор}$ .

К стационарным процессам в узком смысле относятся процессы, у которых сама  $n$ -мерная плотность распределения вероятностей инвариантна к временному сдвигу  $\tau$ :

$$w(x_1, t_1; x_2, t_2; \dots; x_k, t_k) = w(x_1, t_1 + \tau; \dots; x_k, t_k + \tau), \text{ для любого } \tau.$$

Для эргодического процесса усреднение по статистике совпадает с усреднением по времени:

$$\lim_{\tau \rightarrow \infty} K_x(\tau) = 0.$$

Для эргодического процесса усреднение по множеству (статистике) может быть заменено усреднением по времени, а, следовательно, значение спектра (спектральной плотности мощности) процесса  $G(\omega)$  связано со спектральной плотностью сигнала  $S(j\omega)$  соотношением:

$$G(\omega) = \lim_{T \rightarrow \infty} \frac{|S(j\omega)|^2}{\pi T},$$

где  $T$  - время наблюдения процесса.

Используя вычисленную функцию корреляции можно определить спектр (спектральную плотность мощности) стационарного случайного процесса, используя теорему Винера-Хинчина:

$$\begin{aligned} G_x(\omega) &= \int_{-\infty}^{\infty} K_x(\tau) e^{-j\omega\tau} d\tau; \\ K_x(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G_x(\omega) e^{j\omega\tau} d\omega. \end{aligned} \quad (23)$$

Под эффективной шириной спектра случайного процесса с неравномерной спектральной плотностью мощности понимается полоса частот, удовлетворяющая

соотношению:

$$\Delta\Omega_{эф} = \frac{\int_0^{\infty} G(\omega) d\omega}{G_{\max}} = \frac{K(\tau=0)}{G_{\max}}.$$

Случайный процесс, спектр которого непрерывен и сосредоточен возле некоторой фиксированной частоты  $\omega_0$ , а также выполняется условие:

$$\frac{\Delta\Omega}{\omega_0} \ll 1, \quad \text{называется узкополосным.}$$

Используя понятие эффективной ширины спектра можно установить следующую связь между ее значением и интервалом корреляции случайного процесса:

$$\tau_{кор} = \frac{\pi G(0)}{2 G_{\max} \Delta\Omega_{эф}},$$

где  $G_{\max}$ ,  $G(0)$  – значение спектральной плотности мощности процесса в точке максимума и в точке  $\omega = 0$ .

*Примеры корреляционных функций типовых случайных процессов.*

1. Для белого гауссовского шума, определяемого как нормальный случайный процесс  $x(t)$ , его значения в сколь угодно близкие моменты времени остаются не коррелированными, т.е. имеют корреляционную функцию вида:

$$K_x(t_1, t_2) = N(t_1) \delta(t_2 - t_1), \quad (24)$$

где  $N(t)$ - интенсивность БГШ, являющаяся постоянной для стационарного случая и равной единице для стандартного БГШ;

$$\delta(t_2 - t_1) = \left\{ \begin{array}{l} \infty, \text{ при } t_1 = t_2; \\ 0, \text{ в остальных случаях.} \end{array} \right\} \text{ - дельта- функция.}$$

Используя теорему Винера-Хинчина, получим спектральную плотность мощности БГШ:

$$G_x(\omega) = \frac{N}{2\pi} = \text{const}.$$

Из выражения (18) следует, что дисперсия БГШ  $\sigma_w^2 = K_x(0) = \infty$  в силу бесконечного спектра  $\Delta\Omega_{эф}$  этого процесса, а интервал корреляции БГШ равен:

$$\tau_{кор} = \frac{\pi}{2(\Delta\Omega_{эф} = \infty)} = 0.$$

2. Экспоненциально коррелированным процессом называют процесс с функцией корреляции вида:

$$K_x(\tau) = D_x \exp(-\alpha |\tau|). \quad (25)$$

Соответствующая спектральная плотность мощности процесса имеет вид:

$$G_x(\omega) = \frac{D_x \alpha}{\pi(\alpha^2 + \omega^2)}. \quad (26)$$

### 3. Случайные поля.

Векторное случайное поле  $\mathbf{u}(\mathbf{x})$ , у которого аргумент  $\mathbf{x}=(x, y, z, t)$  является вектором, содержащим три пространственные координаты и время  $t$ , может быть задан в рамках корреляционной теории вектором математических ожиданий  $m_{\mathbf{u}}(\mathbf{x}) = [m_{u_x}, m_{u_y}, m_{u_z}]^T$  и матричной корреляционной функцией  $K_{\mathbf{u}}(\mathbf{x}_1, \mathbf{x}_2)$ .

Рассмотрим стационарный, эргодический процесс с нормальным распределением (13) его мгновенных значений  $x(t) = A(t) \cos[\omega_0 t + \theta(t)]$ .

Определим плотности распределения вероятностей его параметров:

$A(t)$  - амплитуды и  $\theta(t)$  - фазы.

Решение задачи целесообразно провести для квадратурного представления сигнала:

$$x(t) = [A_c(t) \cos \theta(t) + A_s(t) \sin \theta(t)] \cos \omega_0 t, \quad (27)$$

где  $A(t) = \sqrt{A_c^2(t) + A_s^2(t)}$  - огибающая сигнала;

$\theta(t) = \arctg \frac{A_s(t)}{A_c(t)}$  - фаза сигнала.

Квадратуры сигнала  $A_c, A_s$  имеют распределение, аналогичное мгновенным значениям сигнала, так как квадратуры формируются на основе линейного сдвига по частоте, поэтому:

$$w(A_{c(s)}) = \frac{1}{\sqrt{2\pi\sigma_{c(s)}^2}} \exp\left(-\frac{A_{c(s)}^2}{2\sigma_{c(s)}^2}\right).$$



Учитывая статистическую независимость квадратур, совместная плотность распределения вероятностей квадратур сигнала имеет вид:

$$w(A_c, A_s) = w(A_c)w(A_s) = \frac{A}{2\pi\sigma_x^2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right), \text{ при этом } \sigma_x^2 = \sigma_c^2 + \sigma_s^2 \quad (28)$$

Плотность распределения (28) может быть без труда декомпозирована на две плотности распределений вероятностей, а именно плотность распределения амплитуды  $w(A)$  и фазы  $w(\theta)$ , которые имеют вид распределения Релея и равномерного распределения в пределах  $[0-2\pi]$  соответственно:

$$w(A) = \frac{A}{\sigma_x^2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right); \quad (29)$$
$$w(\theta) = \frac{1}{2\pi}.$$

### ***Заключение.***

На лекции рассмотрены вопросы классификации случайных факторов, определяющих качество процесса функционирования линий и сетей связи, введены основные понятия и характеристики случайных величин, сигналов, процессов и полей.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 4*

### **Методы математической статистики в задачах принятия решений**

#### СОДЕРЖАНИЕ

##### Введение

##### Учебные вопросы :

1. Постановка задачи и общий алгоритм анализа случайных последовательностей при принятии решений с использованием методов математической статистики.
2. Алгоритмы получения эмпирических оценок числовых характеристик, вероятностей и законов распределения случайных последовательностей и анализ их качества.

##### Заключение

## Литература:

1. Гмурман В. Е. Теория вероятностей и математическая статистика . Учебное пособие для вузов – Изд.. 7-е, стер. – М.: Высш. шк. 2001.
2. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. Учебное пособие для вузов – Изд.. 2-е, стер. – М.: Высш. шк. 2000.
3. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. Учебное пособие для вузов – Изд.. 2-е, стер. – М.: Высш. шк. 2000.

## Введение

На прошлой лекции были рассмотрены случайные факторы, их распределения и параметры оценок величин, то есть что можно получить, имея такие данные. На этой лекции мы рассмотрим вопрос — как это делать и каким математическим аппаратом производить анализ, оценку и оптимизацию данных в задачах принятия решений.

При принятии решений проблемы с оценкой получаемых данных подразделяются на три класса:

1) хорошо структурированные или количественно сформулированные данные и соответственно анализ проблемы, в которых получают численные оценки;

2) неструктурированные или качественно выраженные данные и отсюда проблемы, в которых количественные зависимости между признаками и характеристиками совершенно неизвестны;

3) слабо структурированные или смешанные данные и следовательно проблемы, содержащие как количественные, так и качественные элементы, причем последние имеют тенденцию к доминированию.

***Постановка задачи и общий алгоритм анализа случайных последовательностей с использованием методов математической статистики.***

Наиболее полным описанием случайной последовательности является функция распределения вероятностей ее значений и задача анализа в общем случае сводится к получению эмпирических вероятностных характеристик по доступным выборочным данным и проверке гипотез о их соответствии некоторым стандартным характеристикам, определяющим различные классы случайных последовательностей и отдельные их свойства. Часто в качестве стандартной случайной последовательности (СП)  $X$  выступает последовательность, например, с нормальным распределением  $N(M_X; D_X)$  и числовыми характеристиками:  $M_X$  - математическое ожидание и  $D_X$  - дисперсия случайной последовательности.

*Общий алгоритм анализа случайной последовательности с учетом вводимой стандартной случайной последовательности может включать следующие этапы.*

1. Определение эмпирических вероятностных характеристик анализируемой случайной последовательности (математического ожидания, дисперсии, корреляционного момента, вероятностей событий и функции распределения вероятностей). Важно, чтобы качество полученных эмпирических оценок соответствовало выдвигаемым априорно требованиям к допустимому отклонению от истинных значений характеристик (доверительному интервалу и доверительной вероятности), а также определялось требуемым для этого размером выборки. На основе полученных характеристик могут быть установлены свойства симметрии распределения (совпадение значений среднего, моды и медианы, либо равенство значений вероятностей превышения и

не превышения среднего значения) и близости его формы к некоторому стандартному (см. лекцию 3), например, к нормальному.

2. Построение гистограммы вероятностей и восстановление эмпирического распределения случайной последовательности на основе полученных вероятностных характеристик и выдвижение гипотезы о виде распределения СП.
3. Проверка верности выдвинутой гипотезы по критериям соответствия эмпирических и аналитических вероятностных характеристик, а также определение класса и основных свойств случайной последовательности с оценкой показателей качества полученных оценок и решений.

**Основные этапы анализа случайных последовательностей** в предположении выполнения условия стационарности выборочных данных.

Вероятностной характеристикой  $\theta$  случайной величины  $X$ , определяемой непосредственно путем эксперимента, является некоторое число - математическое ожидание, дисперсия, вероятность события  $\alpha < X < \beta$ . Символ  $\theta$  означает истинное значение характеристики. Путем обработки результатов экспериментального исследования  $X$  получают экспериментальное значение характеристики  $\theta$ , статистическую характеристику или оценку  $\tilde{\theta}$  характеристики  $\theta$ .

Экспериментальное исследование случайной величины  $X$  с целью определения  $\tilde{\theta}$  - оценки (приближенного значения)  $\theta$ , заключается в проведении  $N$  **опытов** (испытаний, наблюдений) и получении (путем соответствующих измерений) ряда значений  $x_1, \dots, x_i, \dots, x_N$  — реализаций  $X$ . В результате обработки экспериментальных данных определяется  $\tilde{\theta} = \Psi_{\theta}(x_1, \dots, x_i, \dots, x_N)$  как функция эксперимента.



Если провести **еще одну серию из  $N$  опытов**, то будет получен ряд других реализаций  $x'_1, \dots, x'_i, \dots, x'_N$  случайной величины  $X$  и другое значение  $\tilde{\theta}'$  оценки искомой характеристики  $\theta$ . Значение  $x_i$  случайной величины  $X$ , полученное в результате  $i$ -ого опыта в серии, можно рассматривать как значение случайной величины  $X_i$  а оценку  $\tilde{\theta}$  - как реализацию более общей случайной величины

$$\tilde{\Theta} = \Psi_{\theta}(X_1, \dots, X_i, \dots, X_N), \quad (1)$$

**Вероятностными характеристиками системы двух случайных величин  $(X, Y)$** , определяемыми непосредственно на основании эксперимента, являются математические ожидания, дисперсии, корреляционный момент, вероятность события  $\alpha_x < X < \beta_x, \alpha_y < Y < \beta_y$ . Эксперимент заключается в проведении  $N$  опытов и получении ряда значений  $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)$  реализаций случайных величин  $X, Y$ . В результате обработки экспериментальных данных получается оценка

$$\tilde{\theta} = \Psi_{\theta}((x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)),$$

как реализация случайной функции аналогичной (1).

$$\tilde{\Theta} = \Psi_{\theta}(X_1, Y_1; \dots, X_i, Y_i; \dots, X_N, Y_N), \quad (2)$$

Погрешность приближения оценки  $\tilde{\Theta}$  к  $\theta$  равная

$$\Delta\tilde{\Theta} = \tilde{\Theta} - \theta, \quad (3)$$

является, как и  $\tilde{\Theta}$ , случайной величиной.

Функцию  $\Psi_{\theta}$  желательно выбирать так, чтобы выполнялось три условия

1. Математическое ожидание  $\Delta\tilde{\Theta}$  равно нулю:

$$M[\Delta\tilde{\Theta}] = 0. \quad (4)$$

2. Дисперсия  $\Delta\tilde{\Theta}$  стремится к нулю с увеличением  $N$

$$\lim_{N \rightarrow \infty} D(\Delta\tilde{\Theta}) = 0. \quad (5)$$

3. Дисперсия  $D(\Delta\tilde{\Theta})$  при данной  $\Psi_{\theta}$  должна быть наименьшей.

### Определения

При выполнении условия (4) **оценка**  $\tilde{\theta}$  называется **несмещенной**,  
условий (4), (5) - **состоятельной**,  
всех трех условий - **эффективной**.

Вследствие случайного характера погрешности (3) для характеристики точности приближенного равенства  $\tilde{\theta} = \theta$  необходимо располагать вероятностью  $p_\Delta$  того, что абсолютное значение погрешности не превзойдет некоторого предела

$$P(|\Delta\tilde{\theta}| \leq \Delta_\theta) = p_\Delta. \quad (6)$$

Интервал от  $\tilde{\theta} - \Delta_\theta$  до  $\tilde{\theta} + \Delta_\theta$ , в котором с вероятностью  $p_\Delta$  находится истинное значение  $\theta$ , называется **доверительным интервалом**, его границы - **доверительными границами**, а вероятность  $p_\Delta$  - **доверительной вероятностью**.

Если число экспериментальных данных  $N$  достаточно велико, то погрешность (3) состоятельной оценки  $\tilde{\theta}$  можно практически считать распределенной нормально с математическим ожиданием (4), дисперсией  $D(\Delta\tilde{\theta})=D(\tilde{\theta})=D_{\theta}$  и средним квадратичным отклонением  $\sigma(\Delta\tilde{\theta})=\sigma(\tilde{\theta})=\sigma_{\theta}=\sqrt{D_{\theta}}$ . При этом выражение (6) имеет

вид:

$$p_{\partial} = \Phi\left(\frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}}\right) - \Phi\left(-\frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}}\right) = 2\Phi\left(\frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}}\right) = 2\Phi(t), \quad (7)$$

где  $\Phi(t)$  - функция Лапласа,  $t = \frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}}$ .

С помощью этой формулы решается задача определения доверительной вероятности  $p_{\partial}$  по известным данным  $\Delta_{\tilde{\theta}}, \sigma_{\tilde{\theta}}$ .

Функция Лапласа  $\tau = \Phi(t)$  выражает зависимость  $\tau$  от  $t$ . Обратная  $t = \Phi^{-1}(\tau)$  выражает зависимость  $t$  от  $\tau$ . При  $t = \frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}}$ ,  $\tau = \frac{p_{\partial}}{2}$  имеем

$$\frac{\Delta_{\tilde{\theta}}}{\sigma_{\tilde{\theta}}} = \Phi^{-1}(\tau). \quad (8)$$

С помощью формулы (8) и обратной функции Лапласа решается задача определения доверительного интервала  $\Delta_{\tilde{\theta}}$  по известным  $p_{\partial}$  и  $\sigma_{\tilde{\theta}}$  и необходимого числа испытаний по известным  $p_{\partial}$  и  $\Delta_{\tilde{\theta}}$ .

При решении первой задачи согласно (8) определяется  $\Delta_{\tilde{\theta}}$ . При решении второй задачи согласно (8) определяется  $\sigma_{\tilde{\theta}}$ , а затем N.

Для проведения анализа СП обычно приводят к стандартному виду. Для случая двоичной ноль - единичной последовательности это достигается перекодировкой исходной последовательности в симметричную -1,1- ю последовательность в соответствии с правилом

$$x_i(k) = 2x_i'(k) - 1.$$

Здесь  $x_i(k)$ ,  $x_i'(k)$  - элементы стандартной и исходной последовательностей соответственно.

***Алгоритмы получения эмпирических оценок числовых характеристик, вероятностей и законов распределения случайных последовательностей и анализ их качества.***

*Определение математического ожидания*

$$\hat{M}_x = \frac{1}{N} \sum_{i=1}^N x_i, \quad \tilde{M}_x = \frac{1}{N} \sum_{i=1}^N X_i, \quad (9)$$

где  $X_i$  - независимые случайные величины с одинаковыми  $M_{x_i} = M_x$  и  $D_{x_i} = D_x$ .

Математическое ожидание погрешности оценки среднего равно

$$M[\Delta \tilde{M}_x] = M[\tilde{M}_x - M_x] = M\left[\frac{1}{N} \sum_i^N X_i\right] - M_x = \frac{1}{N} NM_x - M_x = 0. \quad (10)$$

Дисперсия погрешности оценки среднего равна

$$D(\Delta \tilde{M}_x) = D(\tilde{M}_x) = D_{\tilde{M}_x} = D\left(\frac{1}{N} \sum_i^N X_i\right) = \frac{1}{N^2} D\left(\sum_i^N X_i\right) = \frac{1}{N^2} ND_x = \frac{D_x}{N} = \frac{\sigma_x^2}{N}. \quad (11)$$

Среднеквадратическое отклонение оценки математического ожидания (9)

$$\sigma_{\tilde{M}_x} = \sigma_x \sqrt{N} \approx \tilde{\sigma}_x \sqrt{N}. \quad (12)$$

Оценка (9) – несмещенная, состоятельная и эффективная.

## Определение оценки дисперсии и ее среднеквадратического отклонения

Оценка дисперсии  $D_x$  
$$D_x = \frac{1}{N} \sum_i^N (x_i - M_x)^2 .$$

Так как значение  $M_x$  априори неизвестно, то принимают  $M_x \approx \tilde{M}_x$  и тогда

$$\tilde{D}_x = \frac{1}{N} \sum_i^N X_i^2 - \left( \frac{1}{N} \sum_i^N X_i \right)^2 . \quad (13)$$

Математическое ожидание погрешности оценки равно

$$M[\Delta \tilde{D}_x] = M[\tilde{D}_x - D_x] = -\frac{D_x}{N} , \quad (14)$$

что означает, что оценка (14) является *смещенной*.

Смещение пропорционально  $D_x$  и обратно пропорционально  $N$ . Это означает, что оценка  $D_x$ , полученная согласно (14), - *состоятельная*.

Смещение устраняется с переходом к  $\tilde{D}'_x = \frac{N}{N-1} \tilde{D}_x$ .

При этом вместо (13) имеем 
$$\tilde{D}'_x = \frac{1}{N-1} \sum_i^N x_i^2 - \frac{N}{N-1} \tilde{M}_x^2 . \quad (15)$$

При больших значениях  $N$  результаты расчета по формулам (13) и (15) практически будут одинаковыми.

Зависимость среднего квадратического отклонения  $\sigma_{\tilde{D}_x}$  от его точного значения  $\sigma_x$  определяется выражением

$$\sigma_{\tilde{D}_x} \approx \sqrt{\frac{2}{N-1}} \sigma_x .$$

## Определение корреляционного момента и коэффициента корреляции

Экспериментальное значение корреляционного момента  $R_{xy}$  как оценка смешанного центрального момента  $m_{11}$  системы двух случайных величин равно

$$R_{xy} = \frac{1}{N} \sum_1^N (x_i - M_x)(y_i - M_y). \quad (16)$$

Так как значения  $M_x$ ,  $M_y$  неизвестны, то принимают  $M_x \approx \tilde{M}_x$ ,  $M_y \approx \tilde{M}_y$  и тогда

$$\tilde{R}_{xy} - \frac{1}{N} \sum_1^N (x_i - \tilde{M}_x)(y_i - \tilde{M}_y) = \frac{1}{N} \sum_1^N x_i y_i - \left( \frac{1}{N} \sum_1^N x_i \right) \left( \frac{1}{N} \sum_1^N y_i \right)$$

или

$$\tilde{R}_{xy} = \frac{1}{N} \sum_1^N X_i Y_i - \left( \frac{1}{N} \sum_1^N X_i \right) \left( \frac{1}{N} \sum_1^N Y_i \right). \quad (17)$$

Погрешность оценки  $\tilde{R}_{xy}$

$$\Delta \tilde{R}_{xy} = \tilde{R}_{xy} - R_{xy} \quad (18)$$



Математическое ожидание погрешности (18)

$$M[\Delta \tilde{R}_{xy}] = M\left[\frac{1}{N} \sum_1^N X_i Y_i\right] - M\left[\left(\frac{1}{N} \sum_1^N X_i\right)\left(\frac{1}{N} \sum_1^N Y_i\right)\right] = -\frac{R_{xy}}{N}$$

Это означает, что оценка (17) - смещена и равна

$$\tilde{R}_{xy} = \frac{N-1}{N} R_{xy} . \quad (19)$$

Можно показать, что она является и состоятельной.

Смещение устраняется с переходом от  $\tilde{R}_{xy}$  к  $\tilde{R}'_{xy} = \frac{N-1}{N} \tilde{R}_{xy}$ .

При этом вместо (17) имеем 
$$\tilde{R}'_{xy} = \frac{1}{N-1} \sum_1^N x_i y_i - \frac{N}{N-1} \tilde{M}_x \tilde{M}_y . \quad (20)$$

Среднеквадратическое значение погрешности (18) равно среднему квадратическому отклонению оценки (20):

$$\sigma_{\tilde{R}'_{xy}} \approx \sqrt{(\tilde{R}'_{xy} + D_x D_y) / (N-1)} . \quad (23)$$

Оценка коэффициента корреляции определяется согласно

$$\tilde{r}_{xy} = \tilde{R}'_{xy} / (\tilde{\sigma}_x \tilde{\sigma}_y) . \quad (24)$$

### Определение вероятности события через его повторяемость

Экспериментальное значение вероятности  $P$  некоторого события - это его повторяемость [1-3]

$$W = \frac{n}{N} = \tilde{P} = \frac{1}{N} \sum_1^N x_i, \quad (26)$$

причем число  $n$  появлений события в серии из  $N$  испытаний можно рассматривать как сумму  $N$  независимых случайных слагаемых:

$$W = \frac{1}{N} \sum_1^N X_i, \quad (27)$$

каждое из которых может принимать только два значения 1 и 0 с вероятностями  $P$  и  $1 - P$ .

Математическое ожидание и дисперсия случайной величины  $X_i$ :

$$M_{x_i} = P; \quad D_{x_i} = P(1 - P). \quad (28)$$

Погрешность оценки (26) равна

$$\Delta W = W - P. \quad (29)$$

Математическое ожидание погрешности и ее дисперсия:

$$M[\Delta W] = 0; \quad D(\Delta W) = P(1 - P)/N = D(W). \quad (30)$$

Таким образом, оценка (26) - несмещенная и состоятельная. Среднее квадратическое отклонение оценки (26)

$$\sigma_w = \sqrt{P(1 - P)/N}.$$

На практике принимают

$$\sigma_w \approx \sqrt{W(1 - W)/N}. \quad (31)$$

*Определение законов распределения случайной величины (строим гистограмму).*

Если случайная величина  $X$  - дискретная, то определяются  $\tilde{M}_x$ ,  $\tilde{D}_x$  и оценки  $\tilde{P}_{x_i}$  значений функции вероятности  $P(x_i)$  или оценки  $\tilde{F}(x_i)$  значений функции распределения  $F(x_i)$ .

Если случайная величина  $X$  - непрерывная, то определяются  $M_x$ ,  $D_x$  и оценки  $f_x(x)$ ,  $F_x(x)$  плотности вероятности  $f_x(x)$  и функции распределения  $F_x(x)$ .

При оценивании законов распределения непрерывной случайной величины процесс обработки экспериментальных данных - реализаций  $x_1, \dots, x_N$ , начинается с выбора границ  $a$  и  $b > a$  интервала, заключающего возможные значения  $X$ , и деления этого интервала на  $k$  равных элементарных промежутков  $c = (b - a) / k$ .

При расчете  $c$  значения  $a$  и  $b$  следует для удобства округлять, принимая, например, вместо  $b = 3,341$ ,  $a = -2,63$  значения  $3,4$  и  $-2,7$ . Во всех случаях округление производится в сторону увеличения разности  $b-a$ . Значение  $k$  выбирается в пределах от 8 до 20. Удобно принять  $k=10$ .

После этого определяют границы  $x'_j$  всех элементарных промежутков и составляют таблицу (табл.1), в которой  $x'_0 = a$ ,  $x'_k = b$ . Значение  $\tilde{n}_j$  - это число реализаций  $X$ , оказавшихся в пределах  $j$ -ого интервала от  $x'_{j-1}$ , до  $x'_j$ . Значения  $\tilde{P}_j$  и  $\tilde{F}_j$ :

$$\tilde{P}_j = \tilde{n}_j / N \approx P(x'_{j-1} < X < x'_j); \quad (32)$$

$$\tilde{F}_j = \tilde{P}_1 + \dots + \tilde{P}_{j-1} = P(X < \tilde{x}_j). \quad (33)$$

При группировке реализаций  $X$  по отдельным интервалам может оказаться что некоторые из них придутся точно на границу двух смежных промежутков. В этих случаях необходимо прибавить к числам  $\tilde{n}_j$  и  $\tilde{n}_{j+2}$  смежных интервалов по 1/2.

Таблица 1

$x_i$	$x'_0$	$x'_1$	$x'_2$	...	$x'_{k-1}$	$x'_k$
$n_j$		$\tilde{n}_1$	$\tilde{n}_2$	...	$\tilde{n}_k$	
$\tilde{P}_j$		$\tilde{P}_1$	$\tilde{P}_2$	...	$\tilde{P}_k$	
$\tilde{F}_j$		$\tilde{F}_1$	$\tilde{F}_2$	...	$\tilde{F}_k$	

По данным таблицы могут быть построены эмпирические гистограмма и график функции распределения.

Затем возникает весьма сложная задача подбора аналитического закона распределения, достаточно хорошо согласующегося с результатами эксперимента. Основанием для выбора аналитического выражения плотности вероятности  $f_x(x)$  могут служить соображения о том, чтобы простейшие числовые характеристики теоретической случайной величины были равны экспериментальным значениям этих характеристик. Если, например, теоретический закон определяется двумя параметрами, то их выбирают так, чтобы совпали два момента ( $m_1, \mu_2$ ).

Для оценки существенности или несущественности расхождения между теоретическим и эмпирическим распределениями используют различные критерии

### ***Критерий интервальных оценок***

Располагая результатами эксперимента согласно (31) рассчитывают средние квадратические отклонения:  $\sigma_{\tilde{P}_j} = \sqrt{\tilde{P}_j(1-\tilde{P}_j)N}$ ;  $\sigma_{\tilde{F}_j} = \sqrt{\tilde{F}_j(1-\tilde{F}_j)N}$ . (34)

Согласно (8) рассчитываются доверительные интервалы  $\Delta_{\tilde{P}_j} = 3\sigma_{\tilde{P}_j}$ ;  $\Delta_{\tilde{F}_j} = 3\sigma_{\tilde{F}_j}$  и границы изменения ВВХ  $\tilde{P}_j \pm \Delta_{\tilde{P}_j}$ ;  $\tilde{F}_j \pm \Delta_{\tilde{F}_j}$ , (35)

соответствующие доверительной вероятности  $p_0 = 0,9972$  и  $\Phi^{-1}(0,4986) = 3$ .

Располагая выбранным аналитическим выражением плотности вероятности

$f_x(x)$ , рассчитываются теоретические значения:  $P_j = P(x'_{j-1} < X < x'_j) = \int_{x'_{j-1}}^{x'_j} f_x(x) dx$ ; (36)  
 $F_j = P(x'_{j-1} < X < x'_j) = \int_{x'_{j-1}}^{x'_j} f_x(x) dx$

Критерием согласия теоретического и экспериментального распределения

является соблюдение неравенств:

$$\begin{aligned} \tilde{P}_j - \Delta_{\tilde{P}_j} < P_j < \tilde{P}_j + \Delta_{\tilde{P}_j}; \\ \tilde{F}_j - \Delta_{\tilde{F}_j} < F_j < \tilde{F}_j + \Delta_{\tilde{F}_j}. \end{aligned} \quad (37)$$

**Критерий**  $\chi^2$

Рассчитав  $P_j$  согласно (35), находят значения  $n_j = NP_j$  (38)

и рассчитывают

$$\chi^2 = \sum_1^k (\tilde{n}_j - n_j)^2 / n_j. \quad (39)$$

Если расхождение между экспериментальным и теоретическим распределением несущественно, то распределение случайной величины (39) близко к нормальному с математическим ожиданием  $M_{\chi^2} = s$  и средним квадратическим отклонением  $\sigma_{\chi^2} = \sqrt{2s}$ , где  $s$  - так называемое *число степеней свободы* и согласно (8) с доверительной вероятностью  $p_0 = 0,997$  справедливо неравенство

$$(\chi^2 - s) / \sqrt{2s} < 3. \quad (40)$$

Число степеней свободы  $s = k - u$  - это разность между числом интервалов  $k$ , выбираемых произвольно, и числом условий  $u$ , которым должно удовлетворять эмпирическое распределение случайной величины. Этих условий обычно три: сумма всех  $\tilde{P}_j$  равна единице, математическое ожидание равно  $\tilde{M}_j$ , дисперсия равна  $\tilde{D}_x$ .

### **Заключение.**

В заключение отметим, что все используемые в настоящее время методы анализа случайных последовательностей не выходят за рамки представленного общего подхода, однако в некоторых случаях позволяют существенно упростить процедуры их классификации если учитывают специфику анализируемых последовательностей.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 5*

### **Численные методы оптимизации.**

#### СОДЕРЖАНИЕ

Введение

Учебные вопросы:

1. Структура и постановка задач оптимизации.
2. Условия оптимальности и типы вычислительных процедур оптимизации.
3. Методы одномерной оптимизации
4. Поиска экстремума функций многих переменных. Метод Гаусса-Зайделя.
5. Метод наискорейшего спуска.

Заключение



## Литература:

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.
4. Таха Х. Введение в исследование операций . В двух книгах .Пер. с англ. М.: Мир 1985.

## **Введение**

*Оптимизация* в широком смысле слова - это поиск лучшего из возможных вариантов.

*Методы оптимизации* – это численные методы решения задач оптимизации (задач, имеющих множество допустимых решений, из которых необходимо выбрать одно, лучшее в каком-либо смысле). Численные методы оптимизации, как методы численного (приближенного) программирования основаны на поисковых регулярных либо случайных процедурах выбора одного из множества возможных путей достижения экстремумов критериев.

### ***Структура и постановка задач оптимизации***

*Формулировка задачи оптимизации* включает три этапа:

1. словесное представление о параметрах задачи, множестве ее решений и поставленной цели;
2. запись критерия оптимальности (целевой функции) как функции параметров задачи;
3. запись условий, определяющих область допустимых значений параметров.

Параметры задачи  $x_i$ , ( $i=1,2,\dots,n$ ) - это переменные, значения которых необходимо определить в результате ее решения

Критерий оптимальности может быть представлен в виде функции параметров - целевой функции -  $f(X)$ , где  $X=(x_1, x_2, \dots, x_n)$  – одно из допустимых решений задачи, или функционала  $I(X)$ , который при фиксированных значениях параметров задачи представляет собой не число, а функцию времени или пространственной координаты, например,

$$f(X) = \sum_{i=1}^n (x_i - x_i^{(0)})^2, \quad I(X) = \int_0^t \varphi(X, t) dt.$$

Область допустимых значений параметров  $x_i$ , ( $i=1,2,\dots,n$ ) может быть определена с помощью ограничений -  $g_j(X) \leq 0$ ; ( $j=1,2,\dots,m$ ) - условия типа неравенств и связей -  $h_k(X) = 0$ ; ( $k=1,2,\dots,p$ ) - условия типа равенств.

Таким образом, обобщенная постановка оптимизационной задачи имеет вид: найти решение  $X^*=(x_1^*, x_2^*, \dots, x_n^*)$ , которому соответствует экстремум функции  $f(X)$  - минимум или максимум, при выполнении условий:

$$g_j(X) \leq 0; (j=1,2,\dots,m) \quad (1)$$

$$h_k(X) = 0; (k=1,2,\dots,p) \quad (2)$$

Задачи оптимизации классифицируют по следующим признакам:

1. наличие или отсутствие ограничений и связей - задачи на
  - *безусловный экстремум* - условия (1), (2) отсутствуют и на
  - *условный экстремум* - имеется условие (1), условие (2) или оба;
2. вид критерия оптимальности:
  - *вариационные задачи* (критерий — функционал) и
  - *задачи математического программирования* (критерий - функция);
3. характер функций  $f, g, h$ 
  - *задачи линейного программирования* (все функции - линейные)
  - *задачи нелинейного программирования* (хотя бы одна из функций - нелинейная);
4. характер параметров
  - *непрерывные значения*
  - если параметры задачи могут принимать только строго определенные значения то это задача *дискретного (целочисленного) программирования*,
  - а если число этих значений конечно – *комбинаторная задача*.

## **Условия оптимальности и типы вычислительных процедур оптимизации**

Теоретической основой методов решения задач оптимизации являются условия оптимальности решения различных типов задач, т.е. условия, при которых критерий оптимальности той или иной задачи достигает минимального или максимального значения. Условия оптимальности подразделяются на необходимые и достаточные.

*Примеры необходимых условий оптимальности* - условие А необходимо для выполнения В, если при выполнении В всегда выполняется А):

- если в окрестности некоторой точки  $x^*$  ( $x \in [(x^* - \delta), (x^* + \delta)]$ ,  $\delta > 0$ )  $f(x) > f(x^*)$ , то в точке  $x = x^*$  функция  $f(x)$  имеет локальный минимум (в окрестности другой точки  $x$  значения  $f(x)$  могут быть еще меньше);
- если дифференцируемая функция  $f(x)$  имеет в точке  $x = x^*$  экстремум, то  $f'(x^*) \equiv 0$  для функции многих переменных ( $df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0$ ,  $i = 1, 2, \dots, n$ ).

## Примеры достаточных условий оптимальности (условие А достаточно

для выполнения В если при выполнении А всегда выполняется В):

- если в точке  $x=x^*$  функция  $f(x)$  имеет минимум, то в некоторой окрестности этой точки  $x^*$  ( $x \in [(x^*-\delta), (x^*+\delta)]$ ,  $\delta > 0$ ) выполняется условие  $f(x) > f(x^*)$ ;

- если  $f'(x^*)=0$  и  $f''(x^*) \neq 0$ , то в точке  $x^*$   $f(x)$  имеет экстремум (при  $f''(x^*) > 0$  – минимум, при  $f''(x^*) < 0$  – максимум);

- для того, чтобы дифференцируемая функция  $f(x_1, x_2, \dots, x_n)$  имела в точке  $(x_1^*, x_2^*, \dots, x_n^*)$  минимум достаточно, чтобы  $(df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0, i=1, 2, \dots, n)$  и все миноры матрицы Гессе  $\{ d^2f(x_1^*, x_2^*, \dots, x_n^*)/dx_1^2 \dots d^2f(x_1^*, x_2^*, \dots, x_n^*)/(dx_1 dx_n) ; \dots ; d^2f(x_1^*, x_2^*, \dots, x_n^*)/(dx_n dx_1) \dots d^2f(x_1^*, x_2^*, \dots, x_n^*)/dx_n^2 \}$  были положительны, а если миноры нечетного порядка отрицательны, а четного - положительны, то в точке  $(x_1^*, x_2^*, \dots, x_n^*)$  функция  $f(x_1, x_2, \dots, x_n)$  имеет максимум.

После того как сформулированы условия оптимальности критерия конкретной задачи, необходимо организовать вычислительную процедуру поиска оптимального решения. Имеется **пять основных типов вычислительных процедур решения задач оптимизации**, на основе которых разрабатываются методы оптимизации:

**Сравнение значений целевой функции на сетке значений аргументов.** Сетка образуется в результате разбиения областей допустимых значений аргументов на равные интервалы (рис.1). Оптимальному решению соответствует минимальное или максимальное значение целевой функции в “узлах” сетки. Процедура обычно применяется многократно: вначале шаг сетки “крупный”, а затем вокруг лучшей точки строится более “мелкая” сетка. Аналогом этой процедуры для задач дискретного программирования и комбинаторных является поиск лучшего допустимого решения путем полного или локального перебора. Эта процедура применяется всегда, когда это не связано со слишком большим объемом вычислений.

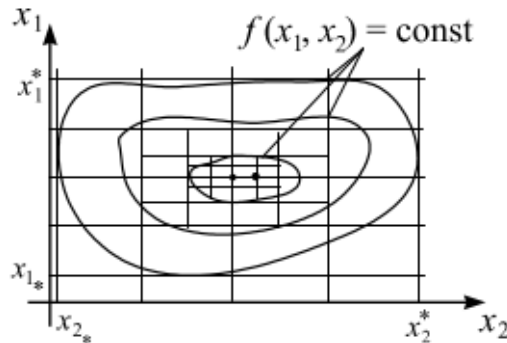


Рисунок 1. Сетка значений аргументов функции  $f(x_1, x_2)$

**Использование необходимых условий экстремума** целевой функции, т.е. формирование и решение систем уравнений вида  $(df(x_1^*, x_2^*, \dots, x_n^*)/dx_i = 0, i=1, 2, \dots, n)$ . Применение этой процедуры ограничено, т.к. аналитические выражения производных целевой функции существуют далеко не всегда.

**Использование достаточных условий оптимальности:** образуется вспомогательная задача, множество решений которой шире допустимого, а критерий оптимальности на допустимом множестве совпадает с критерием исходной задачи - например задача условной оптимизации заменяется задачей безусловной оптимизации, в критерий которой вводится "штраф" за выход из допустимой области, т.е. за невыполнение условий (1), (2).

**Отсечение множеств заведомо неоптимальных решений** (рис. 2) на основе правил, различных для каждой конкретной задачи (метод "золотого сечения", ветвей и границ для экстремальных комбинаторных задач).

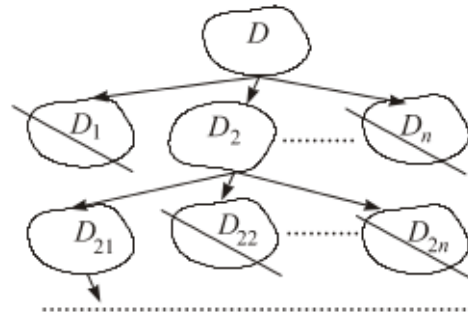


Рисунок 2. Отсечение множеств неоптимальных решений.



## *Построение оптимизирующей последовательности допустимых решений*

задачи: выбирается одно из допустимых решений  $X(0)$ , называемое начальным приближением, и на его базе строится последовательность допустимых решений  $X(0), X(1), X(2), \dots, X(n)$ , где  $X(k+1) = X(k) + \Delta X(k)$ ,  $k=0, 1, \dots, n$ , причем  $\Delta X(k)$  выбирается так, чтобы при поиске  $\min f(X)$  выполнялось условие  $f(X(k)) > f(X(k+1))$ , а при поиске  $\max f(X)$  – условие  $f(X(k)) < f(X(k+1))$ . Приращение  $\Delta X(k)$  чаще всего является функцией одного или нескольких предыдущих членов последовательности:  $\Delta X(k) = \varphi(X(k))$  или  $\Delta X(k) = \varphi(X(k), X(k-1))$ . Построение последовательности заканчивается в момент выполнения необходимых условий  $\text{ext } f(X)$ . Большинство методов оптимизации основаны именно на этой процедуре.

## ***Методы одномерной оптимизации***

К числу наиболее популярных численных методов поиска экстремума функции одного переменного одномерной оптимизации относятся: метод «золотого сечения» и пошаговый метод. Первый из них ориентирован на поиск  $\text{ext } f(x)$  внутри фиксированного интервала  $(a, b)$  оси  $x$ , последний - на поиск  $\text{ext } f(x)$  в окрестности заданной точки  $x_0$ .

### ***Метод «золотого сечения»***

Определение: «золотым сечением» отрезка называется его деление на две части таким образом, что отношение длины отрезка к его большей части равно отношению большей части к меньшей. Следовательно, для отрезка единичной длины:  $1/t=t/(1-t)$

$$\Rightarrow t^2+t-1=0 \Rightarrow t=-1/2 \pm \sqrt{(1/4+1)};$$

$$|t|<1 \Rightarrow t=(\sqrt{5}-1)/2=0.618; \quad (1-t)=0.382$$

**Алгоритм метода «золотого сечения»** при поиске минимума функции  $f(x)$  включает операции, см. рис. 3:

- 1) интервал  $(a, b)$  делится точками  $x_1, x_2$  в отношении «золотого сечения»:  
 $x_1 = a + (b - a) \times 0.382$ ,  $x_2 = b - (b - a) \times 0.382$ ;
- 2) вычисляются значения  $f(x_1)$  и  $f(x_2)$ ;
- 3) если  $f(x_1) < f(x_2)$ , то от интервала  $(a, b)$  отсекается его правая часть:  $b = x_2$ , в противном случае – левая:  $a = x_1$ ;
- 4) в случае  $b = x_2$  точка  $x_1$  осуществляет «золотое сечение» нового интервала  $(a, b)$  и играет в нем роль точки  $x_2$ , а точка  $x_1$  нового интервала определяется аналогично п.1); при  $a = x_1$  наоборот –  $x_1 = x_2$ , а точка  $x_2$  нового интервала  $(a, b)$  определяется как в п.1).

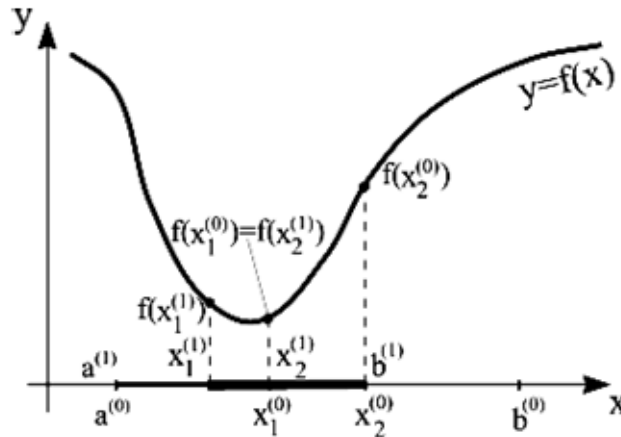


Рисунок 3. Метод «золотого сечения».

Для нового интервала  $(a, b)$  вновь выполняются действия п.п.2)-4), причем в п.2) значение функции  $f(x)$  вычисляется один раз: только для вновь определяемой точки  $x_1$  или  $x_2$ .

Процесс деления интервала продолжается до тех пор, пока его длина не станет меньше заданной точности:  $b-a < \epsilon$ . При завершении процесса поиска за точку минимума принимается значение  $x^* = (a+b)/2$ .

Достаточные условия сходимости алгоритма метода «золотого сечения»:

- функция  $f(x)$  непрерывна внутри интервала  $(a, b)$ ;
- $f(x)$  унимодальна на интервале  $(a, b)$ , т.е. имеет внутри него единственный экстремум;
- в некоторой окрестности искомой точки  $x^*$   $f(x)$  является монотонной (с одной стороны возрастает, с другой — убывает).

## Пошаговый метод

Этот метод применяется в тех случаях, когда интервал  $(a, b)$  оси  $x$ , содержащий точку экстремума функции  $f(x)$  неизвестен, но известно, что экстремум находится в окрестности экспериментально найденной точки  $x_0$ . Этот метод применяется на практике значительно чаще метода «золотого сечения», так как условие сходимости его алгоритма намного проще: для этого достаточно, чтобы функция  $f(x)$  была непрерывна в окрестности точки  $x_0$ .

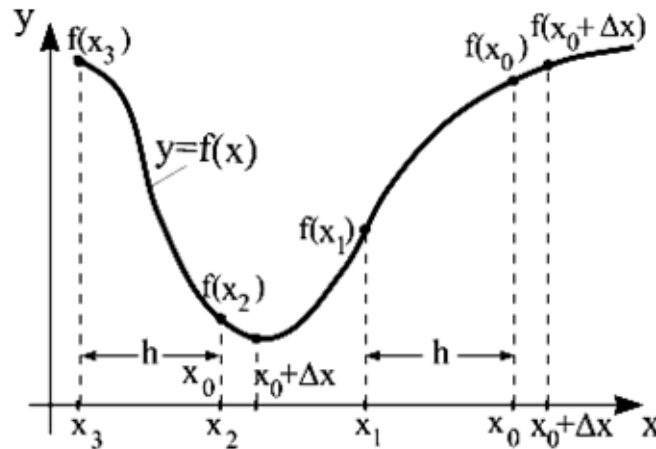


Рисунок 4. Пошаговый метод.

При поиске минимума функции **пошаговый метод** содержит следующее, см. рис. 4:

- 1) выполняется пробный шаг от точки  $x_0$  с целью выбора направления поиска:  $x = x_0 + \Delta x$  ( $\Delta x \sim 0.5 \times \varepsilon$ ) и вычисляются значения  $f(x_0), f(x)$ ;
- 2) если  $f(x) < f(x_0)$ , то величина основного шага, с которым осуществляется движение в направлении убывания функции, положительна ( $h > 0$ ), в противном случае – отрицательна ( $h < 0$ );
- 3) движение в выбранном направлении с шагом  $h$ :  $x_{k+1} = x_k + h$ ,  $k = 0, 1, 2, \dots$  осуществляется до тех пор, пока  $f(x_{k+1}) < f(x_k)$ ;
- 4) если  $f(x_{k+1}) \geq f(x_k)$ , то при выполнении условия  $h < \varepsilon$  процесс поиска заканчивается, а если  $h \geq \varepsilon$ , то шаг дробится:  $h = |h|/p$ ,  $p > 1$  и осуществляется возврат к п.1) с начальной точкой  $x_0 = x_k$ .

В качестве коэффициента дробления шага  $p$  используют 2, 3, 5, но чаще всего  $p = e = 2.71828$ . По завершении процесса поиска за точку экстремума принимается значение  $x^* = (x_{k+1} + x_k)/2$ .

## *Поиска экстремума функций многих переменных*

Несмотря на большой выбор приближенных методов вычислений далее рассмотрим три основополагающих из них: метод Гаусса-Зайделя; метод наискорейшего спуска и методы случайного поиска.

### *Метод Гаусса-Зайделя.*

Данный метод использует информацию об унимодальности функционала качества, что позволяет существенно сократить число просматриваемых точек по сравнению с полным перебором. Суть метода заключается в эквивалентной замене общей многопараметрической задачи поиска экстремума последовательностью однопараметрических задач поиска частных экстремумов. Частная производная оптимизируемого функционала имеет вид:

$$dI(\vec{x})/dx_i = dI(x_1, \dots, x_i, \dots, x_n)_{x_{l \neq i} = \text{const}} / dx_i, i, l = 1, \dots, I. \quad (3)$$

при этом оптимальное значение  $x_i^{opt}$  может быть найдено из условия:

$$dI(\vec{x})/dx_i=0, x_i=x_i^{opt}. \quad (4)$$

Как видно из выражения (3) поиск оптимальных значений параметров  $x_i^{opt}$ , удовлетворяющих экстремуму  $\partial I(\vec{x})/\partial \vec{x}=0$ , может быть осуществлен на основе итеративной последовательной процедуры оптимизации по каждому  $i$ -му параметру при фиксированных значениях остальных  $l$ -х параметров. Сходимость такой процедуры к оптимальному решению по всем оптимизируемым переменным гарантируется при наличии унимодальности и дифференцируемости целевой функции.



**Алгоритм вычисления экстремальных координат критерия** включает следующие шаги:

- выбор начальной оптимизируемой переменной  $x_1$  и отыскание частного экстремума ( $\min, \max$ )  $\partial I(\vec{x})/\partial x_1=0$  при фиксированных значениях остальных переменных;
- фиксация значения  $x_1=x_1^{opt}$  и осуществление поиска экстремума по переменной  $x_2$ , до обращения в нуль частной производной  $\partial I(\vec{x})/\partial x_2=0$  ;
- фиксация значений переменных  $x_1$  и  $x_2$  на уровне частных экстремумов и поиск оптимальных значений оставшихся переменных;
- переход к повторному циклу поиска частных экстремумов (шагу 1) до тех пор, пока найденная точка экстремума окажется общей для всех переменных (момент выполнения условий по точности решения задачи оптимизации для всех переменных).

**Пример 1.** Пусть критерий оптимальности имеет следующий вид:

$$\min_{x_1, x_2} I(x_1, x_2) = x_1^2 + x_2^2 + 1,5x_1x_2.$$

**Шаг\_1.** Поиск начинается из точки  $x_1=3$ ,  $x_2=3$ . Начальной оптимизируемой переменной выбирается  $x_1$ . Тогда экстремальное ее значение определяется из условия

$$\partial I(x_1, 3) / \partial x_1 = 2x_1 + 4,5 = 0,$$

т.е.  $x_1^{opt} = -2,25$ .

**Шаг\_2.** Находим минимум по  $x_2$ .

$$\partial I[-2,25, x_2] / \partial x_2 = (5 + x_2^2 - 3,4x_2)' = 2x_2 - 3,4 = 0,$$

откуда  $x_2^{opt} = 1,7$ .

**Шаг\_3.** Повторяем цикл вычислений для координаты  $x_1$  при фиксации  $x_2 = x_2^{opt} = 1,7$ . Получаем  $x_1^{opt}(2) = -1,27$ .

Далее цикл вычислений повторяется для переменной  $x_2(2)$  и так далее до достижения заданной точности  $\varepsilon$  приближения оптимальных решений к точке общего экстремума ( $x_1^{opt}(\infty) = 0$ ,  $x_2^{opt}(\infty) = 0$ ).

## *Метод наискорейшего спуска*

### *Градиентный метод*

Для нахождения оптимального вектора  $\vec{x}'$  используют следующую рекуррентную схему:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \alpha^{(k)} \vec{L}^{(k)}; k=0, 1, 2, \dots \quad (5)$$

В выражении (5)  $\vec{x}^{(k+1)}$  и  $\vec{x}^{(k)}$  – значения вектора  $\vec{x}$  на  $k$ -м и  $(k+1)$ -м шаге итерационной процедуры;  $\alpha^{(k)}$  – множитель, определяющий длину шага;  $\vec{L}^{(k)}$  – вектор, определяющий направление движения от точки  $\vec{x}^{(k)}$  к точке  $\vec{x}^{(k+1)}$ . В градиентном методе в качестве  $L^{(k)}$  выбран градиент целевой функции  $\nabla f(\vec{x})$  в случае максимизации  $f(\vec{x})$  и антиградиент  $\vec{L}^{(k)} = -\nabla f(\vec{x})$  в случае минимизации  $f(\vec{x})$ . В качестве начала итераций выбирается точка  $\vec{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ . Далее будем предполагать, что решается задача минимизации целевой функции.

Решение задачи градиентным методом заключается в следующем. Находится вектор  $\vec{L}^{(0)} = -\nabla f(\vec{x}^{(0)})$ , тогда луч  $\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}$  (где  $\alpha \geq 0$ ) определяет направление, вдоль которого скорость уменьшения  $f(\vec{x})$  будет наибольшей. При этом шаг  $\alpha^{(0)} = \alpha$  предполагается постоянным. При этом его выбор производится таким образом, чтобы выполнялось условие

$$f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}) < f(\vec{x}^{(0)}) . \quad (6)$$

Если это условие не выполняется, то производится коррекция длины шага, например  $\alpha^{(0)} = \frac{\alpha}{2}$ , и опять проверяется выполнение неравенства. Получив точку  $\vec{x}^{(1)}$ , в которой значение функции будет меньше, чем в точке  $\vec{x}^{(0)}$ , начинаем следующую итерацию с начальной точкой  $\vec{x}^{(1)}$  и длиной шага  $\alpha^{(1)} = \frac{\alpha}{2}$ . Процесс завершается в точке, где  $\nabla f(\vec{x}) = 0$  либо приращение значений переменных на соседних шагах оптимизации не превышает ранее заданной величины погрешности вычислений.

Рассмотрим **пример** решения задачи оптимизации градиентным методом. Пусть есть целевая функция  $\min_{x_1, x_2} I(x_1, x_2) = x_1^2 + x_2^2 + 1,5x_1x_2$ , начальные условия  $\vec{x}^{(0)} = (3, 3)$ , шаг итераций  $\alpha = \alpha^0 = 1$ .

1. Найдем градиент функции в точке  $\vec{x}^{(0)} = (3, 3)$

$$\left. \frac{\partial f(x)}{\partial x_1} \right|_{\vec{x}^{(0)}} = (2x_1 + 1,5x_2) \Big|_{\vec{x}^{(0)}} = 10,5;$$

$$\left. \frac{\partial f(\vec{x})}{\partial x_2} \right|_{\vec{x}^{(0)}} = (2x_2 + 1,5x_1) \Big|_{\vec{x}^{(0)}} = 10,5.$$

$\nabla f(\vec{x}^{(0)}) = (10,5; 10,5)$ , так как  $f(\vec{x}) \rightarrow \min$ , следовательно,  $\vec{L}^{(0)}$

$$= -\nabla f(\vec{x}^{(0)}) = (-10,5; -10,5).$$

2. Уравнение луча имеет вид:  $\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}$

$$=(3;3) + 1 \cdot (-10,5; -10,5) = (-7,5; -7,5).$$

3. Проверим выполнение неравенства  $f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}) < f(\vec{x}^{(0)})$ , для чего произведем вычисление  $f(\vec{x}^{(0)})$  и  $f(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)})$ .

$$f\left(\vec{x}^{(0)}\right) = 31,5, \quad f\left(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}\right) \approx 196,9.$$

Учитывая, что неравенство не выполняется, произведем изменение длины шага  $\alpha = \alpha^{(0)} = \frac{1}{2}$ .

В этом случае  $f\left(\vec{x}^{(0)} + \alpha^{(0)} \vec{L}^{(0)}\right) = f(-2,25; -2,25) \approx 17,7 < f\left(\vec{x}^{(0)}\right)$ , следовательно,

$$\vec{x}^{(1)} = (-2,25, -2,25).$$

Повторяем рассуждения, взяв за начальную точку  $\vec{x}^{(1)} = (-2,25; -2,25)$  и длину шага  $\alpha^{(1)} = \frac{1}{2}$ .

$$\vec{L}^{(1)} = -\nabla f\left(\vec{x}^{(1)}\right) = (7,9; 7,9), \quad \alpha = \frac{1}{2}. \quad \text{В этом случае } f(\vec{x}^{(1)} + \alpha^{(1)}\vec{L}^{(1)}) = 10,1 <$$

$$f\left(\vec{x}^{(1)}\right), \text{ следовательно } \vec{x}^{(2)} = (1,7; 1,7)$$

Продолжим вычисления для точки  $\vec{x}^{(2)} = (1,7; 1,7)$

$$\vec{L}^{(2)} = -\nabla f\left(\vec{x}^{(2)}\right) = (-5,44; -5,44), \quad \alpha = \frac{1}{2}. \quad \text{В этом случае}$$

$$f\left(\vec{x}^{(2)} + \alpha^{(2)}\vec{L}^{(2)}\right) = 3,5 < f\left(\vec{x}^{(2)}\right), \text{ следовательно } \vec{x}^{(3)} = (-1; -1),$$

Эти операции повторяются до тех пор, пока  $\nabla f(\vec{x}) = 0$ .

Для ускорения процесса поиска оптимального вектора  $\vec{x}'$  может быть использован *метод наискорейшего спуска*. Основное отличие этого метода от градиентного состоит в способе выбора шага  $\alpha^{(k)}$ , а поиск оптимального вектора осуществляется с помощью следующей итерационной процедуры:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \alpha^{(k)} \frac{\nabla f(\vec{x}^{(k)})}{\|\nabla f(\vec{x}^{(k)})\|}, \quad (7)$$

где  $\|\nabla f(\vec{x}^{(k)})\| = \sqrt{\sum_{i=1}^n \left(\frac{\partial f(\vec{x}^{(k)})}{\partial x_i}\right)^2}$  – норма градиента целевой функции в точке  $\vec{x}^{(k)}$ , а все остальные элементы выражения (7) соответствуют введенным ранее.

То есть, в отличие от градиентного метода, в методе наискорейшего спуска шаг спуска  $\alpha^{(k)}$  определяется из решения уравнения

$$\frac{\partial f(\vec{x}^{(k)} + \alpha^{(k)} \vec{L}^{(k)})}{\partial \alpha^{(k)}} = 0. \quad (8)$$

Метод хорошо "работает" при минимизации гладких функций и если начальное приближение выбрано достаточно далеко от оптимума. Если же очередная точка  $\vec{x}^{(k)}$



окажется в окрестности оптимума, то уменьшение целевой функции будет очень медленным. Это происходит из-за того, что для получения оптимума с высокой точностью необходимо выполнить большое число мелких шагов [2].

Как видно из существа метода, условием его применения является непрерывная дифференцируемость  $f(\vec{x})$ .

**Пример 2.** Найти координаты минимума целевой функции следующего вида:

$$f(x_1, x_2) = x_1^2 + x_2^2 + 1,5 x_1 x_2.$$

Для начальной точки  $(x_1=2, x_2=3)$  находим градиент (частные производные):

$$\nabla f(\vec{x}) = \left[ \frac{\partial f(\vec{x})}{\partial x_1}; \frac{\partial f(\vec{x})}{\partial x_2} \right]^m = [2x_1 + 1,5x_2; 2x_2 + 1,5x_1]^m.$$

Определим норму градиента:

$$\|\nabla f(\vec{x})\| = \sqrt{(2x_1 + 1,5x_2)^2 + (2x_2 + 1,5x_1)^2} = \sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}.$$

Далее определяется единичный вектор, имеющий направление антиградиента:

$$\vec{L} = -\frac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|} = \left[ -\frac{2x_1 + 1,5x_2}{\sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}}; -\frac{2x_2 + 1,5x_1}{\sqrt{6,25x_1^2 + 12x_1x_2 + 6,25x_2^2}} \right].$$

Выражение, определяющее величину оптимального шага, находим из уравнения (8):

$$\frac{\partial f(\vec{x} + \alpha \vec{L})}{\partial \alpha} = \frac{\partial \left( (x_1 + \alpha L(x_1))^2 + (x_2 + \alpha L(x_2))^2 + 1,5(x_1 + \alpha L(x_1))(x_2 + \alpha L(x_2)) \right)}{\partial \alpha} = 0 .$$

Решая данное уравнение получаем:

$$\alpha = - \frac{x_1 L(x_1) + x_2 L(x_2) + 0,75 x_1 L(x_2) + 0,75 x_2 L(x_1)}{L^2(x_1) + 1,5 L(x_1) L(x_2) + L^2(x_2)} ,$$

где  $L(x_i)$  - проекция вектора  $\vec{L}$  на ось  $x_i$ .

Первый шаг:  $x_1^{(0)} = 2, x_2^{(0)} = 3$  ;

$$\nabla f(\vec{x}^{(0)}) = [2x_1^{(0)} + 1,5x_2^{(0)}; 2x_2^{(0)} + 1,5x_1^{(0)}]^m = [8,5; 9]^m ;$$

$$\|\nabla f(\vec{x}^{(0)})\| = \sqrt{6,25(x_1^{(0)})^2 + 12x_1^{(0)}x_2^{(0)} + 6,25(x_2^{(0)})^2} \approx 12,38 ;$$

$$\vec{L}^{(0)} = \left[ -\frac{8,5}{12,38} ; -\frac{9}{12,38} \right]^m = [-0,69; -0,73]^m$$

$$\alpha^{(0)} \approx 3,5 ;$$

$$x_1^{(1)} = x_1^{(0)} + \alpha^{(0)} L(x_1^{(0)}) = 2 - 3,5 \cdot 0,69 = -0,415 ;$$

$$x_2^{(1)} = x_2^{(0)} + \alpha^{(0)} L(x_2^{(0)}) = 3 - 3,5 \cdot 0,73 = 0,445 .$$

Второй шаг:  $x_1^{(1)} = -0,415, x_2^{(1)} = 0,445$  ;

$\nabla f(\vec{x}^{(1)}) = [-0,16; 0,24]^m$  ;  $\|\nabla f(\vec{x}^{(1)})\| \approx 0,32$  ;  $\vec{L}^{(1)} = [0,5; -0,75]^m$  ;  $\alpha^{(1)} \approx 0,56$  ;

$$x_1^{(2)} = x_1^{(1)} + \alpha^{(1)} L(x_1^{(1)}) = 0,415 - 0,56 \cdot 0,5 = -0,135 ;$$

$$x_2^{(2)} = x_2^{(1)} + \alpha^{(1)} L(x_2^{(1)}) = 0,445 - 0,56 \cdot 0,75 = 0,025 .$$

Сравнение результатов применения первого и второго методов показывает большую скорость сходимости к оптимальным значениям переменных при применении метода наискорейшего спуска, чем метода Гаусса –Зайделя.

Заключение.

Регулярные и случайные методы поиска оптимальных значений переменных обладают тем основным достоинством по сравнению с аналитическими методами оптимизации, что остаются эффективными в более широком классе целевых функций (нелинейные, динамические, нестационарные).

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 6*

### **Векторный анализ эффективности процесса принятия решений, проблемы и методы их преодоления.**

#### СО Д Е Р Ж А Н И Е

##### ВВЕДЕНИЕ

##### УЧЕБНЫЕ ВОПРОСЫ:

1. Постановка задачи векторного анализа эффективности процесса принятия решений.
2. Проблемы векторного анализа эффективности процесса принятия решений в сетях связи и методы их преодоления.
3. Общий алгоритм векторного анализа эффективности функционирования сети связи.

##### ЗАКЛЮЧЕНИЕ

## Литература:

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Петухов Г. Б. Основы теории эффективности целенаправленных процессов. – М.: МО СССР, 1989.
4. Терентьев В.М., Санин Ю.В. Анализ эффективности функционирования автоматизированных сетей многоканальной радиосвязи. - С-Петербург: ВАС, 1992.
5. Терентьев В. М., Илюхин А. А., Куцакин А. И., Осипов А. Н. Основы построения сетей спутниковой связи с подвижными объектами: Учебное пособие. – Орел: Академия Спецсвязи России, 2004.

## *Введение*

Проблемам оценки эффективности сложных целенаправленных систем, систем обмена информацией и принятия решений посвящено большое количество работ, анализ которых позволяет выделить несколько этапов в развитии теории эффективности. К начальному этапу можно отнести период до середины 70-х годов, когда анализ эффективности функционирования систем передачи информации шел по пути оценки отдельных их свойств и где результатом оценки эффективности выступали оценки отдельных ПК сетей. Начиная с 1975 года, анализ эффективности функционирования начал проводиться как комплексная оценка совокупности свойств сетей связи. Окончательно теория эффективности, как самостоятельное направление в науке, оформилась к началу 80-х годов. Применительно к системам связи, наиболее продуктивной явилась методология вероятностного анализа эффективности, развитая в работах Г.Б.Петухова [3].

## *Постановка задачи векторного анализа эффективности процесса принятия решений.*

Понятие «эффективность (результативность) операции», совершаемой объектом, связано с целым рядом составляющих его частных понятий:

- объект,
- процесс функционирования (операция),
- цель,
- задача,
- свойство,
- качество  $(x)$  ,
- показатель качества ,
- система показателей качества и требования к ним,
- показатель эффективности  $\left( \Phi \left( \vec{Y}(x); \vec{Y}_{\text{тр}} \right) \right)$  ,
- критерий оценки эффективности  $\left( I \left( \Phi \left( \vec{Y} \right); \Phi_{\text{тр}} \right) \right)$  .

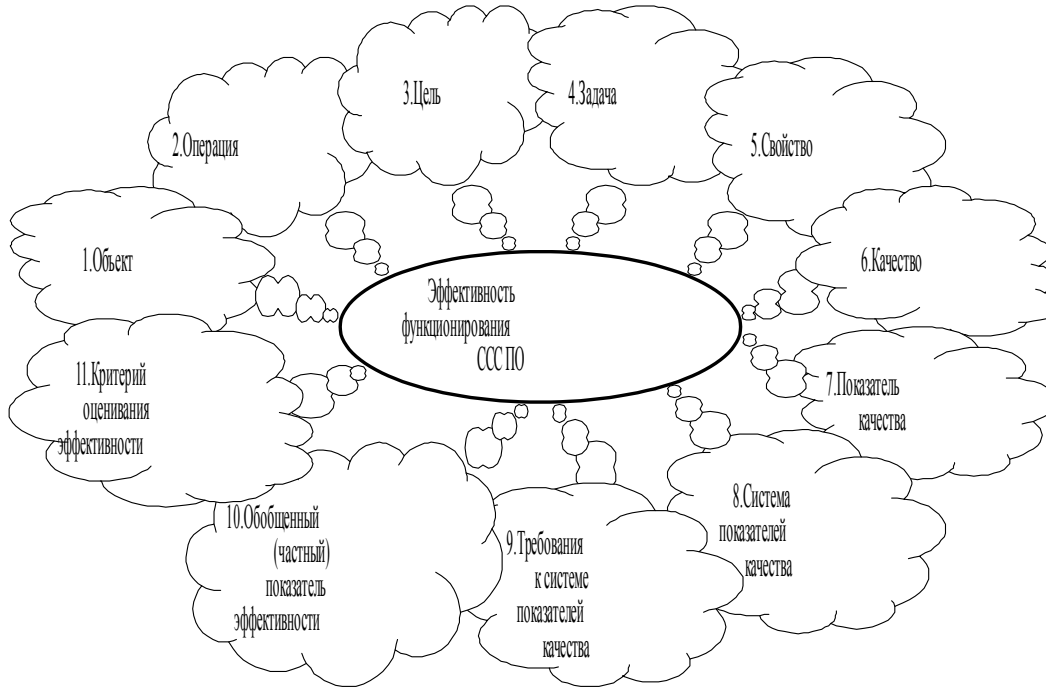


Рисунок 1. Основные понятия теории эффективности решений.

Обобщенный показатель эффективности функционирования (ПЭФ) системы по всем показателям качества - вероятность достижения цели  $P_{\text{ВЫП Ф}} = P(Y_i \in Y_i^{\text{ДОП}})$ .



*Критерий оценивания эффективности* – совокупность условий, определяющих цели анализа эффективности: пригодность, превосходство или оптимальность исследуемого процесса функционирования ССС ПО.

При оценивании качества любого объекта, описываемого  $n$ -мерным векторным показателем, реализуется совокупность критериев, каждый из которых в общем случае может принадлежать одному из классов:

- классу  $\{G\}$  критериев пригодности;
- классу  $\{S\}$  критериев превосходства;
- классу  $\{0\}$  критериев оптимальности.

Многокритериальный характер требований к качеству связи и управления, учет протекающих в системе связи процессов приводят к постановке векторной задачи анализа - эффективности функционирования систем связи.

***Проблемы векторного анализа эффективности процесса  
принятия решений и методы их преодоления.***

Анализ этапов развития методов оценки эффективности позволяет сделать вывод о достаточной общности, объективности и возможности проведения аналитических расчетов большой размерности, основанных на векторном подходе к формированию ОПЭФ.

Наличие множества различных и зачастую противоречивых частных показателей качества СС порождает следующие основные проблемы векторного подхода к анализу эффективности ее функционирования:

- необходимость *сокращения исходной размерности* вектора частных показателей качества,
- их *нормирование* и последующую *свертку (скаляризацию)* в обобщенный показатель эффективности функционирования (ОПЭФ) [1].

*Процесс нормализации* включает этапы перехода к единой размерности (безразмерности), сведения к одной точке отсчета и переход к равноценным шкалам (одному масштабу). Достаточно полно все перечисленные этапы могут быть выполнены при использовании линейного преобразования частных показателей качества

$$\vec{Y}_n^i(\vec{x}; k) = c_n Y_n(\vec{x}; k) + d_n, \quad (1)$$

где  $c_n = \frac{1}{Y_n^{\max}(\vec{x}; k) - Y_n^{\min}(\vec{x}; k)}$  – масштабный коэффициент;

$$d_n = \frac{Y_n^0(\vec{x}; k)}{Y_n^{\max}(\vec{x}; k) - Y_n^{\min}(\vec{x}; k)} - \text{коэффициент сдвига, корректирующий начало}$$

отсчета;

$\vec{Y}_n^i$ ,  $Y_n^{\max}$ ,  $Y_n^{\min}$  – нормированное, наибольшее и наименьшее значения показателя качества.

Использование данного преобразования позволяет:

- *привести все показатели к нулевой точке отсчета,*
- *их изменение ограничивается отрезком  $[0, 1]$ ,*
- *делает все показатели безразмерными.*

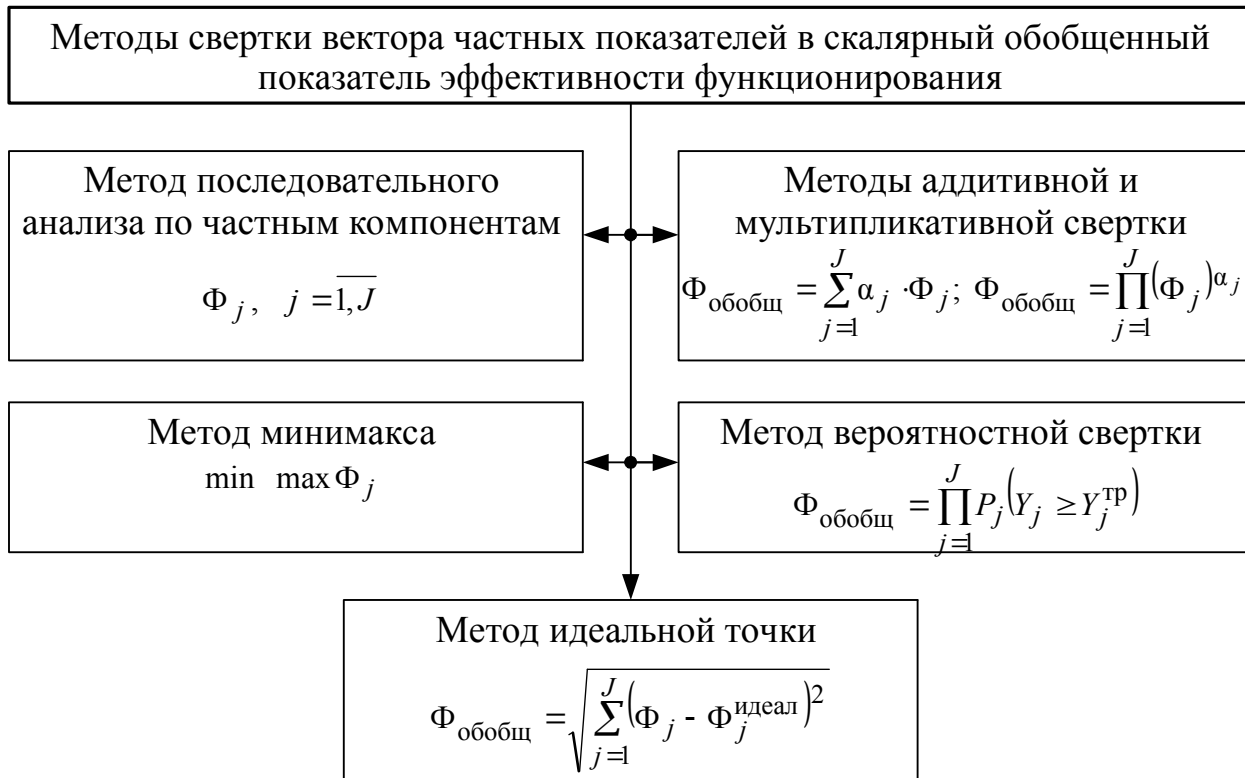


Рис. 2. Основные методы свертки вектора ЧПЭФ в скалярный ОПЭФ

**Общий алгоритм анализа эффективности** функционирования СС включает в себя следующие этапы:

- *определение системы показателей качества (СПК) СС, отражающей все ее основные свойства при использовании по назначению (обычно формируются три группы показателей: временные, показатели основного функционального эффекта и экономические);*
- *формулировку оперативно-технических требований (ОТТ) к каждому из показателей качества;*
- *нормализацию компонентов пространства показателей качества, введение меры в пространство критериев и выбор метода свертки векторного показателя эффективности в обобщенный скалярный показатель эффективности;*
- *проведение оценки частных показателей технико-экономической эффективности и обобщенного показателя эффективности функционирования сети;*
- *формулировку выводов об эффективности (технико-экономической целесообразности) принятых решений и определение выигрыша в технико-экономических показателях сети при использовании разработанных решений.*

Конкретизируем основные этапы общего алгоритма оценки эффективности функционирования применительно к СС.

1. В СПК функционирования СС должны входить четыре иерархически связанных частных системы показателей качества (СПК): СПК ИО, СПК У, СПК СИО, СПК СУ. Учитывая определяющую значимость СПК ИО для оценки эффективности СС, следует рассмотреть ее более детально.

Процесс информационного обмена в сетях связи характеризуется такими свойствами (качествами), как достоверность, своевременность, безопасность, экономичность. Каждое из этих свойств может быть определено количественно с помощью отдельного показателя качества (ПК).

2. Следующим этапом оценки эффективности функционирования СС является определение требований (пороговых или предельных значений) к показателям качества функционирования  $\vec{Y}'$  [4,5]. В основу формирования системы оперативно-технических требований к СС могут быть положены разработанные МСЭ-Т (*ITU*) рекомендации и стандарты.
3. Третьим этапом оценки является этап формирования частных показателей эффективности функционирования СС.

С учетом случайного характера большинства факторов, определяющих эффективность функционирования сети связи частные показатели эффективности ее функционирования (ЧПЭФ) определяются как совместная вероятность достижения предъявляемых к показателям качества требований:

$$P_{\text{вып } \phi} = P(Y_{\phi} \leq \vec{Y}'_{\phi}) = \int_0^{Y'_{\phi 1}} \dots \int_0^{Y'_{\phi r}} \omega(Y_{\phi 1}, \dots, Y_{\phi r}) d\vec{Y}_{\phi}, \quad (2.10)$$

где  $\vec{Y}_{\phi}$  – СПК функционирования СС;

$Y'_{\phi i}$  – требуемое значение вектора ПК СС;

$\omega(Y_{\phi 1}, \dots, Y_{\phi m})$  – совместная плотность распределения вероятностей значений вектора показателей качества  $\vec{Y}_{\phi}$ .

Обычно ЧПЭФ СС формулируются в виде вероятностей выполнения следующих требований к качеству услуг:

- по достоверности (надежности связи), т. е. вероятности того, что величина отношения сигнал/шум на входе линейного устройства не будет меньше допустимого значения  $P(h_{c/ш}^2 \geq h'_{c/ш}{}^2)$  (2.11) во всех направлениях сети;
- своевременности, т.е. вероятности того, что время доставки сообщения в любом направлении сети не превысит предельного значения  $P(t_{\partial} \leq t'_{\partial})$  (2.12);
- помехозащищенности, т. е. вероятности того, что отношение сигнал/шум в линии связи любого направления связи будет меньше предельного значения  $P(h_{рп}^2 < h'_{рп}{}^2)$ ; (2.13);
- экономичности, т. е. вероятности того, что затрачиваемые сетью физические или экономические ресурсы на производство услуг (на бит передаваемой информации в любом из направлений связи) не превысят заданных значений  $P(\vec{Z}_c \leq \vec{Z}'_c)$ . (2.14).



Основная трудность использования вероятностной свертки заключается в необходимости априорного задания многомерных плотностей распределения вероятностей значений показателей качества и их многократном интегрировании, что значительно упрощается при переходе к аппарату условных вероятностей [1,3].

Несомненным преимуществом метода вероятностной скаляризации является автоматическое решение при его применении проблемы нормализации пространства частных показателей эффективности.

Заключительным этапом оценки эффективности является формулировка и оценка обобщенного показателя эффективности функционирования (ОПЭФ) СС.

Это определяет целесообразность использования ОПЭФ, полученного на основе введения в векторное нормированное пространство ЧПЭФ евклидовой метрики – суммы квадратов отклонений ЧПЭФ от идеальных (требуемых) их значений. При этом обобщенный показатель эффективности функционирования записывается в виде [1]:

$$\Phi_{\text{обобщ}}^{\text{эф}} = \sqrt{\sum_{j=1}^J (\Phi_j - \Phi_j^{\text{идеал}})^2}, \quad (2.15)$$

где  $\Phi_j$  – значение  $j$ -го ЧПЭФ СС;

$\Phi_j^{\text{идеал}}$  – идеальное значение  $j$ -го ЧПЭФ СС.

## **Вывод**

Задачу оценки эффективности функционирования автоматизированных систем связи (СС) целесообразно решать в следующие два этапа.

*Этап 1 – анализ частных вероятностных показателей эффективности функционирования СС:*

$P(t_{\partial} \leq t'_{\partial})$  – вероятности своевременной доставки;

$P(h_{\text{с/ш}}^2 \geq h_{\text{с/ш}}^{2'} = f_d^{-1}(P'_{\text{ош.б}}))$  – надежности связи;

$P(h_{\text{рп}}^2 < h_{\text{рп}}^{2'})$  – помехозащищенность;

вероятности обеспеченности ресурсами (полосой частот и мощностью, пространственными и временными ресурсами, прочими материальными и экономическими ресурсами).

*Этап 2 – анализ обобщенного показателя эффективности функционирования СС, полученного на основе метода идеальной точки.*

Данное решение позволяет сохранить достаточно высокую чувствительность ОПЭФ к отдельным ПК при большой размерности СПК.

## *Общий алгоритм векторного анализа эффективности функционирования сети связи*

В алгоритме можно выделить следующие группы блоков:

- вспомогательных расчетов;
- анализа частных вероятностных показателей эффективности функционирования СС;
- анализа ОПЭФ, полученного на основе метода идеальной точки,
- формулировки выводов по критерию пригодности.

## **Заключение**

Овладение методами оценки эффективности функционирования ИТКС позволяет оценить вклад отдельных ее подсистем в выполнение задач, поставленных перед ИТКС в целом, оптимизировать подход к решению задач обоснования оперативно-технических требований, разработки структуры автоматизированной системы управления ИТКС и распределения ресурсов ИТКС между ее подсистемами.

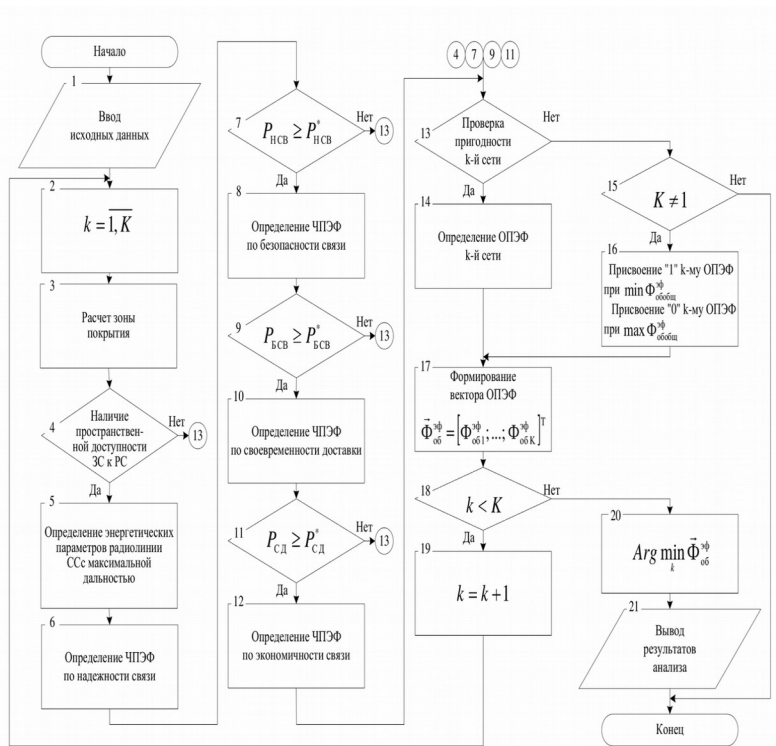


Рисунок 3. Обобщенный алгоритм векторного анализа эффективности функционирования СС.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 7-1-2*

### **Априорная неопределенность вероятностных моделей в задачах принятия решений. Методы динамического программирования.**

#### СОДЕРЖАНИЕ

Введение

Учебные вопросы:

1. Уровни априорной неопределенности относительно статист. характеристик.
2. Основные методы преодоления априорной неопределенности при принятии статистических решений.
3. Характеристика многошаговых распределительных задач.
4. Методы динамического программирования. Постановка задачи прямой и обратной прогонки.
5. Методика реализации принципа оптимальности.
6. Метод множителей Лангранжа для задач с ограничениями в форме равенств.
7. Задачи нелинейного программирования с ограничениями в форме неравенств. Условия Куна-Таккера.

Заключение

## Литература:

1. Теория передачи сигналов: Учебник для вузов/ Зюко А.Г., Кловский Д.Д., Назаров М.В., Финк Л.М. \_М.: Связь, 1980.
2. Терентьев В.М. , Парашук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. СПб.: ВАС 1995.
3. Левин Б.Р. Теоретические основы статистической радиотехники.- М.: Радио и связь, 1989.
4. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
5. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
6. Таха Х. Введение в исследование операций . В двух книгах .Пер. с англ. М.: Мир 1985 .

## Введение

Полнота вероятностной математической модели определяется наличием всех априорных статистических данных относительно вида и параметров плотностей или функций распределения, представляемых ею реальных процессов моделируемого объекта. В случае отсутствия ряда данных о модели говорят о наличии априорной неопределенности, преодоление которой является существом задач математической статистики – построения алгоритмов обработки наблюдаемых выборочных данных с целью получения оптимальных статистических выводов.

Множество задач с подобной постановкой исходных данных в качестве выборочных параметров и построением обобщенных моделей применяются в реальных коммуникационных технических системах и сетях.

Коммуникационное оборудование почти всегда в реальных условиях работает в априорной параметрической неопределенности — это обусловлено как объемами проходящего трафика, так и незнанием параметров исходных распределений сигналов и помех, которые могут оказывать значительное влияние на качественные показатели устройств. Способы решения задач и методы преодоления подобных априорных неопределенностей различны и зависят от вида самой неопределенности.



## ***Уровни априорной неопределенности относительно статистических характеристик.***

*Основными понятиями, определяющими принятие статистических решений в условиях априорной неопределенности, являются:*

- **пространство наблюдений**  $\vec{z}(\vec{x}(t), t) \in Z$  – совокупность всех реализаций наблюдаемого случайного процесса  $\vec{x}(\vec{\theta}, t)$ ;
- **вероятностная мера на пространстве наблюдений** – совместная плотность (функция) распределения выборочных значений случайного процесса  $w(\vec{z}/\vec{\theta})$  - функция правдоподобия, условная по вектору априорно неизвестных параметров модели  $\vec{\theta}$ ;
- **априорное распределение неизвестных параметров**  $w(\vec{\theta})$ .

В зависимости от полноты априорных сведений о виде распределения случайного процесса  $\vec{x}(t)$  и его параметрах различают следующие **уровни априорной неопределенности**:

- **параметрический**, когда вид распределения  $w(\vec{x}(t), t)$  является заданным, а сведения о его параметрах, например, о среднем либо дисперсии в условиях задачи отсутствуют;
- **непараметрический** – нет априорных сведений ни о виде распределения, ни о его параметрах.

*Наличие априорной неопределенности в задаче принятия статистических решений приводит к необходимости проведения дополнительного по сравнению с традиционным подходом этапа обработки наблюдаемых данных (обучения):*  
**идентификации параметров модели** - в случае параметрического уровня и  
**идентификации вида распределения** – в случае непараметрической постановки задачи.

Этап обучения позволяет преодолеть априорную неопределенность за счет формирования оценок неизвестных заранее функций распределения либо их параметров на основе информации, заложенной в выборочных данных. При этом *выборки могут быть классифицированными (тестовыми) и не классифицированными.*

Восстановление априорно неизвестных данных по выборке приводит к возможности приспособления (адаптации) оптимальных для условий полной статистической определенности алгоритмов принятия решений к новым условиям путем использования полученных оценок распределений и их параметров вместо истинных.

Поэтому под *адаптивными алгоритмами* в дальнейшем *будем понимать алгоритмы обработки и принятия решений, использующие обучение (преодоление априорной неопределенности) на основе выборочных данных.*

Общим критерием качества работы адаптивных алгоритмов является их сходимость к алгоритмам, оптимальным в условиях полной априорной определенности.

***Методы принятия решений в условиях риска.***

***Принятие решений при известных априорных вероятностях.***

Будем обозначать вероятности гипотез:  $Q_1=p(P_1), Q_2=p(P_2), \dots, Q_n=p(P_n)$ . Таким образом, мы считаем, что вероятности известны до того, как мы решили принять решение.

Решение – выбор оптимальной стратегии  $A$ . говорят, что это ситуация идеального наблюдателя.

Естественно, в качестве критерия выбирается средний выигрыш, который мы получим, если выберем стратегию  $A_i$ .  $\bar{a}_i = \sum_{j=1}^n a_{ij} Q_j$ . Это аналог математического ожидания. Решение принимается по критерию:  $\max_i \bar{a}_i$  - критерий максимального среднего выигрыша - критерий Лапласа [L].

Если задана матрица рисков, то можно для каждой строки вычислить средний риск:  $\bar{r}_i = \sum_{j=1}^n r_{ij} Q_j$  и оптимальным будет являться решение, которое  $\min_i \bar{r}_i$ .

Слабым местом этого подхода является то, что *надо знать априорные вероятности*. Если они неизвестны, то необходимо их изучить. Это можно делать путём экспериментов, которые учитывают условия природы. Говорят, что мы эту систему обучаем. Такой подход называется принципом адаптации к условиям.

Если априорные вероятности изучить не удаётся, то применяется принцип недостаточности основания: *если не знаем о вероятности, то считаем, что гипотезы природы равновероятны*.

После этого применяем критерий идеального наблюдателя. Так как при равных вероятностях энтропия (неопределённость) максимальна, то мы применяем принцип пессимизма. Если сами значения вероятности неизвестны, но есть информация о предпочтениях гипотез, то существуют методы обработки предпочтений и получения вероятностей.

В некоторых случаях учитывается не только средний выигрыш, но также и дисперсия, т. е. величина разброса выигрыша в каждой строке:

$$\max_i (\bar{a}_i - k \times \sigma_{a_i}); \quad \sigma_{a_i} = \sqrt{D_{a_i}}$$

## **Принятие решений при неизвестной априорной информации.**

Если априорная информация неизвестна или ненадёжна, то применяются другие критерии:

1) **Критерий Вальда** ( $W$ ) : это максимально-минимальный критерий, т.е. выбирается максимальная стратегия  $A_k$ , для которой  $W = \max_i \min_j a_{ij}$ . Мы подходим к этой задаче, рассчитывая на самый худший случай.

2) **Критерий Сэвиджа** ( $S$ ) : это критерий минимаксного риска:  $S = \min_i \max_j r_{ij}$ . Этот критерий не эквивалентен критерию Вальда, т.е. оптимальный по Сэвиджу не обязательно так же будет эквивалентен по Вальду.

3) **Критерий Гурвица** ( $H$ ) : это комбинированный критерий, его так же называют критерием пессимизма-оптимизма:  $H = \max_i \left( \kappa \min_j a_{ij} + (1 - \kappa) \max_j a_{ij} \right)$ ;  $0 < \kappa < 1$  - коэффициент, который выполняет требования критерия быть более или менее оптимистичным. При  $\kappa = 1 \rightarrow H \rightarrow W$ , а при  $\kappa = 0$ ,  $H \rightarrow L$  крайний оптимизм. Существуют другие критерии и однозначный выбор одного критерия невозможен.

## ***Основные методы преодоления априорной неопределенности при принятии статистических решений.***

Для случаев когда объект или модель работает в условиях динамического изменения существенно влияющих на нее параметров и/или правил необходим переход к адаптивным алгоритмам формирования управлений, который предполагает лишь модификацию этапа стохастического оптимального оценивания состояния объекта управления путем добавления к алгоритму оценки процедуры предварительной идентификации неизвестных ранее данных. Ниже рассмотрим наиболее употребляемые алгоритмы идентификации для случаев параметрической и непараметрической априорной неопределенности относительно вероятностной модели объекта.

## *А. Алгоритм оценки – идентификации на основе расширенного фильтра Калмана.*

Пусть для случая полной априорной определенности модель состояния и наблюдения за объектом записывается в виде:

$$\begin{aligned}x(k+1) &= -\alpha(k)x(k) + b(k)u(k) + g(k)v(k), \\z(k) &= h(k)x(k) + w(k), x(0) = \bar{x}.\end{aligned}\tag{1}$$

Здесь  $\alpha(k)$  неизвестный априори параметр, заданный плотностью распределения вероятностей его значений  $N(\bar{\alpha}, \sigma_{\alpha}^2)$  и экспоненциальной функцией корреляции.

Введем расширенный вектор состояния объекта управления  $\vec{x}(k) = (x(k), \alpha(k))^T$ , объединяющий все компоненты, подлежащие как оценке, так и идентификации. При этом уравнения состояния и наблюдения примут вид:

$$\begin{aligned}\vec{x}(k+1) &= F(k)\vec{x}(k) + B(k)u(k) + G(k)\vec{v}(k), \\z(k) &= H(k)\vec{x}(k) + \vec{w}(k), x(0) = \bar{x},\end{aligned}\tag{2}$$

где  $F(k), B(k), G(k), H(k)$  - матрицы состояния, эффективности управления, возбуждения и наблюдения соответствующих размерностей;

$\vec{v}(k), \vec{w}(k)$  - взаимонезависимые векторные шумы возбуждения и наблюдения соответственно, являющиеся белыми гауссовскими последовательностями с нулевыми средними и диагональными матрицами дисперсий  $\sigma_{\vec{v}}^2, \sigma_{\vec{w}}^2$ .

Как видно из выражения (2) задача идентификации – оценивания приведена к классической задаче линейной фильтрации и может быть решена на основе использования векторного алгоритма фильтрации Калмана, обеспечивающего минимум среднеквадратичного отклонения оценок неизвестного параметра  $\hat{\alpha}(k)$  и состояния объекта  $\hat{x}(k)$  от их истинных значений:

$$\vec{\hat{x}}(k) = F(k)\vec{\hat{x}}(k-1) + B(k)u(k-1) + K(k)\left[\vec{z}(k) - H^T(k)(F(k)\vec{\hat{x}}(k-1) + B(k)u(k-1))\right], \quad (3)$$

где  $K(k) = P_2(k)H^T(k)\sigma_w^{-2}$  - коэффициент фильтра Калмана, рассчитываемый для линейного случая заранее, т.к. не зависит от текущих оценок;

$P_2(k)$  - матрица дисперсий ошибок оценивания, элементы которой для установившегося состояния работы фильтра определяются на основе ранее приводимых выражений.

Достоинством данного метода адаптации является универсальность и высокая точность алгоритма идентификации - оценивания, а недостатком - значительные вычислительные затраты в связи с ростом размерности вектора состояния.



## В. Алгоритм идентификации на основе процедуры стохастической аппроксимации Робинса - Монро.

Пусть наблюдения за априори неизвестным параметром  $\theta(k)$  представляют собой временной ряд  $\vec{z}(k)=(z(0), \dots, z(k), \dots, z(K))^T$ , каждый элемент которого описывается уравнением наблюдения  $z(k)=\theta(k)+w(k)$ .

Тогда условное среднее параметра  $M[\theta(k)/z(k)]$  может быть определено в соответствии с рекуррентной процедурой стохастической аппроксимации Робинса – Монро:

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\gamma(z(k)-\hat{\theta}(k-1)), \quad (4)$$

где  $\gamma=\frac{1}{k+1}$ ,  $\gamma'=\frac{1}{k+\sigma_w^2/P_2(k)}$  - весовые коэффициенты фильтра, удовлетворяющие

условиям Дворецкого по сходимости данного алгоритма;  $P_2(k)=\frac{\sigma_w^2}{(k+\sigma_w^2/\sigma_\theta^2(0))}$  -

апостериорная пошаговая дисперсия идентифицируемого параметра  $\theta(k)$ ;

$\sigma_w^2, \sigma_\theta^2(0)$  - априорно задаваемые дисперсии шума наблюдения и параметра  $\theta$ .

Нетрудно видеть, что при использовании второго весового коэффициента сходимость и точность оценки при высоких отношениях сигнал /шум становятся выше, однако это требует дополнительных по сравнению с первым случаем априорных данных о дисперсиях шума и идентифицируемого параметра.

Достоинством алгоритма является его простота, а недостатком низкая точность в силу низкого уровня учитываемых априорных данных об идентифицируемом параметре.

***С. Минимаксный алгоритм оценки состояния объекта при непараметрическом уровне априорной неопределенности модели.***

Пусть уровень непараметрической неопределенности относительно статистики оцениваемого параметра  $x(k)$  задан совокупностью плотностей распределения  $\{w_i(x)\}, i=1, \dots, n$ , ни одна из которых в отдельности не может надежно описывать реальную плотность распределения. Задача непараметрической оценки процесса  $x(k)$  в данной ситуации может быть решена на основе следующего минимаксного критерия:

$$\min_{\hat{x} \in X} \max_{w_i(x)} \left[ \int w_i(x) \int L(x, \hat{x}) w(z/x) dz dx \right], \quad (5)$$

где  $w(x), w(z/x)$  – плотность распределения параметра  $x$  и функция правдоподобия, соответственно;

$L(x, \hat{x})$  – функция потерь.

## ***Д. Алгоритмы идентификации плотностей и функций распределения для непараметрических уровней априорной неопределенности статистических характеристик.***

При неизвестных видах априорных распределений параметров объекта, подлежащих оценке необходимые для принятия решения эмпирические распределения могут быть получены непосредственно из наблюдений. Близость полученных эмпирических распределений известным аналитическим распределениям может быть проверена по критериям согласия следующего вида.

### **1. Критерий Колмогорова - Смирнова.**

Обеспечивает получение оценки апостериорного распределения, оптимальной в смысле минимальной верхней границы:

$$I_1[\hat{w}(\vec{x})] = \inf_{\hat{w}(x)} \sup_{\vec{x} \in X} |w(\vec{x}) - \hat{w}(\vec{x})| \quad (6)$$

по всем значениям модулей их разности с распределением заданного класса  $w(x)$  для заданного ряда значений переменной состояния  $x$ .

### **2. Критерий Ренье (относительный критерий согласия).**

$$I_2[\hat{w}(\vec{x})] = \inf_{\hat{w}(x)} \sup_{\vec{x} \in X} \left| \frac{w(\vec{x}) - \hat{w}(\vec{x})}{\hat{w}(\vec{x})} \right| \quad (7)$$

### **3. Критерий Мизеса (интегральный критерий согласия).**

$$I_3[\hat{w}(\vec{x})] = \inf_{\hat{w}(x)} \sup_{\vec{x} \in X} \int [w(\vec{x}) - \hat{w}(\vec{x})]^2 \Psi[w, \vec{x}] dw(\vec{x}) \quad (8)$$

где  $\Psi(w, \vec{x})$  - вес квадрата отклонения эмпирической и гипотетической плотности для всего ряда значений переменной состояния  $x$  от первого до последнего.

*В качестве статистик при решении задач непараметрического оценивания могут быть использованы знаковые, порядковые либо ранговые.*

Пусть  $\vec{z}=(z_1, \dots, z_n)^T$  наблюдаемая выборка. Введем знаковую функцию

$$\operatorname{sgn} z = \frac{z}{\sqrt{|z|}} = \begin{cases} 1, & z \geq 0, \\ -1, & z < 0. \end{cases} \quad (9)$$

Тогда знаковым вектором выборки  $\vec{z}$  назовем вектор  $\operatorname{sgn} \{ \vec{z} = (\operatorname{sgn} z_1, \dots, \operatorname{sgn} z_n)^T$ , образующий пространство всех знаковых векторов с общим числом точек  $2^n$ . Наконец, *произвольная функция компонент знакового вектора является знаковой статистикой*, а использующий ее алгоритм знаковым алгоритмом. Примером применения знаковой статистики является ее использование в задаче проверки симметричности относительно нуля априорно неизвестного распределения по знаку получаемых выборочных данных.

Если элементы выборки  $\vec{z}$  перегруппировать по мере возрастания их значения, то получим упорядоченную выборку  $\vec{z}'$  – вариационный ряд  $\vec{z}' = (z^1, \dots, z^n)$ , где  $-z^k \leq z^j - n p_{j-k} < j$ , называемую *вектором порядковых статистик*. *Произвольную функцию от данного вектора либо его элемента называют порядковой статистикой*. Совокупность порядковых статистик обычно представляет простую марковскую последовательность, если она была образована на основе однородной независимой выборки размером  $n$  из распределения, имеющего непрерывную плотность.

Рангом  $R_i$  элемента выборки  $z_i$  называется номер этого элемента в вариационном ряду  $\vec{z}' = z_1, \dots, z_n$ . Ранговым вектором  $\vec{R}(\vec{z}) = (R_1, \dots, R_n)^T$  выборки  $\vec{z}$  называется перестановка чисел  $1, 2, \dots$ , которая получается при замене элементов выборки их рангами. *Произвольная функция от рангового вектора называется ранговой статистикой.*

Исходную выборку  $\vec{z}$  можно восстановить, зная вектор порядковых статистик  $\vec{z}'$  и ранговый вектор  $\vec{R}$ . Ранг элемента  $z_i$  выборки размера  $n$  может быть определен на основе функции единичного скачка:

$$R_i = \sum_{k=1}^n u(z_i - z_k), i = 1, \dots, n,$$

$$u(z_i - z_k) = \begin{cases} 1, & z_i \geq z_k, \\ 0, & \text{во всех остальных случаях.} \end{cases}$$

Так как для однородной независимой выборки функция правдоподобия инвариантна к группе перестановок ее аргументов, то все ранговые векторы для указанной выборки равновероятны ( т.е. попадание рангового вектора  $R_i$  в любую точку  $r_i, i=1, \dots, n!$  пространства возможных перестановок равна  $1/n!$ ) независимо от вида распределения, к которому принадлежит выборка. Это обстоятельство определило целесообразность использования ранговых статистик в задаче принятия решения об однородности и независимости выборки в непараметрической ее постановке.

## ***Выводы***

Проблема априорной неопределенности относительно статистических характеристик оцениваемых параметров состояния управляемого объекта разрешается переходом к адаптивным алгоритмам принятия решений. При этом, в зависимости от уровня неопределенности, используются методы параметрической, либо непараметрической идентификации неизвестных априори статистических характеристик на основе обработки выборочных данных (данных наблюдения) в виде различного вида статистик.

Степень различия решений, полученных на основе адаптивных алгоритмов, к решениям, полученным из традиционных оптимальных алгоритмов, зависит от качества выполнения этапа идентификации и асимптотически стремится к нулю при увеличении размера выборки.

## *Методы динамического программирования*

Повышение эффективности вычислений при решении определенного класса задач математического программирования может быть достигнуто путем использования методов динамического программирования. Особенности методов динамического программирования являются использование для их реализации принципов инвариантного погружения и оптимальности. Принцип инвариантного погружения предполагает замену общей задачи на эквивалентную совокупность более простых (пошаговых) задач. Принцип оптимальности определяет возможность получения глобально-оптимальных стратегий (решений) на основе решений пошаговых задач оптимизации. Методы динамического программирования позволяют существенно сократить (по сравнению с полным перебором) число анализируемых вариантов решений в процессе определения глобально-оптимального решения за счет учета априорной информации о решениях, не являющихся допустимыми, и использования информации, полученной на предыдущих шагах оптимизации. Кроме того, достоинством методов динамического программирования является их инвариантность к классу целевой и ограничительных функций.

## ***Характеристика многошаговых распределительных задач.***

В распределительных задачах с большим числом различных результатов производственной деятельности ( $i=1, n$ ) и видов ресурсов ( $j=1, m$ ) общее решение задачи оптимизации может быть при определенных условиях заменено совокупностью последовательно решаемых менее сложных частных задач оптимизации, например, по каждому из отдельных видов производственной деятельности. При этом важными понятиями ДП являются: *последовательность шагов* оптимизации, *состояние* системы распределения ресурсов и *варианты решения* (области изменения оптимизируемых переменных)

### ***Методы динамического программирования. Постановка задачи прямой и обратной прогонки.***

Рассмотрим сущность динамического программирования и введенных выше понятий на примере общей задачи линейного программирования:

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n \Rightarrow \max,$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2,$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m,$$

$$x_1, \dots, x_n \geq 0,$$



где  $z$ -целевая функция, подлежащая максимизации;  
 $x_i$ -оптимизируемые переменные;  
 $i=1, n$ -номер оптимизируемой переменной;  
 $c_i$ - доход от реализации  $i$ -го вида производственной деятельности;  
 $j=1, m$ -номер ограничений на значения переменных;  
 $a_{ij}$ - коэффициенты уравнений-ограничений;  
 $b_j$ -величина  $j$ -го ресурса (правая часть ограничений).

Здесь каждый вид производственной деятельности  $i$  может рассматриваться как отдельный шаг оптимизации; множество возможных значений переменных (допустимая область решений)  $x_i$  как варианты решений, а количество каждого  $j$ -го вида ресурса ( $B_{i1}, \dots, B_{ij}, \dots, B_{im}$ ),  $0 \leq B_{ij} \leq b_j$ , доступного для распределения на всех предыдущих и текущем (либо текущем и всех последующих) шагах (по  $i$ -м типам деятельности) как состояние модели.

Тогда оптимальное значение целевой функции  $z$  для шагов  $i, i+1, \dots, n$  при заданных состояниях  $\{B_{ij}\}$  может быть записано в виде следующей рекуррентной функции Белмана (алгоритма прямой прогонки):

$$f_i(B_{i1}, \dots, B_{im}) = \max \{c_i x_i + f_{i-1}(B_{i1} - a_{i1}x_i, \dots, B_{im} - a_{im}x_i)\}, \quad (11),$$

$$0 \leq a_{ij}x_i \leq B_{ij}$$

$$i = 1, n; j = 1, m, 0 \leq B_{ij} \leq b_j,$$

с начальными условиями  $f_0(B_{01}, \dots, B_{0m}) = 0$ .

Оптимальное значение целевой функции для случая обратной прогонки, т.е. шагов  $n, \dots, i, i-1, \dots, 1$  при заданных состояниях  $\{B_{ij}\}$  может быть записано в виде следующего алгоритма:

$$f_n(B_{n1}, \dots, B_{nm}) = \max \{c_n x_n\},$$

$$0 \leq a_{nj} x_n \leq B_{nj}$$

$$f_i(B_{i1}, \dots, B_{im}) = \max \{c_i x_i + f_{i+1}(B_{i1} - a_{i1} x_i, \dots, B_{im} - a_{im} x_i)\}, i=1, n; j=1, m, \quad (12)$$

$$0 \leq a_{ij} x_i \leq B_{ij} \quad \text{где } 0 \leq B_{ij} \leq b_j.$$

Разница между прямым и обратным способами решения задачи заключается в определении *состояния* модели. В прямой модели  $B'_{ij}$  - количество ресурса  $j$ -го типа, распределяемого от первого шага до  $i$ -го, а для обратной модели  $B_{ij}$  - количество ресурса, распределяемого на шагах от  $n$ -го до  $i$ -го.

Решение задачи (11) основывается на двух основополагающих принципах.

**Принципе инвариантного погружения**, определяющего декомпозицию решения общей задачи на пошаговое решение частных (для каждого вида производственной деятельности) задач, объединяемых общим ресурсом и значением целевой функции.

**Принципе оптимальности**, определяющем независимость решений, получаемых на каждом текущем шаге оптимизации, от решений, полученных на предыдущих (последующих) шагах, а лишь их зависимость от цели оптимизации и состояния ресурсов на  $i$ -м шаге. При этом гарантируется оптимальность глобальной стратегии (последовательности решений) при оптимальных локальных (пошаговых) решениях.

*Процесс решения задачи методом динамического программирования включает два этапа.*

*На первом этапе* пошаговые задачи оптимизации приводят к условно-оптимальным по ресурсу (состояниям) решениям и одному (конечному) безусловно-оптимальному решению.

*На втором этапе* формируется окончательная безусловно-оптимальная стратегия  $(x_1^{opt}, \dots, x_i^{opt}, \dots, x_n^{opt})$  путем учета полученного на первом этапе конечного решения и затрат ресурсов на его реализацию, а также обратного по шагам анализа множества условно-оптимальных решений и выделения из него оптимальной стратегии.

### **Выводы**

Метод динамического программирования предназначен для повышения эффективности вычислений при решении задач математического программирования путем их декомпозиции на относительно простые, а следовательно легче решаемые задачи. Принцип оптимальности является основой поэтапного решения задачи, при этом последовательность и число этапов определяются числом оптимизируемых переменных в общей задаче, возможные варианты решений допустимыми областями их определения, а состояние системы количеством ресурсов, распределяемых на текущем и предыдущих (последующих) шагах оптимизации.

### *Методика реализации принципа оптимальности.*

Большинство практических задач оптимизации связано с наличием ограничений на область допустимых значений переменных. Данные ограничения существенно уменьшают размеры области, в которой осуществляется поиск оптимума, что, однако, не приводит к упрощению задачи. Напротив, наличие ограничений на область допустимых значений аргументов создает дополнительные трудности при попытке непосредственного использования методов, связанных с отысканием стационарных точек. Указанные трудности связаны с тем, что при наличии ограничений экстремум целевой функции может находиться на краю области допустимых значений аргументов, где первая производная (градиент) от целевой функции может быть не равной нулю. Это обстоятельство приводит к необходимости анализа не только всех стационарных, но и всех точек на границе области допустимых решений.



Положим, что функция (21) и функции  $g_i(\vec{x})$  обладают непрерывными частными производными по всем своим аргументам.

В соответствии с методом множителей Лагранжа задача (21-22) преобразуется в следующую задачу безусловной оптимизации: требуется найти вектор  $\vec{x}'$ , при котором

$$L(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{j=1}^m \lambda_j g_j(\vec{x}) \rightarrow \min_{\{\vec{x}\}} (\max) \quad (23)$$

При этом функция  $L(\vec{x}, \vec{\lambda})$  называется функцией Лагранжа, а числа  $\lambda_j$  – неопределенными множителями Лагранжа, т.е. неизвестными величинами, значения которых необходимо определить. На знак  $\lambda_j$  никаких требований не накладывается.

Задача (23) может быть решена путем взятия частных производных от функции Лагранжа  $L(\vec{x}, \vec{\lambda})$  по всем переменным  $x_i (i=\overline{1, n})$  и множителям Лагранжа  $\lambda_j (j=\overline{1, m})$  и приравнивания их нулю. При этом получается система из  $n+m$  уравнений с таким же числом неизвестных, которые определяют необходимые условия наличия экстремума:

$$\begin{aligned}
& \frac{\partial L(x, \lambda)}{\partial x_1} = 0 \\
& \dots\dots\dots \\
& \frac{\partial L(x, \lambda)}{\partial x_i} = 0 \\
& \dots\dots\dots \\
& \frac{\partial L(x, \lambda)}{\partial x_n} = 0 \\
& \frac{\partial L(x, \lambda)}{\partial \lambda_1} = 0 \\
& \dots\dots\dots \\
& \frac{\partial L(x, \lambda)}{\partial \lambda_m} = 0
\end{aligned} \tag{24}$$

Решение системы (24) позволяет определить стационарные точки функции  $L(\vec{x}, \vec{\lambda})$ . В последующем должна быть организована процедура проверки на минимум или максимум. Данная процедура может быть осуществлена, например, путем анализа матрицы Гессе функции  $L(\vec{x}, \vec{\lambda})$  от аргумента  $\vec{x}$ . Стационарная точка  $\vec{x}'$  будет являться точкой минимума, если матрица Гессе в точке  $\vec{x}'$  положительно полуопределена, и точкой максимума, если матрица Гессе в данной точке отрицательно полуопределена.

### Пример.

Пусть требуется минимизировать целевую функцию  $f(x_1, x_2) = x_1^2 + (x_2 + 2)^2$  при ограничении  $g(x_1, x_2) = 2x_1 - x_2 + 4 = 0$ .

Поскольку целевая функция нелинейного типа, а ограничительная функция задана в форме равенства, для решения задачи воспользуемся методом множителей Лангранжа.

1. Составим функцию Лангранжа

$$L(x_1, x_2, \lambda) = x_1^2 + (x_2 + 2)^2 - \lambda \cdot (2x_1 - x_2 + 4).$$

2. Найдем производные от функции Лангранжа по  $x_1, x_2, \lambda$  и приравняем их к нулю:

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_1} = 2x_1 - 2\lambda = 0,$$

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_2} = 2x_2 + 4 + \lambda = 0,$$

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial \lambda} = 2x_1 - x_2 + 4 = 0.$$



Решив полученную систему уравнений, находим

$$x_1 = \frac{4}{5}; \quad x_2 = -\frac{12}{5}; \quad \lambda = \frac{4}{5}.$$

3. Проверим достаточные условия экстремума. Матрица Гессе для рассматриваемой функции Лагранжа имеет вид:

$$\nabla^2 L(\vec{x}) = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}.$$

Так как угловые главные миноры данной матрицы равны соответственно 2 и 4, рассматриваемая матрица Гессе положительно определена, следовательно,  $\vec{x}' = \left\{ \frac{4}{5}, -\frac{12}{5} \right\}$  – точка минимума.

***Задачи нелинейного программирования с ограничениями в форме  
неравенств. Условия Куна-Таккера.***

Естественным обобщением метода множителей Лагранжа является построение критериев оптимальности на случай общей задачи нелинейного программирования с ограничениями как в виде равенств, так и неравенств.

Рассмотрим вначале случай наличия только прямых ограничений, которые соответствуют требованиям неотрицательности переменных. Задача формулируется следующим образом: требуется найти вектор  $\vec{x}'$ , обеспечивающий

$$f(\vec{x}') \rightarrow \max_{\{\vec{x}\}} \quad (25)$$

при следующих ограничениях

$$g_j(\vec{x}) = 0, \quad j = \overline{1, m}, \quad (26)$$

$$x_i' \geq 0, \quad i = \overline{1, n}. \quad (27)$$

Большинство известных методов позволяют определить только локальный оптимум целевой функции. Для практики же наибольший интерес представляют задачи нахождения глобального оптимума. Поэтому важно знать условия, при которых локальный оптимум одновременно является и глобальным. Известно, что если целевая функция вогнутая (для задач максимизации) или выпуклая (для задач минимизации), и область допустимых значений выпукла, то каждый локальный оптимум является и глобальным.

Составим функцию Лагранжа для задачи (25-27):

$$L(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{j=1}^m \lambda_j g_j(\vec{x}) .$$

Тогда, при выполнении вышеназванных условий вектор определяется решением следующей системы:

$$\frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial x_i} = \frac{\partial f(\vec{x}')}{\partial x_i} - \sum_{j=1}^m \lambda_j \frac{\partial g_j(\vec{x}')}{\partial x_i} \leq 0, \quad (28)$$

$$x_i' \frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial x_i} = 0, \quad i = \overline{1, n}, \quad (29)$$

$$x_i' \geq 0, \quad i = \overline{1, n}, \quad (30)$$

$$\frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial \lambda_j} = g_j(\vec{x}') = 0, \quad j = \overline{1, m'}. \quad (31)$$

Рассмотрим теперь более общий случай, когда не только прямые, но и функциональные ограничения представлены в виде неравенств. Задача оптимизации в этом случае формулируется следующим образом: требуется найти вектор  $\vec{x}'$ , обеспечивающий

$$f(\vec{x}') \rightarrow \max_{\{\vec{x}\}} \quad (32)$$

при следующих ограничениях

$$g_j(\vec{x}) \leq 0, \quad j = \overline{1, m}, \quad (33)$$

$$x_i' \geq 0, \quad i = \overline{1, n}. \quad (34)$$

Задача (32-34) может быть сведена к задаче (25-27) путем замены функциональных ограничений в форме неравенств  $g_j(\vec{x}) \leq 0, j = \overline{1, m}$ , в эквивалентные им ограничения в форме равенств  $g_j(\vec{x}) + Z_j = 0, j = \overline{1, m}$ , путем

добавления к исходным ограничениям неотрицательных переменных  $Z_j$ . В новой задаче содержится  $n+m$  переменных, на каждую из которых наложено требование неотрицательности.

Для рассматриваемой задачи строится обобщенная функция Лагранжа

$$L(\vec{x}, \vec{\lambda}, \vec{Z}) = f(\vec{x}) - \sum_{j=1}^m \lambda_j (g_j(\vec{x}) + Z_j) = f(\vec{x}) - \sum_{j=1}^m \lambda_j g_j(\vec{x}) - \sum_{j=1}^m \lambda_j Z_j, \quad (35)$$

которая в точке оптимума численно равна обычной функции Лагранжа  $L(\vec{x}', \{\vec{\lambda}'\})'$

. При заданных ограничениях  $g_j(\vec{x}) \leq 0$  необходимым условием оптимальности является неотрицательность (неположительность)  $\lambda$  в задаче максимизации (минимизации). Этот результат устанавливается следующим образом. Рассмотрим задачу максимизации. Множители  $\lambda$  выражают скорость изменения  $f(\vec{x})$  по отношению к изменениям  $g_j(\vec{x})$ , т.е.  $\lambda_j = \frac{\partial f(\vec{x})}{\partial g_j(\vec{x})}$ . Как только правая часть ограничения  $g_j(\vec{x}) \leq 0$  увеличивается и становится большей нуля, область допустимых решений расширяется. Следовательно, оптимальное значение целевой

функции не может уменьшиться. Это означает, что  $\lambda_j \geq 0$ . Аналогично при увеличении правой части ограничения в задаче минимизации оптимальное значение  $f(\vec{x})$  не может увеличиться, откуда следует что  $\lambda_j \leq 0$ . Если же ограничения заданы в виде равенств  $g_j(\vec{x})=0$ , то на знак  $\lambda_j$  никаких условий не накладывается.

Оптимальный вектор в рассматриваемом случае находится решением следующей системы уравнений:

$$\frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial x_i} \leq 0, \quad i = \overline{1, n}, \quad (36)$$

$$x_i' \frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial x_i} = 0, \quad i = \overline{1, n}, \quad (37)$$

$$x_i' \geq 0, \quad i = \overline{1, n}, \quad (38)$$

$$\frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial \lambda_j} \leq 0, \quad j = \overline{1, m}. \quad (39)$$

$$\lambda_j' \frac{\partial L(\vec{x}', \{\vec{\lambda}'\})}{\partial \lambda_j} = 0, \quad j = \overline{1, m}' \quad (40)$$

$$\lambda_j' \geq 0, \quad j = \overline{1, m}' \quad (41)$$

Условия (36-41) для общей задачи максимизации составляют содержание теоремы Куна-Таккера.

На практике определение принадлежности целевой функции к классу выпуклых (вогнутых) функций, как правило, не представляет особого труда. Несколько сложнее оказывается задача анализа выпуклости множества допустимых решений, которая может быть решена путем исследования функций  $g_j(\vec{x})$ ,  $j=\overline{1, m}$ , составляющих ограничения.

Приведем перечень требований, достаточных для выполнения условий Куна-Таккера, полагая, что задача нелинейного программирования формулируется в самой общей постановке

$$f(\vec{x}) \rightarrow \min_{\vec{x}} (\max)$$

при следующих ограничениях

$$g_j(x) \leq 0, \quad j=1, 2, \dots, r,$$

$$g_j(x) \geq 0, \quad j=r+1, \dots, p,$$

$$g_j(x)=0, j=p+1, \dots, m.$$

При этом обобщенная функция Лангранжа представляется в виде:

$$L(\vec{x}, \vec{\lambda}, \vec{Z})=f(\vec{x})-\sum_{j=1}^r \lambda_j(g_j(\vec{x})+Z_j)-\sum_{j=r+1}^p \lambda_j(g_j(\vec{x})-Z_j)-\sum_{j=p+1}^m \lambda_j g_j(\vec{x})$$

Требования, устанавливающие достаточность условий Куна-Такера, представлены в таблице 1.

Таблица 1

Тип оптимизации	Требования		
	$f(\vec{x})$	$g_j(\vec{x})$	$\lambda_j$
Максимизация	Вогнутая	Выпуклая	$'\geq 0 \quad (1 \leq j \leq r)$
		Вогнутая	$'\leq 0 \quad (r+1 \leq j \leq p)$
		Линейная	Нет ограничений $(p+1 \leq j \leq m)$
Минимизация	Выпуклая	Выпуклая	$'\leq 0 \quad (1 \leq j \leq r)$
		Вогнутая	$'\geq 0 \quad (r+1 \leq j \leq p)$
		Линейная	Нет ограничений $(p+1 \leq j \leq m)$



Следует иметь в виду, что требования, представленные в таблице 1, охватывают не все случаи, соответствующие условиям Куна-Такера, поскольку область допустимых решений может быть выпуклой и в случаях, не соответствующих указанным в таблице требованиям.

Прямое решение задачи нелинейной оптимизации, основанное на решении уравнений Куна-Таккера, сопряжено со значительными вычислительными трудностями. Тем не менее, условия Куна-Таккера имеют важное теоретическое значение для алгоритмов решения задач нелинейного программирования.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 8*

### **Задачи выбора решений. Метод экспертных оценок. Нечеткие множества. Сетевое планирование.**

#### СО Д Е Р Ж А Н И Е

Введение

Учебные вопросы:

1. Задача выбора решений на основе метода экспертных оценок.
2. Задача выбора решений на основе аппарата нечетких множеств.
3. Основные понятия сетевого планирования. Порядок построения сетевого графика. Оценка времени завершения событий, работ и путей сетевого графика. Оптимизация параметров сетевого графика.

Заключение

## Литература:

1. Таха Х. Введение в исследование операций . В двух книгах .Пер. с англ. М.: Мир 1985.
2. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
3. Терентьев В.М. , Парашук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. С –Пб.: ВАС 1995.
4. Терентьев В. М., Илюхин А. А., Куцакин А. И., Осипов А. Н. Основы построения сетей спутниковой связи с подвижными объектами: Учебное пособие. – Орел: Академия Спецсвязи России, 2004.

## **Введение**

Задачи выбора решений составляют значительный пласт задач оптимизации, при котором для построения алгоритма решения используются различные методы.

Довольно часто в реальных задачах множество альтернатив не ясно с самого начала и требует привлечения и организации дополнительных процедур для подготовки к принятию решения. Оценка ценности каждой альтернативы требует специальных методов измерения полезности.

Кроме этого, во многих задачах, особенно в задачах автоматизации управления, присутствуют так называемые нечеткие переменные и нечеткие критерии и альтернативы, для которых рассматриваются вопросы «человекоподобного» - субъективного восприятия информации.

Для поддержки решения таких задач оптимизации применяются различные методы, рассмотрим три из них на основе:

- метода экспертных оценок;
- аппарата нечетких множеств;
- метода сетевого планирования.

### ***Задача выбора решений на основе метода экспертных оценок.***

Во многих практических задачах при принятии решения возникает принципиальная сложность оценить существующие альтернативы. Например, надо дать оценку качества конкретной физической системы: сети связи, аппаратно-программной подсистемы и пр. Количество параметров, которые влияют на оценку системы может быть так велико и они настолько разнообразны, что невозможно раз и навсегда задать какой-то способ установки соответствия между качеством системы и числом.

В этом случае используют ***метод экспертных оценок.***

В этом методе решаются следующие задачи:

- Необходимо построить множество возможных и допустимых альтернатив решения;
- Сформировать набор аспектов, существующих для оценки альтернатив;
- Определить критериальное пространство;
- Необходимо упорядочить альтернативы по аспектам;
- Получить оценку альтернатив по критериям (отобразить множество допустимых решений в критериальном пространстве).

Все эти задачи составляют часть общей задачи оценивания – сопоставления числа некоторой альтернативе.

Метод основан на использовании экспертных процедур. Общая схема экспертизы такова: саму оценку выполняют люди, специалисты в предметной области, которые называются **экспертами**. Для проведения самой экспертизы привлекается консультант. Он определяет множество альтернатив, а иногда и вспомогательное множество для экспертизы. Каждый эксперт выбирает свою оценку и передаёт её консультанту. Эта оценка обрабатывается по специальной схеме и получается единая для всех экспертов оценка для каждой альтернативы. Затем по определённому правилу выбирается оптимальная оценка консультантом. В схеме экспертизы заложен блок, который отвечает за оценку согласованности мнения экспертов или оценку компетентности экспертов.

Эксперты могут взаимодействовать друг с другом в одних видах экспертизы, либо наоборот, отделяются друг от друга в других методах. В известном методе экспертизы Делфи, экспертам «вновь» предлагаются результаты экспертизы и просят посмотреть на них и призадуматься. Этот метод устанавливает «обратную связь» при экспертизе.

В первом туре опроса *методом Дельфи* экспертам предлагаются вопросы, на которые они дают ответы без аргументирования. Полученные от экспертов данные обрабатываются с целью выделения среднего или медианы и крайних значений оценок. Экспертам сообщаются результаты обработки первого тура опроса с указанием расположения оценок каждого эксперта. Если оценка эксперта сильно отклоняется от среднего значения, то его просят аргументировать свое мнение или изменить оценку.

Во втором туре эксперты аргументируют или изменяют свою оценку с объяснением причин корректировки. Результаты опроса во втором туре обрабатываются и сообщаются экспертам. Если после первого тура производилась корректировка оценок, то результаты обработки второго тура содержат новые средние и крайние значения оценок экспертов. В случае сильного отклонения индивидуальных оценок от средних эксперты должны аргументировать или изменить свои суждения, пояснив причины корректировки.

Проведение последующих туров осуществляется по аналогичной процедуре. Обычно после третьего или четвертого тура оценки экспертов стабилизируются, что и служит критерием прекращения дальнейшего опроса.

При проведении опроса в методе Дельфи сохраняется анонимность ответов экспертов по отношению друг к другу. Это обеспечивает исключение влияния конформизма, т.е. подавления мнений за счет “веса” научного авторитета или должностного положения одних экспертов по отношению к другим.

## **Типы задач оценивания.**

**Оценивание** – составление альтернатив какого-то вектора евклидова пространства.

1) Пусть  $X$  – некоторая альтернатива в множестве альтернатив.  $X \in \Omega$ . Имеется  $m$  критериев, тогда требуется каждой альтернативе сопоставить некоторый вектор  $[f_1(X), f_2(X), \dots, f_m(X)] \in F_m$ . Это общая задача оценивания.

2) Пусть  $k_1, k_2, \dots, k_m$  – критерии, учитываемые при выборе. Эти критерии надо установить по возможности, т.е. здесь оценивается система критериев. Система этих критериев сопоставляется значимостью для альтернативы, как перестановка натуральных чисел. Это задачи ранжирования.

3) Пусть некоторое множество  $Q$  разбито на  $l$  подмножеств и для какой-то альтернативы  $X \in Q$  необходимо указать, какому подмножеству она принадлежит, т.е.  $X$  сопоставляется конкретное подмножество. Это задача классификации.

4) Пусть  $X$  – отрезок, длину которого надо измерить; т. е. отрезку надо сопоставить действительное число. Это самая простая и самая распространённая задача оценивания.



Обозначим  $\Omega$  – исходное множество допустимых оценок;  $\Omega_E$  – множество допустимых оценок для экспертов;  $L$  – тип взаимодействия между экспертами;  $Q$  – наличие обратной связи;  $\phi - (\Omega_E^N - \Omega)$  – алгоритм обработки. Все методы обработки экспертной информации можно разбить на три вида:

### 1) Статистический метод

Результаты оценок каждого эксперта можно рассматривать как реализацию некоторой случайной величины из множества и применять к ним методы математической статистики.

### 2) Алгебраический метод

Методы численных оценок альтернатив — когда задача состоит в сопоставлении оцениваемой альтернативе одного числа, - метод попарного сравнения, метод прибавления баллов.

### 3) Методы шкалирования.

Метод ранжирования - выстраивание критериев или альтернатив по важности.

Оценка каждого эксперта рассматривается, как случайная величина. Обработка производится на основе методов математической статистики, которая позволяет определить согласованность методов экспертов и значимость, т.е. качество экспертизы.

**Например:**

Дается *численная оценка* - каждой альтернативе ставится в соответствие одно число.

$\Omega = E_1, \Omega_E = E_1;$                        $L$  – эксперты изолированы;

$Q$  – обратная связь отсутствует;                       $N$  – количество экспертов:

$$\varphi(X_1, X_2, X_3, \dots, X_N) = \frac{\sum_{i=1}^N (X_i * \alpha_i)}{\sum_{i=1}^N \alpha_i} = a \quad , \text{ где}$$

$\alpha_i$  – веса экспертов; при отсутствии информации компетентности  $\alpha_i = 1$ ;

$X_i$  – числовые оценки экспертов.

**Степень согласованности экспертов** определяется выражением:

$W = D/D_{\max}$  - дисперсионный коэффициент конкордации - определяется как отношение оценки дисперсии к максимальному значению этой оценки.

Коэффициент конкордации изменяется от нуля до единицы, поскольку  $0 \leq D \leq D_{\max}$ . Коэффициент конкордации равен 1, если все ранжировки экспертов одинаковы, и равен нулю, если все ранжировки различны.

Максимальное значение дисперсии равно  $D_{\max} = \frac{d^2(m^3 - m)}{12(m - 1)}$

Введем обозначение  $S = \sum_{i=1}^m \left( \sum_{s=1}^d r_{i,s} - \bar{r} \right)^2$

Используя  $S$ , запишем оценку дисперсии  $D$  в виде  $D = \frac{S}{(m - 1)}$ , где  $r_{is}$  - ранг,

присваиваемый  $s$ -м экспертом  $i$ -му объекту,  $\bar{r} = \frac{1}{m} \sum_{i=1}^m r_i$  - оценка математического ожидания,  $m$ -объектов или туров,  $d$ -экспертов.

*Компетентность* – степень квалификации эксперта в определенной области знаний. Компетентность может быть определена на основе анализа плодотворной деятельности специалиста, уровня и широты знакомства с достижениями мировой науки и техники, понимания проблем и перспектив развития. Для количественной оценки степени компетентности используется коэффициент компетентности, с

учетом которого взвешивается мнение эксперта. Коэффициент компетентности определяется по априорным и апостериорным данным.

Существует ряд методик определения коэффициента компетентности по априорным данным. Наиболее простой является методика оценки относительных коэффициентов компетентности по результатам высказывания специалистов о составе экспертной группы. Ряду специалистов предлагается высказать суждение о включении лиц в экспертную группу для решения определенной проблемы. Проведя несколько туров такого опроса, можно составить достаточно полный список кандидатов в эксперты. По результатам проведенного опроса составляется матрица, в ячейках которой проставляются переменные  $x_{i,j}$ , равные  $x_{i,j} = 1$ , если  $j$ -й эксперт назвал  $i$ -го эксперта  $x_{i,j} = 0$ , если  $j$ -й эксперт не назвал  $i$ -го эксперта. Причем каждый эксперт может включать или не включать себя в экспертную группу. По данным матрицы вычисляются коэффициенты компетентности как относительные веса

экспертов по формуле 
$$k_i = \frac{\sum_{j=1}^m x_{i,j}}{\sum_{i=1}^m \sum_{j=1}^m x_{i,j}}, \quad (i=1, m)$$

где  $k_i$  – коэффициент компетентности  $i$ -го эксперта,  $m$  — количество экспертов (размерность матрицы  $\|x_{i,j}\|$ ). Коэффициенты компетентности нормированы так, что их сумма равна единице: 
$$\sum_{i=1}^m k_i = 1$$

## *Задача выбора решений на основе аппарата нечетких множеств.*

### *Основные понятия нечетких множеств*

Пусть  $A$  - некоторое множество. Подмножество  $B$  множества  $A$  характеризуется своей характеристической функцией

$$\mu_B(x) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B. \end{cases} \quad (1)$$

Что такое нечеткое множество? Обычно говорят, что нечеткое подмножество  $C$  множества  $A$  характеризуется своей функцией принадлежности  $\mu_C: A \rightarrow [0,1]$ . Значение функции принадлежности в точке  $x$  показывает степень принадлежности этой точки нечеткому множеству. Нечеткое множество описывает неопределенность, соответствующую точке  $x$  – она одновременно и входит, и не входит в нечеткое множество  $C$ . За вхождение в  $A$  -  $\mu_C(x)$  шансов, за второе (не входит в  $A$ ) –  $(1 - \mu_C(x))$  шансов.

Если функция принадлежности  $\mu_C(x)$  имеет вид (1) при некотором  $B$ , то  $C$  есть обычное (четкое) подмножество  $A$ . Таким образом, теория нечетких множеств является не менее общей математической дисциплиной, чем обычная теория множеств, поскольку обычные множества – частный случай нечетких.

Обычное подмножество можно было бы отождествить с его характеристической функцией. Этого математики не делают, поскольку для задания функции (в ныне принятом подходе) необходимо сначала задать множество. Нечеткое же подмножество с формальной точки зрения можно отождествить с его функцией принадлежности. Однако термин «нечеткое подмножество» предпочтительнее при построении математических моделей реальных явлений.

Теория нечеткости является обобщением интервальной математики. Действительно, функция принадлежности

$$\mu_B(x) = \begin{cases} 1, & x \in [a, b], \\ 0, & x \notin [a, b] \end{cases}$$

задает интервальную неопределенность – про рассматриваемую величину известно лишь, что она лежит в заданном интервале  $[a,b]$ . Тем самым описание неопределённости с помощью нечетких множеств является более общим, чем с помощью интервалов.

Начало современной теории нечеткости положено работой 1965 г. американского ученого азербайджанского происхождения Л.А.Заде.

Л.А.Заде рассматривал теорию нечетких множеств как аппарат анализа и моделирования гуманистических систем, т.е. систем, в которых участвует человек. Его подход опирается на предпосылку о том, что элементами мышления человека являются не числа, а элементы некоторых нечетких множеств или классов объектов, для которых переход от "принадлежности" к "непринадлежности" не скачкообразен, а непрерывен. В настоящее время методы теории нечеткости используются почти во всех прикладных областях, в том числе при управлении предприятием, качеством продукции и технологическими процессами.

Л.А. Заде использовал термин "fuzzy set" (нечеткое множество). На русский язык термин "fuzzy" переводили как нечеткий, размытый, расплывчатый.

### ***Аппарат теории нечетких множеств.***

В качестве примера дадим определения теоретико-множественных операций над нечеткими множествами. Пусть  $C$  и  $D$  - два нечетких подмножества  $A$  с функциями принадлежности  $\mu_C(x)$  и  $\mu_D(x)$  соответственно. Пересечением  $C \cap D$ , произведением  $CD$ , объединением  $C \cup D$ , отрицанием  $\bar{C}$ , суммой  $C+D$  называются нечеткие подмножества  $A$  с функциями принадлежности

$$\mu_{C \cap D}(x) = \min(\mu_C(x), \mu_D(x)), \mu_{CD}(x) = \mu_C(x)\mu_D(x), \mu_{\bar{C}}(x) = 1 - \mu_C(x),$$

$$\mu_{C \cup D}(x) = \max(\mu_C(x), \mu_D(x)), \mu_{C+D}(x) = \mu_C(x) + \mu_D(x) - \mu_C(x)\mu_D(x), \quad x \in A,$$

соответственно.

Как уже отмечалось, теория нечетких множеств в определенном смысле сводится к теории вероятностей, а именно, к теории случайных множеств. Однако при решении прикладных задач вероятностно-статистические методы и методы теории нечеткости обычно рассматриваются как различные.



Для оценки значений показателей, не имеющих количественной оценки, можно использовать методы нечетких множеств. Например, нечеткие множества применялись при моделировании задач ценообразования на электронные обучающие курсы, используемые при дистанционном обучении, а также для определения прогноза рейтинга специальности в вузе с помощью экспертов.

### ***Статистика нечетких множеств***

Нечеткие множества – частный вид объектов нечисловой природы. Статистические методы анализа объектов нечисловой природы описаны в [3]. В частности, среднее значение нечеткого множества можно определить по формуле:

$$M(A) = \frac{\sum_{i=1}^n x_i \mu_A(x_i)}{\sum_{i=1}^n \mu_A(x_i)},$$

где  $\mu_A(x_i)$  - функция принадлежности нечеткого множества А.

Как известно, методы статистики *нечисловых* данных базируются на использовании расстояний (или показателей различия) в соответствующих пространствах нечисловой природы. Расстояние между нечеткими подмножествами  $A$  и  $B$  множества  $X = \{x_1, x_2, \dots, x_k\}$  можно определить как

$$d(A, B) = \sum_{j=1}^k |\mu_A(x_j) - \mu_B(x_j)|,$$

где  $\mu_A(x_j)$  - функция принадлежности нечеткого множества  $A$ , а  $\mu_B(x_j)$  - функция принадлежности нечеткого множества  $B$ . Может использоваться и другое расстояние:

$$d_1(A, B) = \frac{\sum_{j=1}^k |\mu_A(x_j) - \mu_B(x_j)|}{\sum_{j=1}^k (\mu_A(x_j) + \mu_B(x_j))}.$$

(Примем это расстояние равным 0, если функции принадлежности тождественно равны 0.)

В соответствии с аксиоматическим подходом к выбору расстояний (метрик) в пространствах нечисловой природы разработан обширный набор систем аксиом, из которых выводится тот или иной вид расстояний (метрик) в конкретных пространствах [1,2,4]. При использовании вероятностных моделей расстояние между случайными нечеткими множествами само является случайной величиной, имеющей в ряде постановок асимптотически нормальное распределение [2].

### ***Нечеткие множества как проекции случайных множеств***

С самого начала появления современной теории нечеткости в 1960-е годы началось обсуждение ее взаимоотношений с теорией вероятностей. Дело в том, что функция принадлежности нечеткого множества напоминает распределение вероятностей. Отличие только в том, что сумма вероятностей по всем возможным значениям случайной величины (или интеграл, если множество возможных значений несчетно) всегда равна 1, а сумма  $S$  значений функции принадлежности (в непрерывном случае - интеграл от функции принадлежности) может быть любым

неотрицательным числом. Возникает искушение пронормировать функцию принадлежности, т.е. разделить все ее значения на  $S$  (при  $S \neq 0$ ), чтобы свести ее к распределению вероятностей (или к плотности вероятности). Однако специалисты по нечеткости справедливо возражают против такого «примитивного» сведения, поскольку оно проводится отдельно для каждой размытости (нечеткого множества), и определения обычных операций над нечеткими множествами с ним согласовать нельзя. Последнее утверждение означает следующее. Пусть указанным образом преобразованы функции принадлежности нечетких множеств  $A$  и  $B$ . Как при этом преобразуются функции принадлежности  $A \cap B, A \cup B, A + B, AB$ ? Установить это *невозможно в принципе*. Последнее утверждение становится совершенно ясным после рассмотрения нескольких примеров пар нечетких множеств с одними и теми же суммами значений функций принадлежности, но различными результатами теоретико-множественных операций над ними, причем и суммы значений

соответствующих функций принадлежности для этих результатов теоретико-множественных операций, например, для пересечений множеств, также различны.

Авторы, сравнивавшие теорию нечеткости и теорию вероятностей, обычно подчеркивали различие между этими областями теоретических и прикладных исследований. Обычно сравнивают аксиоматику и сравнивают области приложений.

При сравнении различных аксиоматик теории нечеткости и теории вероятностей нетрудно увидеть, что списки аксиом различаются. Из этого, однако, отнюдь не следует, что между указанными теориями нельзя установить связь, типа известного сведения евклидовой геометрии на плоскости к арифметике (точнее к теории числовой системы). Напомним, что эти две аксиоматики - евклидовой геометрии и арифметики - на первый взгляд весьма сильно различаются.

Как оказалось, теория нечетких множеств тесно связана с теорией случайных множеств и первые могут рассматриваться как «проекции» случайных множеств. Можно ожидать, что теория нечеткости как некое общее целое обобщает классическую математику. Однако теория нечеткости в определенном смысле сводится к теории случайных множеств и тем самым является частью классической математики. Другими словами, по степени общности обычная математика и нечеткая математика эквивалентны. Однако для практического применения в теории принятия решений описание и анализ неопределенностей с помощью теории нечетких множеств весьма плодотворны.

## ***Основы сетевого планирования***

1. Основные понятия сетевого планирования.
2. Порядок построения сетевого графика.
3. Оценка времени завершения событий, работ и путей сетевого графика.
4. Оптимизация параметров сетевого графика и временного плана операции.

Сетевая модель отображает взаимосвязи между операциями и порядок их выполнения (отношение упорядочения или следования). Как правило, для представления операции (*работы*) используется стрелка (ориентированная дуга), направление которой соответствует процессу реализации программы во времени. Отношение упорядочения между операциями задается с помощью *событий*.

Сетевое планирование это способ разработки *рационального* плана проведения (осуществления) операции в кратчайший срок и с минимальными затратами. Сетевое планирование является инструментом должностных лиц, планирующих мероприятия по осуществлению сложных многофазных оперативных и технологических операций, а также определению в них наиболее *узких мест* с целью их устранения за счет перераспределения ресурсов с менее напряженных этапов операции.

Примерами задач, решаемых с помощью сетевого планирования являются: задача планирования развертывания и эксплуатационного обслуживания наземных, радиорелейных, тропосферных и спутниковых линий связи; задача разработки учебного плана подготовки специалистов - связистов; задача разработки и проведения определенного вида занятий со студентами в рамках принятой учебной программы.

Постановка перечисленных задач сетевого планирования начинается с выявления узловых моментов (*событий*) операции, определяющих начало и конец основных ее этапов. Затем определяются мероприятия (*работы*), приводящие к свершению каждого события и сопровождаемые затратами времени и ресурсов. Наконец, события и соответствующие им работы выстраиваются в некоторую логическую последовательность, графически представляемую в виде *сетевого графика*, отражающего все *пути* достижения конечной цели операции.

Последующий анализ временных и вероятностных характеристик сетевого графика имеет целью оценить степень достижения всех оперативно-технических требований к каждому этапу и операции в целом, а также выявить *узкие места* в ней с возможной их ликвидацией за счет перераспределения части ресурсов с благополучных этапов операции.

**Событие** – начальная (либо конечная) точка работы, не связанная с затратами времени и ресурсов.

**Работа** – конечный во времени процесс выполнения мероприятия, связанный с затратами времени и расходом ресурса.

Основные свойства сетевого графика:

- событие происходит тогда, когда все входящие в него работы являются завершёнными;
- исходящая из события работа может начаться только после его свершения;
- последующая работа может начаться только после окончания предшествующей работы.

Построение сетевого графика начинается с определения перечня событий и их нанесением на график в соответствии с рангом. Ранг события определяется его порядковым номером в общей последовательности событий, так например,



исходное событие имеет нулевой ранг, следующее за ним - первый и т.д.. Затем на график наносятся работы в виде линий, соединяющих исходящие и последующие события. При этом работы кодируются буквой латинского алфавита и двухзначным кодом, первый разряд, которого означает ранг исходящего события, а второй - последующего, например  $A_{ij}$ ,  $i \neq j$ . В общем случае сетевой график (рисунок №1) представляет собой граф, узлы которого  $a_i$  - события, а ориентированные дуги  $A_{ij}$  - работы. При этом в графе любая дуга отражает только одну работу, а в случае возникновения ситуации, когда две работы имеют одни и те же исходящие и входящие узлы (события) вводят дополнительную фиктивную работу, не связанную с затратами времени и других ресурсов.

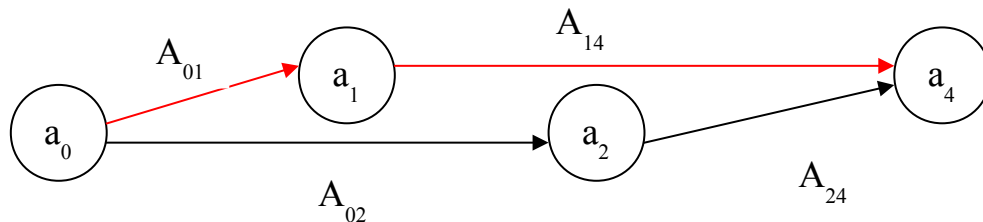


Рис. 1.

Значимыми параметрами такого сетевого графа являются *раннее* и *позднее время* свершения *события*, которые определяют ранее — как время, ранее которого событие не состоится и соответственно позднее, как время, позже которого событие произойти не может. Резервное время определяется, как разница между поздним и ранним временем.

Одна из возможных последовательностей событий и работ от исходного события до завершающего называется *путем* сетевого графика. Путь, имеющий максимальное время выполнения каждой из входящих в него работ и суммарное время выполнения всей операции называется *критическим путем*.

Рассмотрим пример составления сетевого графика для задачи развертывания линии радиорелейной связи между двумя пунктами управления. Исходные данные зададим в виде таблиц с перечнем событий и работ в планируемой операции по развертыванию линии связи.

a <sub>0</sub>	Сбор личного состава
a <sub>1</sub>	Вручены схемы-приказы на обеспечение связи между ПУ
a <sub>2</sub>	Прибыла в район развертывания РРС№1
a <sub>3</sub>	Прибыла в район развертывания РРС№2
a <sub>4</sub>	РРС№1 развернута
a <sub>5</sub>	РРС№2 развернута
a <sub>6</sub>	Установлена служебная связь между РРС№1 и 2
a <sub>7</sub>	Завершена регулировка каналов и они сданы в эксплуатацию

Обозн. работ	Наименование работ	Время выполнения работы(ч)	Предшествующие работы	Последующие работы
A <sub>01</sub>	Прибытие в район сбора	0,5	-	A <sub>12</sub> , A <sub>13</sub>
A <sub>12</sub>	Совершение марша в район развертывания РРС№1	1	A <sub>01</sub>	A <sub>24</sub>
A <sub>13</sub>	Совершение марша в район развертывания РРС№2	2	A <sub>01</sub>	A <sub>35</sub>
A <sub>24</sub>	Развертывание РРС№1	0,5	A <sub>12</sub>	A <sub>46</sub>
A <sub>35</sub>	Развертывание РРС№2	0,5	A <sub>13</sub>	A <sub>56</sub>
A <sub>46</sub>	Установление служебной связи с РРС№2	20мин	A <sub>24</sub>	A <sub>67</sub>

$A_{56}$	Установление служебной связи с РРС№1	20мин	$A_{35}$	$A_{67}$
$A_{67}$	Регулировка и сдача каналов в эксплуатацию	10мин	$A_{46}, A_{56}$	-

Соответствующий исходным данным сетевой график представлен на рисунке №2.

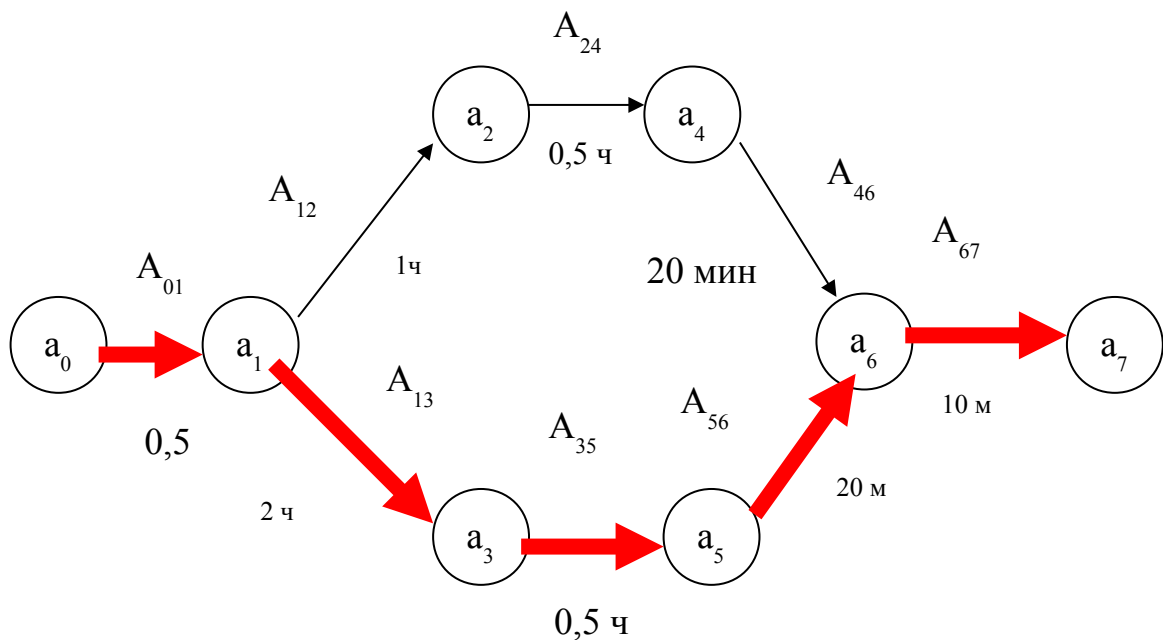


Рис. 2.

Анализ параметров сетевого графика сводится к расчету:  $T_p(j) = \max \{T_p(i) + t_{ij}\}$  - раннее возможное (прямое) время наступления  $j$ -го события ;  $T_n(i) = \min [T_n(j) - t_{ij}]$  - позднее допустимое (обратное) время наступления  $i$ -го события;  $R_i = T_n(i) - T_p(i)$  - резерв времени  $i$ -го события;  $r_n(ij) = T_n(j) - T_p(i) - t_{ij}$  - полный резерв  $ij$ -й работы ;  $r_c(ij) = T_p(j) - T_n(i) - t_{ij}$  - свободный резерв времени для  $ij$ -й работы;  $t_{ij}$  - продолжительность выполнения  $ij$ -й работы. Путь, для которого полные резервы равны нулю ( т.е. суммарное время максимально) выделяется как критический.

Результаты расчета временных параметров сетевого графика можно представить в виде следующей таблицы:

Событие $i(j)$	$T_p(j)$	$T_n(i)$	$R_i$	$r_n(ij)$	$r_c(ij)$
0	0	0	0	$r_n(01)=0$	$r_c(01)=0$
1	0,5	0,5	0	$r_n(12)=1$	$r_c(12)=0$
2	1,5	2,5	1	$r_n(24)=1$	$r_c(24)=0$
3	2,5	2,5	0	$r_n(36)=1$	$r_c(36)=0$
4	2	3	1	$r_n(43)=0$	$r_c(43)=0$
5	3	3	0	$r_n(56)=0$	$r_c(56)=0$
6	3ч 20м	3ч 20м	0	$r_n(67)=0$	$r_c(67)=0$
7	3ч 30м	3ч 30м	0	$r_n(77)=0$	$r_c(77)=0$

Как видно из таблицы критическим путем в сетевом графике является путь 0-1-3-5-6-7, для которого  $r_n(ij)=r_c(ij)=0$  и суммарное время выполнения всех работ  $T \Sigma = 3$  часа 30 минут.

Оптимизация процесса развертывания радиорелейной линии связи на основе параметров сетевого графика начинается с выявления *узких* мест, определяющих временные задержки на критическом пути. Такими работами являются: совершение марша и развертывание РРС№2. Для коррекции плана развертывания линии в сторону уменьшения общих временных затрат целесообразно построить (временной) календарный план, в котором по оси абсцисс отложено текущее время, а по оси ординат критический путь и отрезки времен выполнения не критических работ по ранним и поздним срокам их выполнения. На рисунке №3 представлен временной график выполнения работ по развертыванию радиорелейной линии связи. Красным цветом показан критический путь (0-1-3-5-6-7), который проходит РРС№2, а пунктиром отрезки не критического пути (0-1-2-4-6-7), проходимого РРС№1 до завершения операции развертывания. Как видно из графика узким местом в развертывании линии связи является марш, совершаемый РРС№2 до места

развертывания. Возможный вариант перераспределения ресурса личного состава связан с включением в экипаж РРС№2 лучшего из двух имеющихся водителя-электромеханика, что позволит пройти марш без остановки на отдых, а также дополнительного радиорелейного механика из состава РРС№1. При этом время марша РРС№2 может быть сокращено на 0,5 часа, а время развертывания станции №2 на 10 минут. В результате общее время выполнения задачи станцией №2 будет сокращено с 3 часов 30 минут до 2 часов 50 минут. Вместе с тем время марша РРС№1 с учетом неопытности молодого водителя может быть увеличено на 5 минут, а допустимое время развертывания РРС№1 экипажем неполного состава на 10 минут, что также обеспечит полное время выполнения операции за 2 часа 50 минут. Следовательно выигрыш по времени при скорректированном плане развертывания составит 3 часа 30 минут – 2 часа 50 минут = 40 минут или более 20% от прежнего времени выполнения операции.

В предположении нормального закона распределения временных параметров сетевого графа при многократной реализации рассматриваемой операции, вероятность свершения  $j$ -го события в расчетный срок может быть определена по формуле

$$p_j(T \geq T_j) = \Phi \left\{ \frac{(T - T(j))}{\sqrt{\sum \sigma}} \right\},$$

где  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{tt}{2}) dt$  - функция нормального распределения;

T-заданный срок свершения события;

T(j)- время раннего свершения j-го события;

$\sigma_{ij}$ - среднеквадратическая ошибка в определении продолжительности работ, приводящих к раннему свершению j-го события.

## **Заключение**

Таким образом, использование методов сетевого планирования позволяет:

- а) наглядно представлять последовательность и сроки выполнения работ каждым исполнителем (в виде пути на графике) при совместном участии всех в достижении цели операции;
- б) определять *узкие места* в общем алгоритме реализации данной операции и границы временных ресурсов для ненапряженных работ на ее различных этапах;
- с) корректировать временной план выполнения работ каждому исполнителю на основе перераспределения временных и людских ресурсов с целью выполнения операции в кратчайшие сроки и с минимальными затратами ресурсов.



**«Теория принятия решений»**  
ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

*Лекция 9*

**Теория графов в задачах принятия решений.  
Основы теории графов.**

Учебные вопросы:

Введение. Историческая справка.

1. Основные понятия. Элементы теории графов.
2. Матричное представление графа. Матрица смежности. Матрица инцидентий.  
Матрицы достижимостей и контрдостижимостей.
3. Линейные графы сигналов и передача графа. Эквивалентные преобразования графов. Передача графа.

Заключение

## Литература:

1. Харари, Ф. Теория графов / Ф. Харари ; Под ред. Г.П. Гаврилова. — М. : Мир, 1973.
2. Басакер, Р. Конечные графы и сети / Р. Басакер, Т. Саати. — М. : Наука, 1974.
3. Кристофидес, Н. Теория графов. Алгоритмический подход / Н. Кристофидес. — М. : Мир, 1978.
4. Берж, К. Теория графов и ее применения / К. Берж. — М. : Изд-во Иностранной литературы, 1962.
5. Мэзон, С. Электронные цепи, сигналы и системы / С. Мэзон, Г. Циммерман ; Под ред. проф. П.А. Ионкина. — М. : Издательство иностранной литературы, 1963.
6. Деев, В.В. Методы модуляции и кодирования в современных системах связи / В.В. Деев. — Спб. : Наука, 2007. — ISBN: 978-5-02-025182-3.

Введение.

### ***Историческая справка.***

Основоположником теории графов является Леонард Эйлер (1707–1782), который решил в 1736 г. известную в то время задачу о семи кенигсбергских мостах. Впоследствии эта задача стала одной из классических задач теории графов. Впоследствии методы теории графов применялись и другими известными учеными в самых различных областях науки: в 1847г. Кирхгоф использовал теорию деревьев для расчета силы тока в электрической цепи; в 1857г. Кэли использовал деревья в решении задач органической химии. И таких примеров можно привести ещё много .

*Теория графов* является разделом дискретной математики, изучающим свойства графов. Эта область науки крайне обширна и столь же интересна, сколь и сложна. Этот математический инструмент находит применение во многих отраслях науки и промышленности. В качестве примеров использования теории графов можно привести поиск кратчайшего маршрута на карте, построение блок-схем алгоритмов, схем размещения элементов на печатных платах, принципиальные и структурные схемы устройств и многое другое.

## Основные понятия. Элементы теории графов.

В общем смысле *граф*  $G$  задается множеством точек или *вершин* (узлов)  $x_1, x_2, \dots, x_n$  (обозначается символом  $X$ ) и множеством линий или *ребер*  $a_1, a_2, \dots, a_m$  (обозначается  $A$ ), которые соединяют между собой все или часть этих вершин. То есть, граф полностью задается и обозначается парой  $(X, A)$ . Существует и другое обозначение. Граф, содержащий  $n$  вершин и  $m$  ребер, называется  $(n, m)$ -графом, а  $(1, 0)$ -граф называется *тривиальным*. Пример графа показан на рис.1.

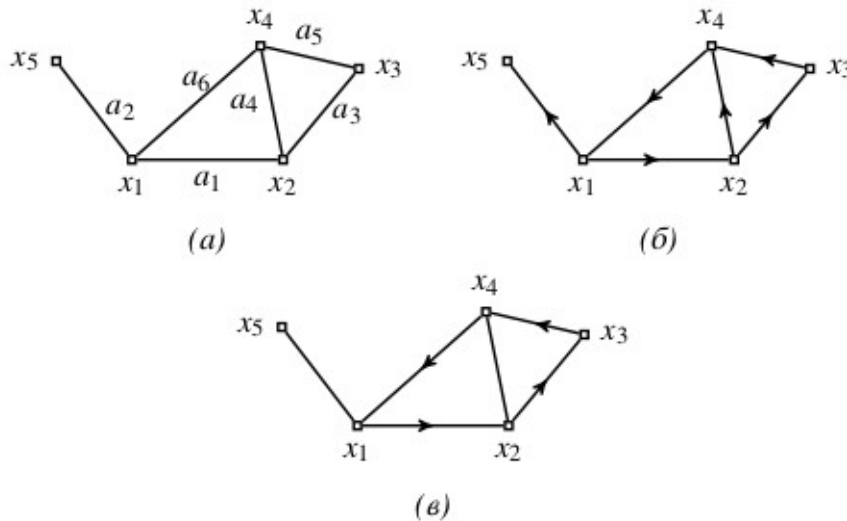


Рис.1. Примеры графа: (а) неориентированный граф; (б) ориентированный граф; (в) смешанный граф.

Две вершины  $x_i$  и  $x_j$ , соединенные ребром  $a_k$ , называются *смежными*. Их иногда обозначают как  $x_i \text{adj} x_j$ . При этом говорят, что вершина  $x_i$  и ребро  $a_k$  *инцидентны*, как и  $x_j$  и  $a_k$ . Два различных ребра, инцидентных одной и той же вершине, также называются *смежными*. Так, на рис.1 вершины  $x_1$  и  $x_2$  смежные, также смежными являются ребра  $a_1$  и  $a_3$ , которые инцидентны вершине  $x_2$ .

Если в графе  $G=(X,A)$  ребра из множества  $A$  ориентированы (обычно показывается стрелкой), то такой граф называется *ориентированным графом* или *орграфом*, а сами ребра называются *дугами*. Граф, ребра которого не имеют ориентации, называется *неориентированным*, а содержащий и ориентированные, и неориентированные ребра — *смешанным*. Неориентированный граф, соответствующий орграфу  $G=(X,A)$ , обозначается как  $G=(X,A)$  и называется *неориентированным дубликатом* или *неориентированным двойником* графа  $G$ . Примеры всех трех видов графа показаны на рис.1. При этом, граф на рис.1(а) является неориентированным дубликатом графа на рис.1(б).

Дуги орграфа удобно обозначать парами вершин вида  $(x_i, x_j)$ , указывая от какой вершины к какой направлена дуга. Например, на рис.1(б) можно выделить дугу  $(x_1, x_2)$ , соответствующую ребру  $a_1$  неориентированного графа на рис.1(а).

Отдельно выделяют два вида графов. *Мультиграф* (см. рис.2(а)), в котором нет петель, но две вершины могут быть соединены более чем одним ребром — такие ребра называются *кратными*. *Петлей* называется ребро (дуга) начальная и конечная вершины которой совпадают. Граф, в котором допускаются и кратные ребра и петли, называется *псевдографом* (рис.2(б)).

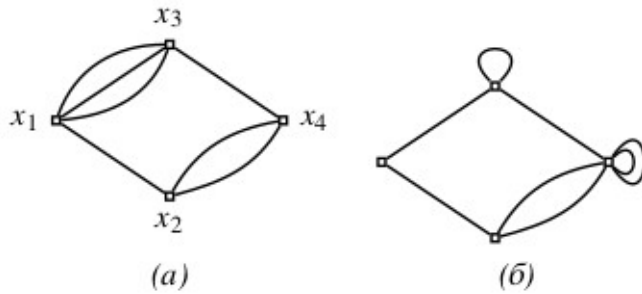


Рис.2. Мультиграф (а) и псевдограф (б).

Орграф, не имеющий симметричных пар дуг, называется *направленным графом*. При этом *симметричными* называют дуги, соединяющие одну и ту же пару вершин, но ориентированные в противоположных направлениях, например  $(x_i, x_j)$  и  $(x_j, x_i)$ . На рис.3(а) показан пример направленного графа. Дуги  $(x_2, x_4)$  и  $(x_4, x_2)$  ненаправленного графа на рис.3(б) являются симметричными.

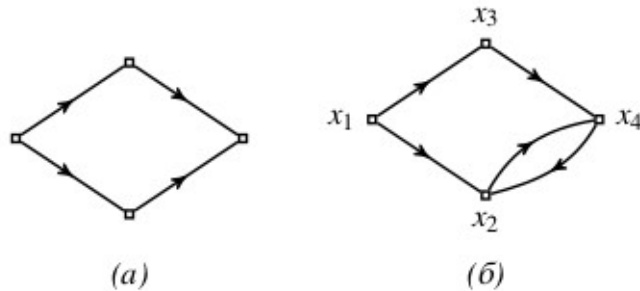


Рис.3. Направленный (а) и ненаправленный (б) графы.

Если вершины графа имеют какие-либо метки, отличающие их друг от друга, то такой граф называется *помеченным* или *перенумерованным*. Например, графы на рис.2(а) и рис.3(б) являются помеченными, а графы на рис.2(б) и 3(а) — нет.

Граф  $G_1$ , все вершины и ребра которого принадлежат графу  $G$ , называется *подграфом* графа  $G$ , а  $G$ , в свою очередь, является *надграфом* для  $G_1$ . Подграф графа  $G$ , содержащий все его вершины, называется *остовным подграфом* или *частичным графом*. Графы на рис.4(б) и 4(в) являются подграфами графа 4(а), который является для них надграфом. При этом, граф 4(в) является остовным подграфом графа 4(а).

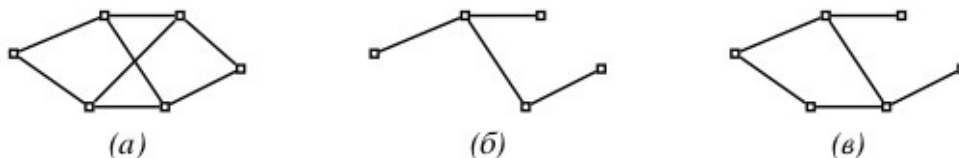


Рис.4. Граф (а) и его подграфы (б) и (в).

*Маршрутом* в графе  $G$  называют чередующуюся последовательность вершин и ребер  $x_1, a_1, x_2, a_2, \dots, x_{n-1}, a_{n-1}, x_n$ , которая начинается и заканчивается вершиной, а каждое ребро инцидентно двум вершинам — непосредственно предшествующей и непосредственно следующей за ним. Фактически, достаточно говорить просто про последовательность вершин или про последовательность ребер — одно подразумевает другое. Такой маршрут, соединяющий вершины  $x_1$  и  $x_n$ , иногда называют  $(x_1 - x_n)$  - *маршрутом*. Если  $x_1 = x_n$ , то маршрут называется *замкнутым*, в ином случае — *открытым*. Маршрут, все ребра которого различны, называется *цепью*. Если при этом различны все вершины, то маршрут называется *простой цепью*. Замкнутая цепь называется *циклом*, а если все  $n$  вершин замкнутого маршрута различны и  $n \geq 3$ , то он называется *простым циклом*.

В случае, если речь идет об ориентированных графах, используют понятие *ориентированный маршрут* или *путь*, под которым понимают последовательность дуг (и, соответственно, вершин), в которой конечная вершина всякой дуги, отличной от последней, является начальной вершиной следующей. Аналогично понятиям цепи и простой цепи выделяют *ориентированную цепь (орцепь)* и *простую орцепь*.

*Длина (мощность) маршрута* определяется количеством ребер в нем. При этом, каждое ребро считается столько раз, сколько оно встречается в этом маршруте. *Расстоянием*  $d(x_i, x_j)$  между вершинами  $x_i$  и  $x_j$  графа называется длина кратчайшей простой цепи, соединяющей их. Для несоединенных  $x_i$  и  $x_j$  полагают  $d(x_i, x_j) = \infty$ .

Иногда дугам  $(x_i, x_j)$  графа  $G$  ставится в соответствие некоторое число  $c_{ij}$ , называемое *весом (длиной, стоимостью)* дуги. Такой граф  $G$  называется *графом со взвешенными дугами*. Если же некоторые веса  $v_i$  приписаны вершинам  $x_i$ , то такой граф называется *графом со взвешенными вершинами* или просто *взвешенным*. Иногда *взвешенным* называют граф, дуги и вершины которого имеют соответствующие им веса.

В том случае, если рассматривается некоторый путь  $\mu$ , представленный последовательностью дуг, то за его *вес* принимается число  $l(\mu)$ , равное сумме весов всех дуг, входящих в  $\mu$ :

$$l(\mu) = \sum_{(x_i, x_j) \in \mu} c_{ij}.$$



## Матричное представление графа

Для алгебраического задания граф удобно представлять в виде матриц.

Для примера возьмем граф  $G$ , представленный на рис.5.

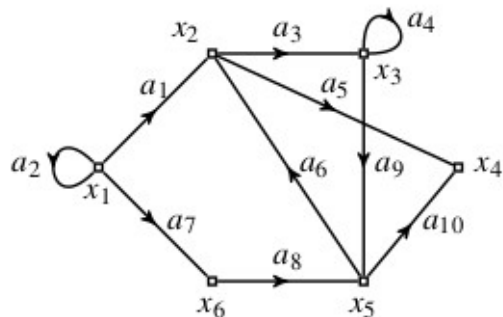


Рис.5. Пример графа представления

## Матрица смежности

Матрица смежности графа  $G$  полностью определяет структуру графа. Она обозначается как  $\mathbf{A}=[a_{ij}]$ , где

$a_{ij}=1$ , если в  $G$  существует дуга  $(x_i, x_j)$ ,

$a_{ij}=0$ , если в  $G$  отсутствует дуга  $(x_i, x_j)$ .

Таким образом, для графа  $G$  на рис.5 матрица смежности будет равна

$$\mathbf{A} = \begin{array}{c|cccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \hline x_1 & 1 & 1 & 0 & 0 & 0 & 1 \\ x_2 & 0 & 0 & 1 & 1 & 0 & 0 \\ x_3 & 0 & 0 & 1 & 0 & 1 & 0 \\ x_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_5 & 0 & 1 & 0 & 1 & 0 & 0 \\ x_6 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}.$$

## **Матрица инциденций**

Матрица инциденций графа  $G$  с  $n$  вершинами и  $m$  дугами, обозначается как  $\mathbf{B}=[b_{ij}]$ . Она имеет размерность  $n \times m$  и определяется как

$$\begin{aligned} b_{ij} &= 1, & \text{если } x_i &\text{ — начальная вершина дуги } a_j, \\ b_{ij} &= -1, & \text{если } x_i &\text{ — конечная вершина дуги } a_j, \\ b_{ij} &= 0, & \text{если } x_i &\text{ не инцидентна } a_j \text{ или } a_j \text{ — петля.} \end{aligned}$$

Таким образом, для графа  $G$  на рис.5 матрица инциденций будет равна

$$\mathbf{B} = \begin{array}{c|cccccccccc} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} \\ \hline x_1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ x_2 & -1 & 0 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ x_3 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ x_4 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ x_5 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 1 \\ x_6 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \end{array} .$$

Так как каждая дуга инцидентна двум различным вершинам, кроме случая, когда она образует петлю, каждый столбец матрицы инциденций либо содержит и  $\langle\langle 1 \rangle\rangle$  и  $\langle\langle -1 \rangle\rangle$ , либо все его элементы равны  $\langle\langle 0 \rangle\rangle$ .

Для неориентированного графа матрица инциденций определяется аналогично, за исключением того, что  $\langle\langle -1 \rangle\rangle$  заменяется на  $\langle\langle 1 \rangle\rangle$ .

## **Матрицы достижимостей и контрадостижимостей**

Матрица достижимостей  $\mathbf{R}=[r_{ij}]$  определяется как

$$\begin{aligned} r_{ij} &= 1, & \text{если вершина } x_j &\text{ достижима из } x_i, \\ r_{ij} &= 0, & \text{в противном случае.} \end{aligned}$$

Множество вершин  $R(x_i)$  графа  $G$ , достижимых из заданной вершины  $x_i$ , состоит из таких  $x_j$ , для которых элемент  $r_{ij}$  в матрице  $\mathbf{R}$  равен  $\langle\langle 1 \rangle\rangle$ . Все диагональные

элементы  $\mathbf{R}$  равны  $\ll 1 \gg$ , так как каждая вершина достижима из себя самой с помощью пути длины 0. Матрица достижимостей  $\mathbf{R}$  для графа  $G$  на рис.5 равна

$$\mathbf{R} = \begin{array}{c|cccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \hline x_1 & 1 & 1 & 1 & 1 & 1 & 1 \\ x_2 & 0 & 1 & 1 & 1 & 1 & 0 \\ x_3 & 0 & 1 & 1 & 1 & 1 & 0 \\ x_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ x_5 & 0 & 1 & 1 & 1 & 1 & 0 \\ x_6 & 0 & 1 & 1 & 1 & 1 & 1 \end{array}.$$

Матрица контрадостижимостей (или матрица обратных достижимостей)  $\mathbf{Q} = [q_{ij}]$  определяется как

$$\begin{aligned} r_{ij} &= 1, & \text{если вершина } x_i \text{ достижима из } x_j, \\ r_{ij} &= 0, & \text{в противном случае.} \end{aligned}$$

Таким образом, контрадостижимое множество  $Q(x_i)$  графа  $G$  — это множество таких вершин, что из любой вершины этого множества можно достигнуть вершины  $x_i$ .

Из определений матриц  $\mathbf{Q}$  и  $\mathbf{R}$  следует, что матрица контрадостижимостей  $\mathbf{Q}$  равна транспонированной матрице достижимостей  $\mathbf{R}^T$ :

Матрица контрадостижимостей  $\mathbf{Q} = \mathbf{R}^T$ .  $\mathbf{Q}$  для графа  $G$  на рис.5 равна

$$\mathbf{Q} = \begin{array}{c|cccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \hline x_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ x_2 & 1 & 1 & 1 & 0 & 1 & 1 \\ x_3 & 1 & 1 & 1 & 0 & 1 & 1 \\ x_4 & 1 & 1 & 1 & 1 & 1 & 1 \\ x_5 & 1 & 1 & 1 & 0 & 1 & 1 \\ x_6 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}.$$

## Линейные графы сигналов и передача графа

*Граф сигналов* — это графическое представление соотношений между несколькими переменными. В случае линейности этих соотношений, граф выражает систему линейных алгебраических уравнений и называется *линейным графом сигналов*. Такой способ представления позволяет наглядно выразить систему уравнений и решить ее непосредственно путем анализа графа.

Граф сигналов представляет из себя ориентированную цепь, в которой каждая дуга<sup>1</sup>  $(x_j, x_k)$  связана с числом  $t_{jk}$ , называемым *передачей дуги*, а каждому узлу  $x_j$  соответствует так называемый *узловой сигнал*  $X_j$ .

В дальнейшем, говоря в этом разделе о графах, будем иметь в виду именно линейные графы сигналов.

Узловые сигналы определяются уравнениями вида:

$$X_k = \sum_j X_j t_{jk}, \quad k=1, 2, 3, \dots$$

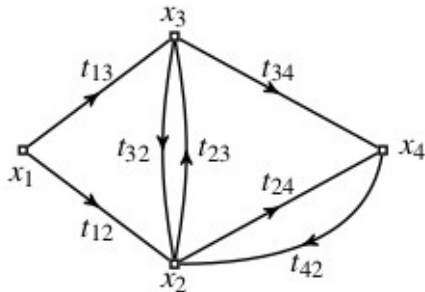


Рис.6. Пример графа сигналов

1 В источниках (например, [5] и [6]) также используют термины <<ребро>> и <<ветвь>>.

Для примера рассмотрим граф сигналов на рис.6. Согласно формуле его узловые сигналы будут равны

$$X_2 = X_1 t_{12} + X_3 t_{32} + X_4 t_{42};$$

$$X_3 = X_1 t_{13} + X_2 t_{23};$$

$$X_4 = X_2 t_{24} + X_3 t_{34}.$$

Таким образом, можно сказать, что линейный граф сигналов представляет из себя систему линейных уравнений, представленную особым образом — вместо символов  $\langle\langle + \rangle\rangle$ ,  $\langle\langle = \rangle\rangle$  используются направленные дуги и узлы. Каждое уравнение в этой системе имеет форму  $\langle\langle \text{причина} — \text{следствие} \rangle\rangle$ . При этом каждый зависимый узловой сигнал выражен один раз в виде явного *следствия*, вызванного другими узловыми сигналами, действующими в качестве *причин*. Совокупность этих свойств приводит к тому, что уравнения вида могут быть решены непосредственно путем вычисления графа .

## Эквивалентные преобразования графов

При решении графов сигналов используются эквивалентные преобразования, позволяющие упростить структуру графа и уменьшить сложность расчетов. На рис.7 приведены элементарные эквивалентные схемы графов.

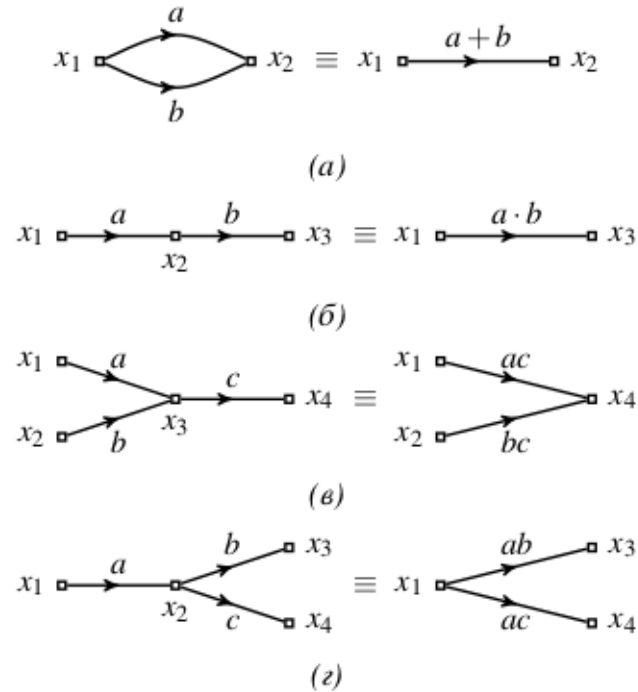


Рис.7. Элементарные эквивалентные схемы графов:

- (а) сложение; (б) умножение; (в) распределение (разложение) на множители справа;  
(г) распределение (разложение) на множители слева

Преобразования на рис.7 являются обратимыми.

На рис.8 приведено исключение узла в звезде. Такое преобразование в общем случае не является обратным. То есть, если задан граф в виде квадрата (рис.8(б)), то не следует ожидать, что эквивалентный ему граф будет иметь форму звезды (рис.8(а)).



Рис. 8.8. Исключение узла в звезде:  
(а) исходная звезда; (б) итоговый квадрат

Если преобразуемый граф содержит любые замкнутые цепи, то при его преобразовании появляются одна или больше петель, как показано на рис.9.

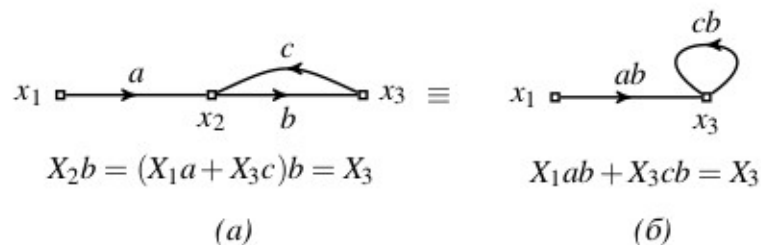


Рис. 8.9. Пример образования петли при разделении дуги:  
(а) исходный граф; (б) итоговый граф с петлей

Уравнение для узлового сигнала, приведенное на рис.9(б) можно преобразовать к виду

$$X_1 a b = X_3 (1 - c b) \Rightarrow X_3 = \frac{X_1 a b}{1 - c b}.$$

Таким образом, граф на рис.9(б) можно свести к одной дуге  $(x_1 x_3)$  с передачей, равной  $\frac{X_1 a b}{1 - c b}$ .

### **Передача графа**

*Передача  $T$  графа* равна сигналу, возникающему в некотором зависимом узле, на единицу сигнала в некотором заданном узле-источнике. Если в графе содержится только один узел - сток  $x_k$ , то есть узел, имеющий только входящие ветви, и один узел - источник  $x_j$  только с исходящими ветвями, то передача графа определяется

однозначно по формуле 
$$T = \frac{X_k}{X_j}.$$

Отметим, что индексы у передачи графа в этом случае не указываются.

Если же в графе есть источник, но нет стока, или стоков несколько, а также в случае произвольной структуры графа при отсутствии источников и стоков, необходимо указывать между какими узлами считается передача графа. В этом случае она обозначается как  $T_{jk}$  — передача графа между узлами  $x_j$  и  $x_k$ .

При определении передачи графа используются его топологические свойства. Важными топологическими параметрами графа являются его пути и контуры.

*Путь* — это непрерывная последовательность дуг (в указанном направлении), вдоль которой каждый узел встречается не более одного раза. Произведение передач дуг вдоль этого пути образует *передачу пути*  $P$ .



*Контур* (или контур обратной связи) — это простой замкнутый путь, вдоль которого каждый узел встречается не более одного раза за один обход контура. *Передача контура*  $L$  равна произведению передач ветвей в этом контуре.

Для примера рассмотрим граф на рис.6. Он содержит четыре различных пути от узла  $x_1$  до узла  $x_4$ :  $P_1=t_{13}t_{34}$ ,  $P_2=t_{13}t_{32}t_{24}$ ,  $P_3=t_{12}t_{23}t_{34}$  и  $P_4=t_{12}t_{24}$ . Также в нем присутствует 3 контура:  $L_1=t_{23}t_{32}$ ,  $L_2=t_{24}t_{42}$  и  $L_3=t_{23}t_{34}t_{42}$ .

Зная все пути между узлами  $x_j$  и  $x_k$  и все контуры в графе можно полностью выразить передачу графа  $T_{jk}$ .

Для расчета передачи любого графа  $T_{jk}$  через  $p$  путей и  $m$  контуров используется общее правило, которое часто называют формулой Мэсона–Циммермана.

$$T_{jk} = \frac{[(P_1 + P_2 + \dots + P_p)(1 - L_1)(1 - L_2) \dots (1 - L_m)]^*}{[(1 - L_1)(1 - L_2) \dots (1 - L_m)]^*}.$$

Знак угла означает, что при умножении коэффициентов внутри скобок приравнивается к нулю любой член, который содержит произведение передач двух контуров или произведение передач пути и контура, которые касаются друг друга в графе.

Формулу можно переписать в виде

$$T_{jk} = \frac{\left[ \sum_i P_i \prod_v (1 - L_v) \right]^*}{\left[ \prod_{w=1}^m (1 - L_w) \right]^*} = \frac{\left[ \sum_i P_i \prod_v (1 - L_v) \right]^*}{1 - \sum_{w=1}^m L_w + \left[ \sum_{s,t} L_s L_t \right]^* - \left[ \sum_{s,t,u} L_s L_t L_u \right]^* + \dots},$$

где  $P_i$  — передача  $i$ -го прямого пути (без замкнутых циклов) между узлами  $x_j$  и  $x_k$  ( $i=1..p$ );  $L_v$  — передача  $v$ -го контура.

В числителе формулы должны находиться только те сомножители  $(1-L_v)$ , для которых контур  $L_v$  не соприкасается с путем  $P_i$ . Знак угла за скобкой показывает, что среди членов соответствующих сумм в произведениях должны отсутствовать соприкасающиеся или пересекающиеся контуры.

Для примера рассмотрим нахождение передачи графа для диаграммы состояний сверточного кода, которая используется для решения задачи определения весового спектра сверточного кода.

На рис.10 приведены диаграмма состояний сверточного кода (рис.10(а)) и модифицированная диаграмма состояний (рис.10(б)), в которой узел  $x_1$  разделен на начальный узел  $x_1$  и конечный узел  $x_1'$ . При этом рассматриваются только те пути, которые выходят из узла  $x_1$  и в него же возвращаются, не попадая в это состояние в промежуточные моменты — именно поэтому на модифицированной диаграмме (рис.10(б)) отсутствует петля  $t_{11}$ .

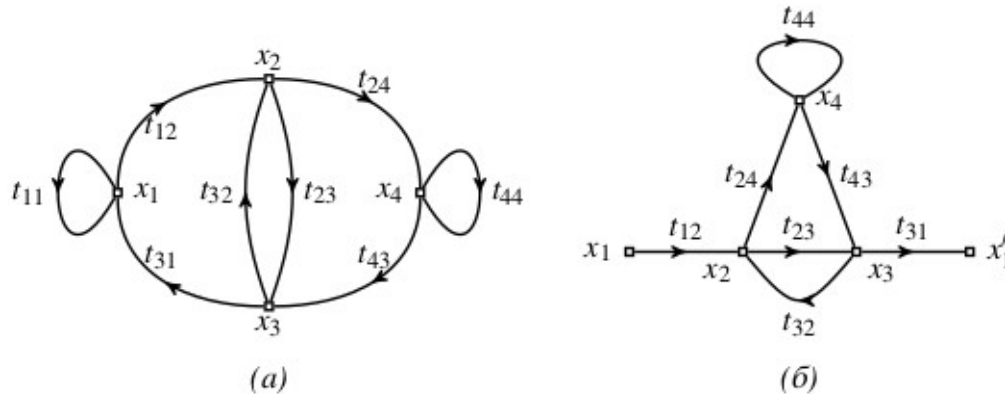


Рис.10. Диаграмма состояний сверточного кода:  
(а) исходная; (б) модифицированная

Модифицированная диаграмма на рис.10(б) содержит два пути между узлами  $x_1$  и  $x_1'$

$$P_1 = (x_1 x_2 x_3 x_1') = t_{12} t_{23} t_{31};$$

$$P_2 = (x_1 x_2 x_4 x_3 x_1') = t_{12} t_{24} t_{43} t_{31}.$$

Также она содержит три контура

$$L_1 = (x_4 x_4) = t_{44};$$

$$L_2 = (x_2 x_3 x_2) = t_{23} t_{32};$$

$$L_3 = (x_2 x_4 x_3 x_2) = t_{24} t_{43} t_{32}.$$

Таким образом, согласно формуле , передача графа для диаграммы на рис.10(б) между узлами  $x_1$  и  $x_1'$  равна

$$T_{11}' = \frac{P_1(1-L_1)+P_2}{1-(L_1+L_2+L_3)+L_1L_2}.$$

В практической деятельности постоянно возникают задачи “наилучшего” - оптимального размещения оборудования в сетях или графах – с объективным требованием минимизации наибольшего расстояния от произвольной вершины графа до пункта размещения. Такие задачи называются минимаксными задачами размещения.

Второй тип задач для таких моделей, ориентированный чаще на поиск максимальных значений - это задачи сетевого планирования и управления.

Еще один большой класс задач теории графов – задача о развозке(доставке), технической трассировке электрических цепей, маршрутные задачи и подобные, приведенные к ним задачи нахождения кратчайших путей и оптимизации маршрутов, включая задачи максимальных и минимальных потоков.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 10*

### **Методы векторной динамической оптимизации**

#### СО Д Е Р Ж А Н И Е

Введение

Учебные вопросы:

1. Проблемы векторной оптимизации принятия решений, редукция системы критериев оптимальности, нормирование составляющих векторного критерия, скаляризация (свертка) векторного критерия оптимальности.
2. Отыскание парето-оптимальных решений.
3. Принцип разделения в стохастической задаче принятия решений (оптимального управления). Принцип оптимальности Беллмана. Функциональное уравнение Беллмана.

Заключение

## Литература:

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСи, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.
4. Терентьев В.М., Санин Ю.В. Анализ эффективности функционирования автоматизированных сетей многоканальной радиосвязи. - С-Петербург: ВАС, 1992.
5. Терентьев В. М., Илюхин А. А., Куцакин А. И., Осипов А. Н. Основы построения сетей спутниковой связи с подвижными объектами: Учебное пособие. – Орел: Академия Спецсвязи России, 2004.

## Введение

Рассматривая особенности теории принятия решений с точки зрения стохастического оптимального управления можно сказать, что к ним относятся:

- использование для описания объектов методов теории переменных состояния;
- использование среднеквадратичных целевых функций, записанных в рекуррентной форме;
- фундаментальная роль принципа разделения;
- построение алгоритмов стохастического оценивания состояния объекта на основе фильтров Калмана;
- использование известных детерминированных методов при принятии управленческих решений (отыскании управляющих воздействий).

### ***Проблемы векторной оптимизации процесса принятия решений.***

Формализованная цель задачи - критерий оптимальности  $I_n = \min L(Y, Y_{mp})$  - формулируется на основе целевой функции  $L(Y, Y_{mp})$ , включающей в себя систему показателей и требования к их значениям, а также указания по поиску их экстремумов (*min, max, minmax, maxmin* и др.).

В практических задачах, как нами рассматривалось ранее, используются следующие основные виды целевых функций:

-простая  $L=(Y-Y_{mp})$  ;

-модульная  $L=|Y-Y_{mp}|$  ;

-квадратичная  $L=|Y-Y_{mp}|^2$  .

Наличие множества различных и зачастую противоречивых критериев оптимальности порождает проблему многокритериальной (векторной) оптимизации процесса выбора решения. Основными операциями на пути ее решения являются необходимость сокращения размерности векторного критерия оптимальности (ВКО), нормализация и последующая скаляризация (свертка) его составляющих.

### *Редукция системы критериев оптимальности.*

Уменьшение размерности системы показателей (критериев оптимальности) значительно упрощает решение задачи ВКО. Одним из наиболее распространенных методов редукции является метод, основанный на оценке степени линейной независимости отдельных составляющих векторного критерия.

Вычисление матрицы коэффициентов корреляции ВКО проводится на основе следующего выражения [1,3]:

$$M \rho \{ \rho_{nn'} \} ;$$

$$\rho_{nn'} = \frac{K(I_n, I_{n'})}{\sigma(I_n)\sigma(I_{n'})} = \frac{\sum_{l=1}^M \sum_{i=1}^M [I_{nl} - \bar{I}_n][I_{n'i} - \bar{I}_{n'}] P_{li}}{\sqrt{\sum_{l=1}^M [I_{nl} - \bar{I}_n]^2 P_{nl} \sum_{i=1}^M [I_{n'i} - \bar{I}_{n'}]^2 P_{n'i}}}, \quad (1)$$

где  $K(I_n, I_{n'}) = M[I_n, I_{n'}]$  – матричная взаимная корреляционная функция  $n$ -го и  $n'$ -го критериев диагональные члены которой являются дисперсиями  $n$ -х критериев, а остальные члены характеризуют степень линейной независимости любой пары критериев;

$n, n'$  – номера критериев оптимальности;

$l = 1, \dots, M$  – номера дискретных значений критериев;

$\bar{I}_n = \sum_{l=1}^{\infty} I_n P_{nl}$  – среднее значение критерия;

$P_{nl}, P_{n'l'}$  – вероятности принятия  $n$  ( $n'$ )-м критерием значения  $l$ ;

$P_{ll'}$  – совместная вероятность принятия  $n$ -м критерием  $l$ -го значения и  $n'$ -м критерием  $l'$ -го значения.

Редукция системы критериев осуществляется путем удаления из исходной задачи тех критериев  $I_n$ , которым в матрице коэффициентов корреляции  $\{\rho_{nn'}\}$  соответствуют такие недиагональные элементы, которые превышают величину 0,9 и выше. Следует отметить, что критерии оптимальности в исходной задаче должны быть предварительно ранжированы по степени их важности для принятия решения. Для случая непрерывных критериев вероятности в выражении (1) должны быть заменены на плотности распределения, а суммы на интегралы.



## ***Нормирование составляющих векторного критерия.***

Процесс нормирования включает этапы перехода к единой размерности или безразмерности, сведения разных составляющих к одной точке отсчета и переход к равноценным шкалам (одному масштабу). Достаточно полно все перечисленные этапы могут быть выполнены при использовании следующего линейного преобразования:

$$I_n^{(H)}(\vec{x}; k) = c_n I_n(\vec{x}; k) - d_n, \quad (2).$$

где  $c_n = \frac{1}{(I_n'(\vec{x}; k) - I_n^0(\vec{x}; k))}$  – масштабный коэффициент;

$d_n = \frac{I_n^0(\vec{x}; k)}{(I_n'(\vec{x}; k) - I_n^0(\vec{x}; k))}$  – коэффициент сдвига, корректирующий начало отсчета;

$I_n^{(H)}, I_n', I_n^0$  – нормированное, наибольшее и наименьшее значения критериев.

Использование преобразования (2) позволяет привести все критерии к нулевой точке отсчета, их изменение ограничивается отрезком  $[0, 1]$ , а также делает все критерии безразмерными. Для изменения цели в критерии с *max* на *min* необходимо заменить в выражении (2) знаки при коэффициентах  $c_n$  и  $d_n$  на обратные.

## *Скаляризация (свертка) векторного критерия оптимальности.*

Задача оптимизации по векторному критерию состоит в отыскании решений, удовлетворяющих экстремуму одновременно всех составляющих ВКО. Существует два основных пути решения данной задачи: поиск компромиссных решений, оптимальных по Парето и поиск решений, оптимальных в смысле обобщенного скалярного критерия, полученного путем свертки (скаляризации) всех компонент ВКО. Первый путь связан с трудностями использования строгих математических методов оптимизации для широкого круга задач, а также отсутствием, как правило, единственности искомого решения [2,3]. В связи с этим этап поиска компромиссных решений имеет вспомогательное значение и используется лишь для предварительного уменьшения размерности исходного множества решений до этапа свертки ВКО.

Суть второго метода заключается в сведении векторной задачи оптимизации к скалярной. При этом формируется обобщенный критерий, значение которого для различных вариантов управления является проекцией всех составляющих ВКО на одну числовую ось, что значительно облегчает окончательный выбор оптимального решения, так как существует множество конструктивных скалярных методов оптимизации.

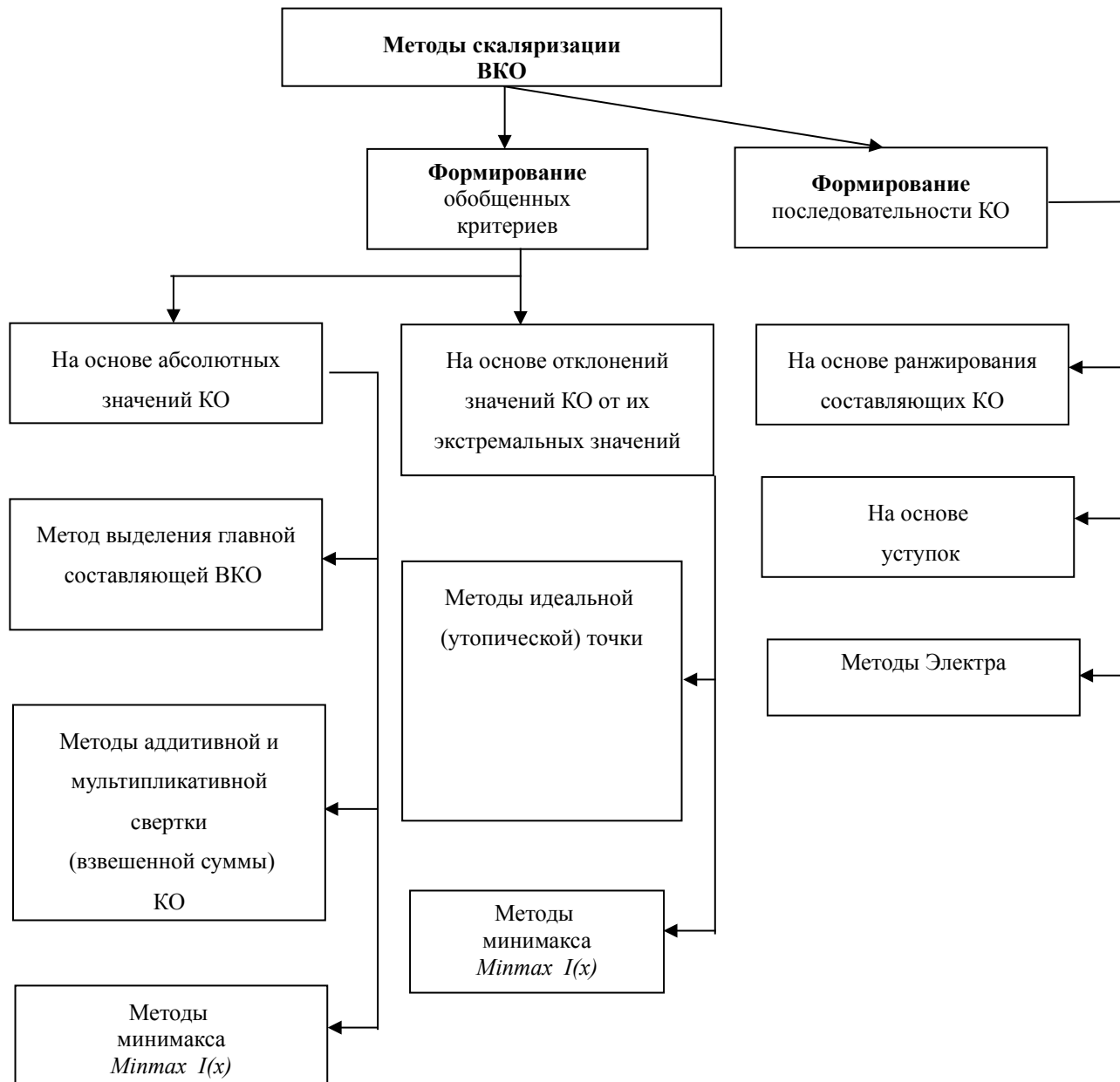


Рис. 2. Методы скаляризации векторных критериев  
ОПТИМАЛЬНОСТИ

К основным методам свертки ВКО (см. рис.2.) относятся: методы, основанные на последовательной оптимизации по частным критериям (метод ведущей составляющей, оптимизация по ранжированной последовательности критериев, метод последовательных уступок, методы «Электра-1»); методы, основанные на получении обобщенных скалярных критериев (метод аддитивной свертки составляющих ВКО с весовыми коэффициентами, метод идеальной (утопической) точки, метод вероятностной свертки).

Особенностями первой группы методов является последовательный (по всем компонентам ВКО) характер решения задачи оптимизации, что приводит к возможности потери компромиссно-оптимального решения уже на первых шагах оптимизации.

**Метод главного критерия.** Здесь сначала выбирают главный критерий и объявляют его единственным. Все остальные критерии становятся управляемыми переменными, на которые накладываются ограничения, чтобы они были не хуже заданной величины. Обычная задача ЛП с ограничениями.

$$\max F(x) = \sum_{j=1}^n c_j x_j = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ &\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \end{aligned}$$

Основным недостатком **метода взвешенной суммы** является субъективный характер выбора весовых коэффициентов определяющих важность различных компонент ВКО и обобщая, как следствие, субъективность получаемых решений.

Свободным от большинства указанных недостатков является **метод идеальной точки**, в котором формирование обобщенного критерия оптимальности осуществляется согласно выражению:

$$I^{обобщ.}(\vec{x}) = \left[ \sum_{n=1}^N (I_n^{(n)}(\vec{x}) - I_n^{(n)идеал.})^q \right]^{\frac{1}{q}}, \quad (3)$$

где  $q = 1, 2, \dots$  - степень целевой функции;

$\vec{x}$  – вектор, оптимизируемых по ВКО параметров.

Следует отметить, что в качестве идеальных значений критериев  $I_n^{идеал.}$  могут выступать либо экстремальные значения  $n$ -х критериев, либо требования к их значениям со стороны пользователей ТКС.

Для случая стохастических систем в выражении (3) выборочные значения критериев должны быть заменены на их средние значения -  $\bar{I} = M\{\bar{I}\}$ .  
Здесь  $M\{ \cdot \}$  - знак математического ожидания.

Кроме того, при наличии косвенных наблюдений за состоянием ТКС эти средние становятся условными по их выборочным значениям:

$$\bar{I}^{обобщ.}(\vec{x}) = M \left\{ \left[ \sum_{n=1}^N (I_n^{(H)}(\vec{x}) - I_n^{(H)идеал.})^q \right]^{1/q} / z(\vec{x}) \right\}, \quad (4)$$

где  $\bar{I}^{обобщ.}(\vec{x})$  - среднее значение обобщенного критерия условное по наблюдениям  $z(\vec{x})$ .

**Метод последовательных уступок.** Проранжируем все критерии и расположим их в порядке убывающей важности:  $K_1, K_2, K_3, \dots, K_L$

Пусть все критерии нужно максимизировать. Сначала отбросим все критерии, кроме  $K_1$ , и найдем допустимое решение, обращающее в максимум  $K_1$ . Как правило, при этом другие критерии не получают своих наилучших значений. Поэтому исходя из практических соображений и точности исходных данных, назначается уступка  $K_1$ , которую мы согласны допустить, для того чтобы обратиться в максимум  $K_2$ . Наложим на критерий  $K_1$  ограничение, чтобы он был не меньше, чем  $K_1_{max} - \Delta K_1$  ( $K_1_{max}$  — максимально возможное значение  $K_1$ ), и при этом ограничении ищем решение, обращающее в максимум  $K_2$ . Далее снова назначается уступка в показателе  $K_2$ , ценой которой находится максимум  $K_3$  и т.д.

Такой способ хорош тем, что здесь сразу видно, ценой какой «уступки» в одном показателе приобретается выигрыш в другом. Заметим, что иногда даже при малом  $\Delta K_{i-1}$  «свобода» выбора позволяет достичь существенной оптимизации  $K_i$ . Это зависит от вида границы критериального пространства. Очевидно, успех такого метода зависит от того, насколько «тупой» максимум. Процедура может быть повторена для других  $K_i$ .

**Методы ЭЛЕКТРА**, предложенные профессором Б. Руа (Франция). В этих методах отношения предпочтения полученных на множестве Парето альтернатив строятся следующим образом.

Для каждого из  $m$  критериев (предполагается, что все критерии числовые) определяется вес – положительное число, характеризующее его важность. Например, при назначении весов критериям выбора автомобиля: цена (критерий 1), важнее комфортности (критерий 2), а та, в свою очередь, важнее, чем скоростные качества (критерий 3) и внешний вид автомобиля (критерий 4). Кроме того, критерии 3 и 4 имеют одинаковую важность, а, рассматриваемые совместно, имеют большую важность, чем критерий 1, т.е.:

$$w_1 > w_2 > w_3 = w_4, \quad w_3 + w_4 > w_1.$$

Назначим весовые коэффициенты:  $w_1 = 5; w_2 = 4; w_3 = w_4 = 3$ .

Выполним нормирование весовых коэффициентов:

$$\lambda_i = \frac{w_i}{\sum_{i=1}^m w_i}, \quad i=1, \dots, m; \quad \sum_{i=1}^m \lambda_i = 1.$$

Для определения, превосходит ли альтернатива  $A_k$  альтернативу  $A_l$ , необходимо:

1) Разбить множество критериев на три подмножества:

- критерии, по которым  $A_k$  превосходит  $A_l$ ;
- критерии, по которым  $A_k$  и  $A_l$  имеют одинаковые оценки;
- критерии, по которым  $A_l$  превосходит  $A_k$ .

2) Определить относительную важность каждого из этих подмножеств

$W_{kl}^+, W_{kl}^-, W_{kl}^-$  как сумму весов входящих в них критериев и считать, что вариант

$A_k$  превосходит  $A_l$  если, выполняется условие: 
$$\frac{W_{kl}^+}{\sum_{i=1}^m w_i} > \frac{W_{kl}^-}{\sum_{i=1}^m w_i}$$

Это условие является необходимым, но не достаточным условием превосходства  $A_k$  над  $A_l$ . Некоторые методы ЭЛЕКТРА формулируют и дополнительные условия оптимизации, которые учитывают не только оценки  $A_k$  и  $A_l$  по критериям, но и значения разностей этих оценок.

## ***Отыскание парето-оптимальных решений.***

**Определение.** Множество безусловно несравнимых альтернатив, оставшихся после отбрасывания всех безусловно худших альтернатив, называется множеством Парето. Парето оптимальное множество еще называют областью компромиссов. Ясно, что множество Парето можно получить в результате анализа критериального пространства.

Таким образом, векторная оптимизация включает два этапа:

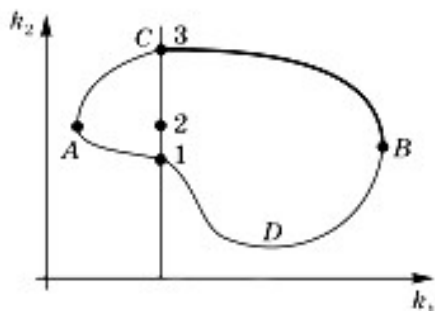
1. Безусловная оптимизация. Здесь анализируется критериальное пространство, отсеиваются безусловно худшие варианты и получают множество Парето.

2. Условная оптимизация. Так как множество Парето, как правило, состоит из более чем одной точки, то для получения единственного решения необходимо применить дополнительные принципы оптимальности (условия согласования критериев).

### **Этап безусловной оптимизации.**

#### ***Определение множества Парето***

#### **Непрерывный случай.**



каждый по своему критерию.

Рассмотрим область возможных значений критериев см.рисунок. Точки, лежащие на одной вертикали или на одной горизонтали, всегда, безусловно,

сравнимы (точки 1, 2, 3). Получается, что все внутренние точки можно отсеять. Дуга  $AB$  состоит из лучших точек по критерию  $K_2$  при постоянном  $K_1$ . Но часть точек дуги  $AB$ , а именно точки дуги  $AC$  можно отбросить, так как они хуже точек на дуге  $CB$ . Поэтому точками множества Парето будут точки, находящиеся на дуге  $CB$ . При анализе характерными являются точки касания вертикалей и горизонталей к области критериев (точки  $A, B, C, D$ ).

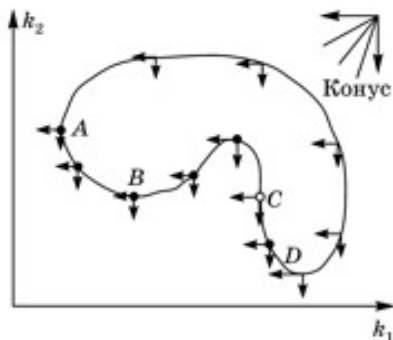
Существует несколько алгоритмов нахождения множества Парето. Они зависят от того, в каком виде задано множество вариантов критериев. В практических задачах сначала приходится получать само множество критериев, затем множество Парето в критериальном пространстве и только затем можно определить Парето-оптимальное множество в области допустимых решений  $X$ .

Существуют два основных метода нахождения множества Парето:

1) **Метод обхода конусом** применяется для непрерывной области критериев.

Назовем конусом пространственный угол, образуемый лучами, идущими из общей вершины и ограниченными в каждой плоскости углом  $90^\circ$ . Направление ограничивающих лучей соответствует направлению оптимизации критериев.

Для получения множества Парето достаточно установить вершину конуса на все точки границы области критериев, если при этом ни один луч не пересекает область, то вершина лежит в точке Парето, а если пересекает, то нет.

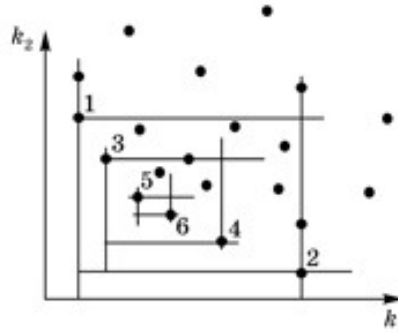


Пример:  $K_1 \Rightarrow \min, K_2 \Rightarrow \min$  см. рисунок.  
Множество Парето:  $[AB] \cup [CD]$ .



## Дискретный случай.

- 2) **Метод прямоугольников** применяют, когда критериальное пространство представляет собой отдельные точки или табличные значения. Сформулируем алгоритм для случая двух критериев, когда  $K_1 \Rightarrow \min$ ,  $K_2 \Rightarrow \min$ , а критериальное пространство — точки на плоскости.



1. Фиксируем самые левые точки. Если их несколько, выбираем среди них самую нижнюю. Эта точка является точкой Парето, фиксируем ее.
2. Выберем самую нижнюю точку. Если их несколько, то выбираем самую левую. Это точка Парето, фиксируем ее.
3. Через полученные точки проводим вертикаль и горизонталь. Отбрасываем сами точки Парето, точки, лежащие на границе полученного прямоугольника, и точки вне прямоугольника.
4. К внутренним точкам полученного прямоугольника применяем алгоритм из пункта 1.

Алгоритм заканчивается тогда, когда внутри прямоугольника не останется ни одной точки.

**Решение:** множество Парето: **1, 2, 3, 4, 5, 6.**

Изложенный алгоритм легко обобщается на случай большего числа критериев, при других направлениях оптимизации, а также на случай табличного задания критериев.

## Этап условной оптимизации.

После нахождения множества Парето, если количество точек в нём  $\geq 2$ , встаёт вопрос о выборе единственной точки, которую необходимо выбрать в качестве оптимальной. Так как точки на множестве Парето отбираются так, что каждая из них лучше по одному критерию, но хуже по другому, то совместное улучшение по двум критериям невозможно. Условная оптимизация предполагает введение условия согласованности в компонентах критерия.

Все это подробно рассмотрено выше в проблемах и задачах векторного многокритериального отбора.

### *Принцип разделения в задаче стохастического оптимального управления.*

Сведение векторной задачи оптимизации к скалярной, выполняемой по обобщенному критерию (4), делает целесообразным рассмотрение общего подхода к решению задач скалярной динамической оптимизации [1,2,3].

Для линейных стохастических процессов, описываемых уравнениями состояния и наблюдения вида:

$$\begin{aligned}\bar{x}(k) &= A(k)\bar{x}(k-1) + B(k)\bar{u}(k) + \bar{v}(k), \\ \bar{z}(k) &= H(k)\bar{x}(k) + \bar{w}(k),\end{aligned}\quad (5)$$

а также при квадратичном критерии оптимальности типа:

$$I^{\text{обобщ}} = \min_{\bar{u} \in U} M \left\{ [\bar{x}^T(K)\rho\bar{x}(K) + \sum_{k=1}^{K-1} \bar{x}^T(k)Q_1\bar{x}(k) + \sum_{k=1}^{K-1} \bar{u}^T(k)R_{1-1}\bar{u}(k)] / \bar{z}(\bar{x}(k)) \right\}, \quad (6)$$

где  $\rho$  и  $Q_1$ - неотрицательно определенные матрицы;

$R_1$ - положительно определенная матрица. Остальные параметры имеют введенный выше смысл.

Как показано в работах [1, 3], для линейной постановки задачи (3, 4.) справедлив принцип разделения, который разбивает решение общей стохастической динамической задачи управления на два отдельно выполняемых, но взаимосвязанных этапа:

- этап получения стохастической оптимальной оценки  $\bar{x}(k) = M[\bar{x}(k)/\bar{z}(k)]$  состояния процесса  $\bar{x}(k)$  на основе наблюдений  $\bar{z}(k)$  ;
- этап формирования детерминированных управляющих воздействий  $\bar{u}^{opt}(k) = -L(k)\bar{x}(k)$  , линейно связанных с оценкой состояния.

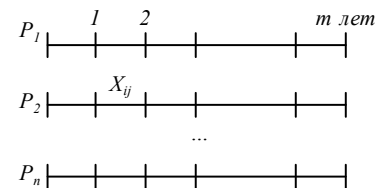
Наиболее конструктивным при решении задачи динамической оптимизации управления оказывается динамическое программирование.

Это особый метод, он специально приспособлен для оптимизации динамических задач, в которых операция состоит из элементов, сильно влияющих друг на друга. ДП связано с именем Ричарда Беллмана, который сформулировал принцип Беллмана. Он позволяет существенно сократить перебор решений в многоэтапных нелинейных задачах.

### Принцип оптимальности Беллмана

Принцип оптимальности Беллмана ставит вопрос о том, что такое оптимальность отдельного элемента системы с точки зрения оптимальности всей системы:

*каково бы ни было состояние системы в результате какого-либо числа шагов, на ближайшем шаге нужно выбирать управление так, чтобы оно в совокупности с оптимальным управлением на всех последующих шагах приводило к оптимальному выигрышу на всех оставшихся шагах, включая выигрыш на данном шаге.*



Принимая решение на отдельном этапе, мы должны выбирать управление на этом этапе с прицелом на будущее, т.к. нас интересует результат в целом по всем шагам. Беллман предложил рассматривать величину выигрыша от  $i$ -го шага и до конца, если  $i$ -ый шаг начинается с некоторого состояния  $S$ . Такую величину называют *условным оптимальным выигрышем*  $W_i(S)$ .

Тогда, принимая решение на каждом  $i$ - шаге, мы должны выбрать решение  $X_i$  так, чтобы условный оптимальный выигрыш был максимальным от  $i$ - шага и до конца.

**Определение:** оптимальность в малом понимается через оптимальность в большом.

Любой процесс имеет где-то окончание, т.е. имеет горизонт планирования. Тогда последний этап «не имеет будущего». Вот именно его можно оптимизировать только из условий данного этапа. После этого приступают к оптимизации  $(m-1)$ -го этапа. Выбирают такое  $X_{m-1}$ , чтобы при применении этого  $X_{m-1}$  его можно было внести в управление последнего шага. При этом мы задаём состояние, с которого начинается  $(m-1)$ -ый шаг. Поэтому функцию  $W_i(S)$  называют *условным оптимальным выигрышем*. Таким образом, процесс оптимизации разворачивается от конца к началу, и начальное состояние становится известно. Принцип Беллмана нашёл применение в методе программно-целевого планирования (любое действие планируется некоторым проектом).

### **Функциональное уравнение Беллмана.**

Назовём *состоянием системы* некоторый вектор координат:  $S=(\xi_1, \xi_2, \dots, \xi_L)$ . В некоторых задачах состояние системы – одна величина. Тогда работу системы можно представить как движение в фазовом пространстве – пространстве состояний. Назовём это - *шаговым управлением* – управлением на  $i$ -ом шаге. Рассмотрим процесс управления системой за  $m$  шагов. Функция

$W_i(S, U_i)$  называется *выигрышем на  $i$ -ом шаге*. Здесь  $S$  – состояние перед  $i$ -ым шагом, а  $U_i$  – управление на  $i$ -ом шаге.

Величина  $W_i(S, U_i)$  должна быть известна до начала динамического программирования. Если состояние перед  $i$ -ым шагом было  $S$  и мы приняли какое-то управление  $U_i$ , то система перейдёт в новое состояние  $S' = \phi_i(S, U_i)$ . Эта функция должна быть так же известна. Если эти функции не заданы, то их надо сформулировать. Введём  $W_i(S)$  – *условный оптимальный выигрыш*. Это выигрыш на всех этапах от  $i$  до конца, если  $i$ -ый шаг начинается с состояния  $S$ .

Рассмотрим  $m$  шагов. Начнём с  $(i+1)$ -го шага. Предполагаем, что мы системой управляем оптимально, величина оптимального выигрыша  $W_{i+1}(S')$ . На  $i$ -ом шаге – любое управление.  $\tilde{W}_i(S)$  – неоптимальный выигрыш, т. к. на  $i$ -ом шаге мы применяем управление  $U_i$ .

$$W_i(S) = \max_{U_i} \left\{ W_i(S, U_i) + W_{i+1}(S') \right\}; \quad S' = \phi_i(S, U_i);$$

$$\underset{\text{неизвестно}}{W_i(S)} = \max_{U_i} \left\{ \underset{\text{известно}}{W_i(S, U_i)} + W_{i+1} \left[ \underset{\text{неизвестно}}{\phi_i(S, U_i)} \right] \right\}$$

Это *функциональное уравнение Беллмана*. Для использования уравнения Беллмана оптимизацию начинают с конца:

- этап условной оптимизации

1)  $i = m$

$$W_m(S) = \max_{U_m} \{ f_m(S, U_m) \} .$$

2)  $i = m - 1$

$$W_{m-1}(S) = \max_{U_{m-1}} \left\{ f_{m-1}(S, U_{m-1}) + W_m[\phi_{m-1}(S, U_{m-1})] \right\}$$

Итак, идя от конца к началу, мы получаем последовательно:

$$W_m(S), W_{m-1}(S), \dots, W_1(S)$$

$$U_m(S), U_{m-1}(S), \dots, U_1(S)$$

Придя в начальное состояние  $W_1(S)$ , мы можем подставить  $S = S_0$  и  $W_1(S_0) =$

$W_{max}$  – это безусловный выигрыш.

- этап безусловной оптимизации

Теперь необходимо получить, идя от начала к концу по цепочке, на каждом шаге (этапе) безусловно оптимальные уравнения:

$$U_1(S) = U_1(S_0) = U_1'$$

$$\phi_1(S, U_1) = \phi(S_0, U_1) = S_1'$$

$$U_2(S) = U_2'$$

$$\phi_2(S, U_2) = \phi(S_1', U_2')$$

Итак, в конце мы получаем безусловно оптимальное решение:

$$U_1', U_2', \dots, U_m'; W_{\max}$$

**Пример:** транспортная задача.

## Заключение

Задача векторной динамической оптимизации принятия решения связана с преодолением проблем редукции, нормализации и скаляризации. Предварительная редукция множества решений может быть осуществлена непосредственно в пространстве критериев на основе методов поиска множеств компромисса (множеств Парето).

Фундаментальную роль в синтезе оптимальных управлений для линейной постановки задачи и среднеквадратического критерия оптимальности имеет принцип разделения. В соответствии с ним задача стохастического оптимального управления решается в два этапа: этап стохастического оценивания состояния объекта и этап поиска детерминированных управляющих воздействий, линейно связанных с оценками состояния.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 11*

### **Методы теории игр в задачах принятия решений**

#### СОДЕРЖАНИЕ

Введение

Учебные вопросы:

1. Схема подготовки и принятия решения в организационных системах.  
Элементы теории игр.
2. Классификация игр. Антагонистические и матричные игры.
3. Игры с чистыми и смешанными стратегиями.

Заключение

## Литература:

1. Щекотихин В.М., Терентьев В.М. Прикладная математика.- Орел: Академия ФАПСИ, 2002.
2. Терентьев В.М. Методика обоснования требований к показателям качества АСМКРС. -Л.: ВАС, 1991.
3. Терентьев В.М., Паращук И.Б. Теоретические основы управления сетями многоканальной радиосвязи. - С-Петербург: ВАС, 1995.
4. Таха Х. Введение в исследование операций . В двух книгах .Пер. с англ. М.: Мир 1985 .
5. Дюбин Г.Н., Суздаль В.Г. Введение в прикладную теорию игр. – М.: Наука, 1981.
6. Петросян Л.А. и др. Теория игр: Учебное пособие.- М.: Высш. Шк., Книжный дом «Университет», 1998.



## Введение

Теория, описывающая конфликтные ситуации с количественной стороны, называется *теорией игр*. Интересы между сторонами могут быть полностью противоположными. Такие модели называются *антагонистическими играми*. Но во многих ситуациях в игре могут принимать участие три и более сторон. Такие игры называются *множественными*. Некоторые стороны могут объединяться по интересам. Такие игры называются *коалиционными*.

*Игра* – модель ситуации, где зафиксированы сами игроки, правила игры, определённые выигрыши после каждого хода, правила окончания игры. В более сложных играх совокупность ходов определяет некоторую *стратегию*. Мы будем рассматривать только парные игры, в которых есть два игрока, и интересы которых полностью противоположны – *антагонистические парные игры*. Если игрок *A* выиграл  $a$ , то игрок *B* выиграл  $-a$  (потерял  $a$ ). Поэтому такие игры называются так же *играми с нулевой суммой*.

Главным в игровой модели является то, что другая сторона – противник, активно противодействует вам в выборе оптимального решения. При этом меняется само понятие оптимального решения. В дальнейшем мы покажем, что принцип *согласованного оптимума* является основой игры.

Ходы в игре могут быть *чистые* и *случайные*. Чистый ход зависит от сознательного решения стороны, а случайный ход – результат случайного механизма, который иногда применяется специально, а иногда случайно вовлекается в игру. Например, часто азартные игры состоят из одних случайных ходов. *Стратегией игрока* называется совокупность правил, определяющих выбор вариантов действий при каждом личном ходе игрока в зависимости от сложившейся ситуации. Если количество стратегий конечно – *игра конечная*, в противном случае – *бесконечная игра*.

Задачей теории игр является обоснование выбора оптимальных стратегий обоих игроков. В теории игр считается, что игра повторяется многократно и игроков интересует наибольший средний выигрыш. С учетом интересов игроков, при подходе к выработке оптимального решения приходится применять тот или иной принцип оптимальности.

## *Схема подготовки и принятия решения в организационных системах.*

### *Элементы теории игр.*

Основными понятиями теории игр являются:

- множество игроков;
- пространство стратегий игроков;
- множество допустимых исходов (состояний объекта приложения усилий);
- функция выигрыша (потерь) и множество ее значений, которое соответствует принимаемым игроками пошаговым решениям и определяет платежную матрицу и конечную цену игры;
- седловая точка игры, соответствующая компромиссным стратегиям игроков и обеспечивающая равенство нижней и верхней цены игры.

Целью игры является получение каждым игроком чистых либо смешанных стратегий управления, гарантирующих выигрыш (проигрыш) в числовых показателях качества в худших условиях функционирования объекта, т.е. цену игры, соответствующую ее седловой точке. Значительный вклад в развитие теории игр внесли Э.Борель, Дж.Фон Нейман, Дж.Нэш, Давыдов, Воробьев и др.

**Задача относится к теории игр, если:**

- *результат решения задачи зависит от решения двух или более лиц, которые принимают эти решения независимо;*
- *решения ЛПР-участниками-игроками принимаются в условиях неопределенности.*

**Постановка обобщенной задачи:**

Пусть имеется два лица (первое и второе) и оба эти лица стремятся получить максимальную выгоду. Следовательно, имеется две целевые функции:

$Q_1(x,y)$  – функция выигрыша первого лица,

$Q_2(x,y)$  – функция выигрыша второго лица,

где  $x,y$  – решения, принимаемые соответственно первым и вторым лицом.

Значение целевой функции  $Q_1(x,y)$  первого игрока зависит не только от его решения  $x$ , которое примет он, но и от решения  $y$ , которое примет второй игрок. То же самое можно сказать и о целевой функции  $Q_2(x,y)$  второго лица.

Если бы решение  $y$  второго игрока было бы точно известно, то для первого игрока выбор оптимального решения  $x^*$  был бы традиционным

$$x^* = \operatorname{argmax} Q_1(x,y), \quad x \in X$$

где  $y'$  – известное решение второго лица,  $X$  – множество возможных решений первого лица.

Дополнительно можем ввести новую оценочную функцию  $Q_1(x, y)$ , которая позволяет сравнивать какое из решений  $x_1$  или  $x_2$  первого лица "лучше".

Оценивать "превосходство" решения  $x$  можно по среднему значению целевой функции  $Q_1$ :

$$\bar{Q}_1(x) = \int_{y_n}^{y_n} Q_1(x, y) P(y) dy.$$

Эта оценка учитывает вклад каждого решения второго лица и вероятность принятия таких решений. Однако, в этом случае необходимо знать вероятность принятия решения вторым лицом или плотность распределения этой вероятности. Если  $P(y)$  не известно, то вычислить  $\bar{Q}_1(x)$  невозможно и следует выбрать новую оценку "хорошего" решения.

Естественной оценкой в этом случае является "наихудшее" (минимальное) значение целевой функции  $Q_1(x)$ :

$$\tilde{Q}_1(x) = \min_{y \in Y} Q_1(x, y).$$

Такая оценка слишком осторожна: ее называют *гарантированным выигрышем*.

Очевидно, первому игроку хотелось бы выбрать такое  $x$ , при котором гарантированный выигрыш  $Q_1(x)$  принял бы максимальное значение

$$\tilde{Q}'_1 = \max_{x \in X} \tilde{Q}_1(x) = \max_{x \in X} \min_{y \in Y} Q_1(x, y).$$

Решением игровой задачи является нахождение наилучшей равновесной ситуации. Эта ситуация называется оптимальной.

### ***Классификация игр. Антагонистические и матричные игры.***

Классификация игр, используемых в игровых задачах, представлена на рисунке.

По степени временной зависимости множества стратегий игроков и элементов платежной матрицы различают динамические (дифференциальные или многошаговые) и статические игры. По количеству возможных стратегий игры их разделяют на конечные и бесконечные. По количеству имеющейся априорной информации определяют: игру с полной информацией о состоянии игры, о стратегиях действий сторон и о платежной матрице; игру с неполной (вероятностной) информацией о функциях выигрыша, о стратегиях сторон, об осведомленности сторон (каналах наблюдения); игры в условиях полной неопределенности.

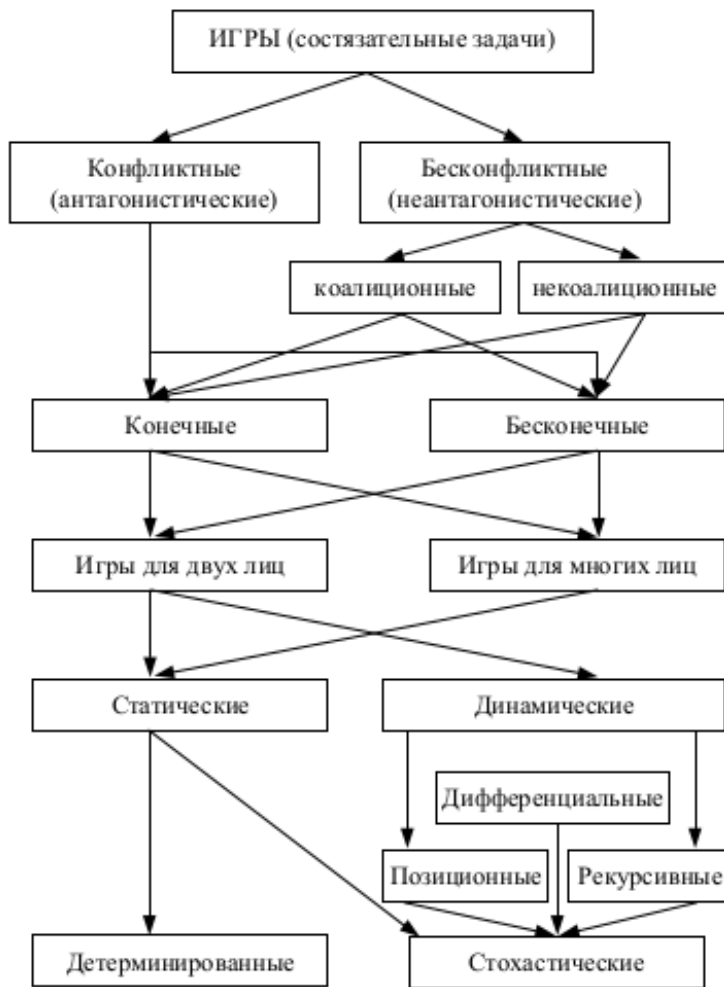


Рис. 1. Классификация игровых задач.

Если в игре участвует два игрока с противоположными интересами (целями) игру называют антагонистической, а если выигрыш одного из них равен проигрышу другого то антагонистической игрой с нулевой суммой. Если ряд игроков объединяются в группы с целью совместного достижения общей цели в игре с другой группой, то игра называется коалиционной. Если процесс игры разворачивается во времени с поочередным ходом каждого из игроков, то игра называется позиционной, а принятие решения на каждом ее шаге может быть связано с наличием либо отсутствием информации о предыдущих (будущих) стратегиях игроков, о текущем состоянии объекта и величине дохода. Наконец, антагонистическая одношаговая игра двух игроков с противоположными интересами с известной платежной матрицей называется матричной игрой с нулевой суммой. Так как к данному типу игр в нормальной форме сводятся большинство позиционных игр далее детально рассмотрим постановку и возможные методы решения матричных игр.

Пусть задана матричная игра, в которой первый игрок  $X=\{x_i\}$  имеет  $m$  возможных стратегий управления, второй игрок  $Y=\{y_j\}$  имеет  $n$  стратегий, а прямоугольная матрица платежей  $G=\{g_{ij}\}$  размерностью  $\{n \times m\}$  определяет выигрыши (проигрыши) каждого из игроков при применении ими стратегий  $x_i$  и  $y_j$ . Целью игры является отыскание первым из игроков стратегии, доставляющей максимум выигрыша  $g_{ij}$ , а вторым игроком стратегии соответствующей минимуму проигрыша  $g_{ij}$ . В зависимости от первенства принятия решений игроками возможно два сценария игры, определяющие различные значения функции потерь.



**Нижней ценой** матричной **игры** с нулевой суммой – минимальным гарантированным выигрышем первого игрока в наихудших условиях назовем величину выигрыша (потерь), соответствующую случаю, когда вначале ходит второй игрок, формируя вектор размера  $\{j\}$  минимальных доходов  $\{g_{(j)imin}\}$ , элементы которого определены по  $i$ -м строкам матрицы  $\{g_{ij}\}$ , соответствующий его условно оптимальным стратегиям  $\{j_{opt}/i\}$ , а затем первый игрок, выбирает  $i_{opt}$ -ю стратегию, соответствующую максимальному элементу вектора доходов  $\{g_{i(j)min}\}$ , определяя тем самым наилучшую пессимистичную стратегию:

$$\alpha = \max_i \min_j g_{ij} \quad (1).$$

**Верхней ценой** **игры** – максимальным гарантированным выигрышем первого игрока назовем величину выигрыша (потерь), соответствующую обратной последовательности принятия решений, т.е. вначале первый игрок определяет максимальные элементы  $j$ -х столбцов матрицы доходов  $\{g_{ij}\}$ , формируя тем самым вектор доходов  $\{g_{(i)j \max}\}$ , соответствующий условно оптимальным стратегиям первого игрока  $\{i_{opt}/j\}$ , а затем второй игрок определяет стратегию  $j_{opt}$ , соответствующую минимальному проигрышу в этом векторе доходов и определяющую безусловную оптимистичную  $j$ -ю стратегию второго игрока:

$$\beta = \min_j \max_i g_{ij}. \quad (2)$$

Нижняя и верхняя цены игры связаны соотношением:

$$\alpha \leq \beta. \quad (3)$$

В случае строгого равенства в выражении (3) говорят о наличии седловой точки  $g_{i_0j_0}$  в матрице  $G$ , а соответствующие ей стратегии  $x_{i_0}$  и  $y_{j_0}$  называют оптимальными (компромиссными) решениями игры в чистых стратегиях.

В случае  $\alpha < \beta$  матрица  $G$  не имеет седловой точки (стабильной игры) в чистых стратегиях и необходимо переходить к ее поиску в смешанных стратегиях. Смешанной стратегией первого игрока называется вектор  $p = \{p_1, \dots, p_n\}$ ,

где  $p_i$  – вероятность применения чистой стратегии  $x_i$ , удовлетворяющая условиям  $0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1$ . Аналогично смешанная стратегия второго игрока представляет собой вектор  $\vec{q} = \{q_1, \dots, q_m\}$ , где  $q_j$  представляет собой вероятность применения вторым игроком чистой стратегии  $y_j$ . Цена игры должна теперь определяться математическим ожиданием выборочных значений выигрыша (проигрыша) игроков с учетом вероятностей появления различных стратегий игроков за период игры:

$$Z(\vec{p}, \vec{q}) = \sum_{i=1}^n \sum_{j=1}^m p_i q_j g_{ij}. \quad (4)$$

## *Игры с чистыми и смешанными стратегиями.*

### **Основная теорема теории игр:**

«Каждая конечная игра имеет по крайней мере одно решение, возможно в области смешанных стратегий. Цена игры заключена в пределах  $\alpha \leq G \leq \beta$  ». Тогда, согласно теореме Нэша [1,3] игра в смешанных стратегиях всегда имеет седловую точку (равенство нижней и верхней цен игры), а соответствующая ей цена игры  $\gamma$  в смешанных стратегиях имеет вид:

$$\gamma(p^{opt}, q^{opt}) = \max_{\vec{p}} \min_{\vec{q}} Z(\vec{p}, \vec{q}) = \min_{\vec{q}} \max_{\vec{p}} Z(\vec{p}, \vec{q}). \quad (5)$$

Для поиска нижней цены игры (средний член выражения (5) *второй* игрок определяет вектор *чистых* стратегий  $\{y_i\}$ ,  $i=1, \dots, n$ , доставляющий минимум выигрыша по всем  $i$ -м строкам матрицы доходов  $\{g_{i \min}\}$ , а далее ищется вектор вероятностей применения стратегий (смешанная стратегия) *первого* игрока  $\{p_i^{opt}\}$ ,

максимизирующих его выигрыш  $Z(\vec{y}, \vec{p}^{opt}) = \max_{\vec{p}} \sum_{i=1}^n g_{i \min} p_i \leq \gamma$ , с учетом ограничений.

Поиск верхней цены игры включает выбор первым игроком вектора *чистых* стратегий  $\{x_j\}$ ,  $j=1, \dots, m$ , соответствующего максимальным элементам  $j$ -х столбцов матрицы доходов  $\{g_{j \max}\}$ , и последующий поиск значений компонент

вектора  $\vec{q}^{opt} = (q_j^{opt})$ , минимизирующего доход первого игрока. Полученное при этом значение выигрыша оказывается равным  $Z(\vec{q}^{opt}, \vec{x}) = \min_{\vec{q}} \sum_{j=1}^m g_{j \max} q_j \geq \gamma$ .

Для случая равенства нижней и верхней цен игры получаем оптимальные или компромиссные решения в смешанных стратегиях, физический смысл которых заключается в соблюдении игроками оптимальных значений вероятностей появления имеющихся в их распоряжении чистых стратегий на всем периоде игры. Отклонение от найденных компромиссных стратегий ведет к проигрышу отклонившегося игрока.

Так как решение игры в смешанных стратегиях является общим случаем по сравнению с игрой в чистых стратегиях, то далее рассмотрим методы решения игровых задач для общего случая.

## Применение математических методов в игровых задачах

### 1. Применение симплекс – метода в задачах поиска компромиссных стратегий.

В соответствие с введенным понятием цены игры в седловой точке

$z(\vec{p}^{opt}, \vec{q}^{opt}) = \chi(\vec{p}^{opt}, \vec{q}^{opt}) = \sum_{i=1}^n \sum_{j=1}^m p_i^{opt} q_j^{opt} g_{ij}$ . и записи его для случая применения

произвольных стратегий со стороны одного из игроков можно записать:

$$\begin{aligned} Z_j(y_i, \vec{p}^{opt}) &= \sum_{i=1}^n g_{i(j)} p_i \geq \gamma, \\ Z_i(x_j, \vec{q}^{opt}) &= \sum_{j=1}^m g_{(i)j} q_j \leq \gamma. \end{aligned} \quad (6)$$

где учтено, что при применении вторым игроком произвольной чистой стратегии  $j$  (т. е. при отклонении его поведения относительно оптимальной вероятностной стратегии, соответствующей седловой точке) первый игрок, придерживающийся оптимальной стратегии  $\vec{p}^{opt}$ , достигает в игре результат, превосходящий  $\gamma$  и наоборот (6).

Одновременно, учитывая условие нормирования для искомым векторов вероятностей стратегий первого и второго игроков

$$\begin{aligned} \sum_{i=1}^n p_i &= 1; \\ \sum_{j=1}^m q_j &= 1, \end{aligned} \quad (7)$$

и с учетом масштабирования выражений (6, 7) на величину  $\gamma$ , получим две (прямую и двойственную) задачи линейного программирования:

$$\begin{aligned} \sum_{i=1}^n p'_i = \frac{1}{\gamma} = U &\Rightarrow \min_{\bar{p}}, \\ \sum_{i=1}^n g_{ij} p'_i &\geq 1; \\ \sum_{j=1}^m q'_j = \frac{1}{\gamma} = W &\Rightarrow \max_{\bar{q}}, \\ \sum_{j=1}^m g_{ij} q'_j &\leq 1, \end{aligned} \quad (8)$$

где  $p'_i, q'_j$  - нормированные на величину  $\gamma$  вероятности использования стратегий игроков.

Выбор цели оптимизации переменных в прямой и двойственной задачах (8) определяется стремлением к равенству в ограничениях для них.

Решение данных задач ЛП могут быть найдены обычным симплекс – методом, а пересчет найденных оптимальных векторов в значения вероятностей появления стратегий игроков осуществляется следующим образом:

$$\gamma = \frac{1}{W} = \frac{1}{V}; p_i = p'_i \cdot \gamma; q_j = q'_j \cdot \gamma.$$

Обычно начинают решение той задачи из двух, для которой меньше число ограничений (проще многоугольник решений).

## 2. Итерационный метод решения игровой задачи

Итерационный метод включает в себя следующие этапы:

- а) выбор произвольной стратегии первого игрока  $x_i$  в первой партии;
- б) выбор стратегии второго игрока  $y_j$ , при которой первый игрок получает минимальный выигрыш  $\min g_{ij}$ ;
- в) стратегия  $y_j$  принимается активной для второго игрока, для него определяются потери при любой стратегии первого игрока и определяется та из них  $x'_i$ , которая доставляет максимум выигрыша  $\max g_{ij}$ ;

г) стратегию  $x'_i$  принимают в качестве активной для первого игрока и отыскивается та из стратегий второго игрока  $y'_j$ , которая обеспечивает минимум суммарного за два шага выигрыша первого игрока.

Далее розыгрыш продолжается до момента совпадения суммарного выигрыша и проигрыша за все шаги игры либо до момента когда их разность станет меньше некоторой наперед заданной величины  $(\beta_\Sigma - \alpha_\Sigma) \leq \varepsilon$ .

Вероятности применения тех либо других стратегий игроков в ходе всей партии принимаются в качестве векторов вероятностей стратегий первого и второго игроков, а число возможных итераций может достигать величин порядка сотен и тысяч шагов.



## Пример.

Задача №1. Отыскание пессимистической стратегии выбора для корреспондента частоты передачи в сети оповещения в условиях радиоподавления.

В целях повышения надежности передачи циркулярных сообщений с заданным качеством в радиосети выделяется группа частот  $\{f_1, \dots, f_i, \dots, f_N\}$ . На каждой из частот известно значение отношения сигнал/помеха для  $j$ -го корреспондента сети  $h_{ij}$ , определяемое воздействием преднамеренной помехи на данной частоте. Задача заключается в выборе из заданной группы частот той из них, для которой отношение сигнал/помеха у корреспондента радиосети, находящегося в наихудших по помехе условиях, окажется наибольшим. Решим задачу методом прямого перебора в терминах матричной игры.

Пусть матрица доходов (потерь) – отношений сигнал/помеха  $\{h_{ij}\}$  на  $i$ -х частотах у  $j$ -х корреспондентов в радиосети передачи циркулярных сообщений имеет вид

	Кор.1	Кор.2	Кор.3	Кор.4	max min $h_{ij}$	min max $h_{ij}$
F <sub>1</sub>	5	10	18	25		25
F <sub>2</sub>	8	7	28	23		
F <sub>3</sub>	21	18	12	20		
F <sub>4</sub>	30	22	19	15	15	
F <sub>5</sub>	3	40	10	7		
F <sub>6</sub>	50	5	15	22		

В соответствии с целью решаемой задачи и алгоритмом поиска чистых стратегий игры находим максиминные стратегии  $i$  и  $j$ , для чего вначале на каждой  $i$ -й частоте определяем минимальные  $j$ -е элементы матрицы доходов  $\{h_{i(j)min}\}$ , а затем

выбором максимального из них получаем искомые оптимальные стратегии  $i_{opt}=4$  и  $j_{opt}=4$  (осуществлять передачу циркулярных сообщений в радиосети на частоте  $f_4$  и ставить помеху в направлении корреспондента №4), а также соответствующее им гарантированное в условиях наихудших помех значение отношения сигнал/помеха  $h_{i_0j_0низк}=15$ , являющееся нижней цены игры.

Наряду с пессимистичным вариантом игры представляет интерес поиск оптимистичного (минимаксного) варианта, соответствующего верхней цене игры. Поэтому для каждой  $j$ -й стратегии подавления определяем максимальные по  $i$  элементы столбцов матрицы доходов, а затем из полученного вектора максимальных доходов  $h_{(i)jmax}$  ищем  $j$ -й элемент, соответствующий минимальному значению отношения сигнал/помеха, а следовательно определяющий оптимистичные стратегии игры. Ими являются стратегии  $i=1$  и  $j=4$  (использовать для передачи частоту  $f_1$  и подавлять корреспондента №4), а верхняя цена игры оказывается при этом равной  $h_{i_0j_0верх}=25$ .

Как следует из сравнения нижней и верхней цены игры в ней отсутствует седловая точка, т.е. нет решения в чистых стратегиях.

### Пример.

Задача №2. Пример использования итерационного метода в задаче поиска смешанных стратегий для сигнала ППРЧ в условиях радиоподавления.

Пусть задано множество возможных частот, используемых для случайных скачков рабочей частоты сигнала с ППРЧ и являющихся - стратегиями игрока  $\{A_{ij}\}$ , и варианты их подавления узкополосной помехой противника  $\{B_{ij}\}$ , а также соответствующая им платежная матрица  $\{C_{ij}[\text{мбит/с}]\}$  с элементами, равными пропускной способности направления связи при работе на  $i$ -й рабочей частоте в условиях  $j$ -го варианта подавления.

	$B_1$	$B_2$	$B_3$
$A_1$	1	4	2
$A_2$	7	3	4
$A_3$	6	6	0

Необходимо определить компромиссные стратегии участников игры (закон прыгания по частотам сигнала в линии связи и закон изменения частоты помехи со стороны противника), удовлетворяющие седловой точке игры. В таблице №1 представлены результаты пошагового решения игры итерационным алгоритмом.

Таблица №1.

Номер партии	Выбор стратегии	Суммарный выигрыш $C_j$			Ниж. цена игры $\alpha$	Выбор стратегии	Суммарный выигрыш $C_i$			Верх. цена игры $\beta$	$\beta - \alpha = \varepsilon$
		$B_1$	$B_2$	$B_3$			$A_1$	$A_2$	$A_3$		
1	$A_2$	7	3	4	3	$B_2$	4	3	6	6	3
2	$A_3$	13	9	4	4/2	$B_3$	6	7	6	7/2	3/2
3	$A_2$	20	12	8	8/3	$B_3$	8	11	6	11/3	3/3

4	A <sub>2</sub>	27	15	12	12/4	B <sub>3</sub>	10	15	6	15/4	3/4
5	A <sub>2</sub>	34	18	16	16/5	B <sub>3</sub>	12	19	6	19/5	3/5
6	A <sub>2</sub>	41	21	20	20/6	B <sub>3</sub>	14	23	6	23/6	3/6
7	A <sub>2</sub>	48	24	24	24/7	B <sub>2</sub>	18	26	12	26/7	2/7
8	A <sub>2</sub>	55	27	28	27/8	B <sub>2</sub>	22	29	18	29/8	2/8
9	A <sub>2</sub>	62	30	32	30/9	B <sub>2</sub>	26	32	24	32/9	2/9
10	A <sub>2</sub>	69	33	36	33/10	B <sub>2</sub>	30	35	30	35/10	2/10

Решение игры находим статистической обработкой результатов игры:

1. Цена игры:  $v = \frac{\alpha + \beta}{2} = \frac{33/10 + 35/10}{2} = 3,4$  [мбит/с].

2. Вероятности применения стратегий:  $p_i = \frac{k}{n}$ , где  $k$ -число применений  $i$ -й стратегии;  $n$ - общее число шагов игры.

3. Окончательно имеем:

*компромиссные стратегии игроков*  $P_A = [0;0,9;0,1]$ ;  $P_B = [0;0,5;0,5]$ .

Таким образом, для получения гарантированной пропускной способности линии связи  $C_{компр.} = 3,4$  мбит/с, использующей сигналы с ППРЧ в условиях радиоподавления на отдельных частотах, необходимо организовать режимы смены рабочих частот в линии (работать девять десятых длительности сеанса связи на частоте  $f_2$ , одну десятую времени - на частоте  $f_3$  и частоту  $f_1$  - не использовать) и режимы смены частоты постановки помех (ставить помеху на частотах  $f_2$  и  $f_3$  в равных временных пропорциях, а подавление частоты  $f_1$  не применять) в соответствии с компромиссными стратегиями. Отклонение от указанных стратегий одной из сторон приведет только к выигрышу противоположной стороны, поэтому полученная цена игры является гарантированной.

## **Заключение.**

Класс задач оптимизации решений, характеризующихся наличием антагонизма сторон, априорной неопределенности относительно конкретных пошаговых внешних воздействий на объект (из множества возможных) с целью снижения качества его функционирования, а также набором собственных стратегий управления объектом, принимаемых с целью компенсации наносимого ущерба, относится к классу задач, решаемых методами теории игр [1,6,7].

Рассмотренные методы теории игр позволяют проводить оптимизацию решений в условиях, когда методы экстремального управления оказываются малоэффективными, например, в случае большой инерционности контура управления объекта по сравнению с динамикой изменения стратегий противника. Полученные на основе игрового подхода компромиссные стратегии игроков обеспечивают гарантированный выигрыш (проигрыш) в значении показателей качества функционирования объекта в худших условиях его функционирования. При этом инерционность и затраты ресурсов на реализацию смешанных стратегий оказываются несравненно ниже, чем для экстремальных систем управления, что может быть определяющим при выборе способа построения системы управления.

# «Теория принятия решений»

ст. преп. каф. СС и ПД  
Владимиров Сергей Александрович

## *Лекция 12*

### **Методы анализа временных рядов. Марковские процессы и модели.**

#### СОДЕРЖАНИЕ

Введение

Учебные вопросы:

1. Модели временных рядов.
2. Рекуррентный алгоритм оценки параметров временного ряда, оптимальный по критерию наименьших квадратов.
3. Методы прогноза временных рядов.
4. Марковские процессы и модели. Марковские модели непрерывных и дискретных процессов.

Заключение

## Литература:

1. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания.- М.: Наука, 1987.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. – М.:Наука,1973.
3. Терентьев В.М. и др. Управление качеством обслуживания в корпоративных сетях спутниковой связи. - Учебное пособие. – Орел: Академия ФСО России, 2006.
4. Анфилатов В.С. Системный анализ в управлении [Электронный ресурс] : учебное пособие / - Москва : Финансы и статистика, 2013.
5. Бородачёв С. М. Теория принятия решений [Электронный ресурс] : учебное пособие / Бородачёв С. М. - Екатеринбург : Уральский федеральный университет, 2014.
6. Демидова, Л. А. Принятие решений в условиях неопределенности [Электронный ресурс] / Л. А. Демидова, В. В. Кираковский, А. Н. Пылькин. - М. : Горячая линия–Телеком, 2012.

## Введение

Надежность оценок параметров потоков требований на ресурс используемого оборудования, получаемых на основе широко используемых в настоящее время параметрических методов стохастического оценивания, относительно невысока из-за наличия априорной неопределенности относительно их вероятностно-временных характеристик [1-3]. Повысить степень обоснованности прогноза трафика разноприоритетных пользователей в этих условиях возможно на основе обработки реальных наблюдений (статистических временных рядов) за параметрами потоков запросов или требований методами математической статистики [1].

К основным задачам моделирования временных рядов и прогнозирования их параметров относятся: выбор подходящей параметрической модели временного ряда - временной последовательности значений интенсивности потока запросов, оценка ее параметров, диагностика качества, а также получение выражений для прогноза параметров ряда (экстраполяции значений интенсивности нагрузки на упреждающий момент времени). Для обеспечения состоятельности, достаточности и эффективности результатов обработки имеющихся статистических данных необходимо выполнение ряда условий для модели, например, стационарности и эргодичности и обеспечение времени обработки (оценивания параметров потока), меньшего, чем интервал корреляции значений временного ряда.



## *Модели временных рядов.*

### *Модель линейной регрессии.*

Под этапом выбора модели понимается обоснование некоторого класса стохастической модели и идентификации ее параметров на основе знания автокорреляционных функций элементов временного ряда, а также метода диагностической проверки модели.

Наличие статистической зависимости значений одного случайного процесса в два различных момента времени  $x(t_k)$ ,  $x(t_{k-1})$  либо значений двух различных процессов в совпадающие моменты времени  $y(t_k)$ ,  $x(t_k)$  делает возможным введение моделей изменения их состояния на основе уравнения прямой линейной (приближенной для негауссовского случая) регрессии  $x(t_k) = f_{xx}(x(t_{k-1}))$ ;  $y(t_k) = f_{yx}(x(t_k))$  в следующем виде [1-3]:

$$\begin{aligned} x(t_k) &\simeq \alpha_{xx} x(t_{k-1}) + \beta_{xx}(t_k); \\ y(t_k) &\simeq \alpha_{yx} x(t_k) + \beta_{yx}(t_k), \end{aligned} \quad (1)$$

где  $\alpha_{xx} = (n \sum_{k=1}^n x(k-1)x(k) - \sum_{k=1}^n x(k-1) \sum_{k=1}^n x(k)) / (n \sum_{k=1}^n x^2(k-1) - (\sum_{k=1}^n x(k-1))^2)$ ,

$\alpha_{yx} = (n \sum_{k=1}^n x(k)y(k) - \sum_{k=1}^n x(k) \sum_{k=1}^n y(k)) / (n \sum_{k=1}^n x^2(k) - (\sum_{k=1}^n x(k))^2)$  - коэффициенты

регрессии  $X(t_k)$  на  $X(t_{k-1})$ , либо  $Y$  на  $X$  в совпадающие моменты времени;

$n$  - общее число обрабатываемых элементов (размер) выборки;

$$\beta_{xx} = \frac{(\sum_{k=1}^n x^2(k-1) \sum_{k=1}^n x(k) - \sum_{k=1}^n x(k-1) \sum_{k=1}^n x(k-1)x(k)) / (n \sum_{k=1}^n x^2(k-1) - (\sum_{k=1}^n x(k-1))^2)}{}$$

$$\beta_{yx} = \frac{(\sum_{k=1}^n x^2(k) \sum_{k=1}^n y(k) - \sum_{k=1}^n x(k) \sum_{k=1}^n x(k)y(k)) / (n \sum_{k=1}^n x^2(k) - (\sum_{k=1}^n x(k))^2)}{}$$
 - коэффициенты

сдвига.

Знак приближенного равенства говорит об аппроксимации точной, но неизвестной априори зависимости поведения процессов, прямой линейной регрессии. Коэффициенты уравнения регрессии в выражении (1) определены на основе метода наименьших квадратов, то есть оптимальны в смысле минимума среднего квадратичного отклонения истинного значения прямой регрессии от значения, полученного на основе модели. Первое уравнение в выражении (1) является уравнением авторегрессии, а второе – взаимной регрессии  $y$  на  $x$ .

Недостатком представленной модели является то, что вычисления коэффициентов уравнения  $\alpha$  и  $\beta$  проводятся по одноразовым измерениям пар значений  $(x,y)$  в пределах каждой выборки, что делает их оценку недостаточно состоятельной. При возможности проведения серии набора статистических данных о парах  $(x,y)$  большого размера  $n$ , в которой их значения повторяются  $n_{xy}$  раз, целесообразно ввести понятие выборочного коэффициента взаимной корреляции

$r_{yx} = \frac{\sum_{i=1}^n n_{xy} xy - n \bar{x} \bar{y}}{n \tilde{\sigma}_x \tilde{\sigma}_y}$ , который связан с коэффициентом регрессии соотношением

$a_{yx} = r_{yx} \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}$ , а также использовать известные тождества

$$\sum_{i=1}^n x_i = n \bar{x}; \quad \sum_{i=1}^n y_i = n \bar{y}; \quad \sum_{i=1}^n x^2 = nM(x^2); \quad \sum_{i=1}^n xy = \sum_{i=1}^n n_{xy} xy,$$

где  $n_{xy}$ - число случаев, в которых наблюдалась пара чисел  $x$  и  $y$ ;

$M(y) = m_y = \bar{y}$  - математическое ожидание процесса  $y$ ;

$M(x) = m_x = \bar{x}$  - математическое ожидание процесса  $x$ ;

$\tilde{\sigma}_y, \tilde{\sigma}_x$  - выборочные средние квадратические отклонения процессов.

Можно показать [1], что при этом уравнения (1), будут более представительно отражать статистические связи процессов и могут быть переписаны в виде

$$\begin{aligned} x(t_k) &\simeq \alpha_{xx} x(t_{k-1}) + \beta_{xx}(t_k) = m_{x_k} + r_{xx} \frac{\tilde{\sigma}_{x_k}}{\tilde{\sigma}_{x_{k-1}}} (x(t_{k-1}) - m_{x_{k-1}}); \\ y(t_k) &\simeq \alpha_{yx} x(t_k) + \beta_{yx}(t_k) = m_y + r_{yx} \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x} (x(t_k) - m_x), \end{aligned} \quad (2)$$

где  $r_{xx}, r_{yx}$  - коэффициенты авто и взаимной корреляции процессов.

Алгоритмы получения эмпирических (выборочных) оценок числовых характеристик случайных последовательностей (среднего, дисперсии, коэффициентов корреляции) и анализ их качества представлены в лекции 4.

## Смешанная модель авторегрессии и скользящего среднего (АРСС(p,q))

Общая модель временного ряда представлена как результат действия механизма обработки предшествующих членов ряда и некоторого числа членов возбуждающей белой последовательности, т.е. в виде объединенной модели авторегрессии и скользящего среднего [1-3]:

$$x(k) = - \sum_{i=1}^p a_i x(k-i) + \sum_{i=0}^q b_i \varepsilon(k-i), \quad (3)$$

где  $a_i, b_i$  - параметры авторегрессии и скользящего среднего, оцениваемые по временному ряду на основе методов математической статистики;

$\varepsilon(k)$  - случайная белая гауссовская последовательность независимых величин с нулевым математическим ожиданием, дельта-функцией корреляции и известной дисперсией.

Если  $q=0$ ,  $b(0)=1$ , то уравнение описывает модель *авторегрессии*, которая интерпретируется как механизм, генерирующий временной ряд, в котором наблюдение переменной  $x(k)$  в некоторый момент  $k$  выражается через систематическую зависимость от той же самой переменной в  $p$  моментах времени, непосредственно предшествующих  $k$  моменту плюс значение случайного возмущения  $\varepsilon(k)$  в момент  $k$ .

Таким образом, текущее значение  $x(k)$  выходного ряда выражается через линейную комбинацию из  $p$  предшествующих отсчетов, которые называют *предсказанными наблюдениями (прогнозом)*. С целью минимизации ошибки прогноза в процессе идентификации параметров авторегрессии будем пользоваться

критерием минимума среднеквадратической погрешности (дисперсии ошибки) прогноза.

Процесс авторегрессии будет стационарным только в случае, если его параметры лежат в определенном диапазоне. Например, если имеется только один параметр, то он должен находиться в интервале  $-1 < a_1 < +1$ . В противном случае, предыдущие значения будут накапливаться и значения последующих  $x(k)$  могут быть неограниченными, следовательно, ряд не будет *стационарным*. Для нескольких параметров авторегрессии можно определить условия, обеспечивающие стационарность.

Частным случаем авторегрессионной модели при  $p=1$  является марковский процесс

$$x(k) = a(k/k-1)x(k-1) + \varepsilon(k). \quad (4)$$

В этом случае связь последовательности выборочных значений коэффициентов корреляции с соответствующими значениями коэффициентов авторегрессии имеет вид [1]

$$r_{xx}(k) = a(1)r_{xx}(k-1) = a^k(1), \text{ при } r_{xx}(0) = 1, k \geq 0. \quad (5)$$

При этом зависимость  $r_{xx}(k)$  имеет монотонный экспоненциальный характер при  $0 \leq a(1) \leq 1$  и экспоненциальный с изменением знака при  $0 > a(1) > -1$ .

Дисперсия процесса имеет вид  $\sigma_x^2 = \sigma_\varepsilon^2 / (1 - a(1)r_{xx}(1)) = \sigma_\varepsilon^2 / (1 - a^2(1))$ , где  $\sigma_\varepsilon^2 = \rho_w$  - дисперсия шума возбуждения.

При  $p=0$  смешанная модель преобразуется в модель *скользящего среднего*, реализующую механизм учета динамики изменения значений возбуждающей последовательности  $\varepsilon(k)$ . В отличие от процесса авторегрессии, в процессе скользящего среднего текущее наблюдение ряда представляет собой сумму случайного компонента  $\varepsilon(k)$  в данный момент времени и линейной комбинации взвешенных значений случайных воздействий в предыдущие  $q$  моментов времени. Определение оптимального значения параметров скользящего среднего основано на учете корреляций значений случайных воздействий на глубину  $q$  шагов.

Функциональная схема модели временного ряда на основе процессов авторегрессии и скользящего среднего представлена на рисунке 1.

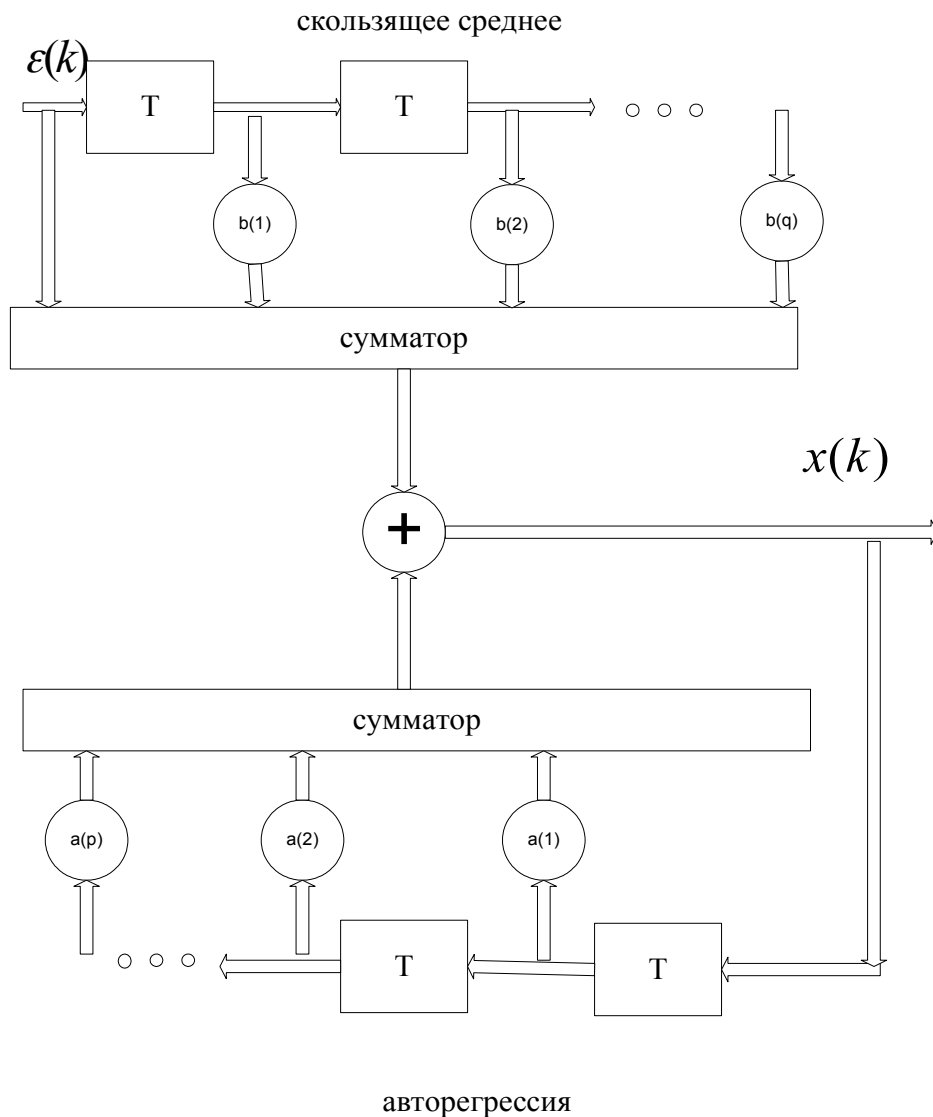


Рис.1. Функциональная схема модели временного ряда на основе АРСС (p, q)

Очень часто для получения оценок коэффициентов авторегрессии и скользящего среднего, а также дисперсии шума возбуждения в модели АРСС(p,q) используют отдельные субоптимальные процедуры их определения [1]. Вначале оценивают коэффициенты АР на основе процедуры наименьших квадратов и уравнения Юла-Уолкера, затем на их основе формируют модель временного ряда и получают разностный ряд между принятым и смоделированным, а последовательность остаточных ошибок (разностный ряд) используется в дальнейшем для определения коэффициентов СС.

### *Методы оценки параметров авторегрессии*

Рассмотрим методы оценки параметров авторегрессии и скользящего среднего на основе информации, содержащейся в выборочных данных (3-5) и рекуррентных уравнений Юла-Уолкера [1].

Если обе части уравнения (3) умножить на  $x(n-m)$  и взять математическое ожидание, то получим уравнение для частных (на определенный временной сдвиг  $m$ ) автокорреляционных функций и коэффициентов автокорреляции:

$$M(x(n)x(n-m)) = - \sum_{k=1}^p a_k M(x(n-k)x(n-m)) + \sum_{k=0}^q b_k M(\varepsilon(n-k')x(n-m)) \quad \text{или}$$

$$r_{xx}(m) = - \sum_{k=1}^p a_k r_{xx}(m-k) + \sum_{k=0}^q b_k r_{\varepsilon x}(m-k'), \quad (6)$$

где  $r_{\varepsilon x}(m-k') = \rho_w$ , при  $\{k'=0\}$  - дисперсия шума возбуждения.

Можно показать [1], что это выражение, связывающее параметры авторегрессии, скользящего среднего и временную последовательность

выборочных значений коэффициента автокорреляции, можно представить в следующем виде:

$$r_{xx}(m) = \begin{cases} r'_{xx}(-m), & m < 0; \\ -\sum_{k=1}^p a(k)r_{xx}(m-k) + \rho_w \sum_{k=m}^q b(k)h'(k-m), & 0 \leq m \leq q; \\ -\sum_{k=1}^p a(k)r_{xx}(m-k), & m \geq q, \end{cases} \quad (7)$$

где  $h(k), h(0)=1$  - импульсная характеристика формирующего фильтра.

Полагая в выражении (7)  $q=0$  и записывая полученное выражение в матричной форме, получим следующую систему из  $P$  уравнений для определения коэффициентов авторегрессии [1]

$$\begin{bmatrix} 1 \\ a(1) \\ a(2) \\ \dots \\ a(p) \end{bmatrix}^T \begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & \dots & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & \dots & r_{xx}(-p+1) \\ \dots & \dots & \dots & \dots \\ r_{xx}(p) & r_{xx}(p-1) & \dots & r_{xx}(0) \end{bmatrix} = \begin{bmatrix} \rho_w \\ 0(1) \\ \dots \\ 0(p) \end{bmatrix}$$

или в другой форме

$$\begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ \dots \\ r_{xx}(p) \end{bmatrix} = \begin{bmatrix} a(1) \\ a(2) \\ \dots \\ a(p) \end{bmatrix}^T \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(p-1) \\ r_{xx}(1) & r_{xx}(0) & \dots & r_{xx}(p-2) \\ \dots & \dots & \dots & \dots \\ r_{xx}(p-1) & r_{xx}(p-2) & \dots & r_{xx}(0) \end{bmatrix}. \quad (8)$$



Выражение (8) можно также представить в следующей векторно-матричной форме относительно вектора  $\vec{a}$

$$\vec{a} = P_p^{-1} \vec{r}_p,$$

где

$$P_p = \begin{bmatrix} 1 & r_{xx}(1) & \dots & r_{xx}(p-1) \\ r_{xx}(1) & 1 & \dots & r_{xx}(p-2) \\ \dots & \dots & \dots & \dots \\ r_{xx}(p-1) & r_{xx}(p-2) & \dots & 1 \end{bmatrix}$$

- матрица коэффициентов корреляции.

Для случая  $p=2$  система уравнений Юла-Уолкера (8) имеет вид [1]

$$\begin{aligned} r_1 &= a_1 + a_2 r_1, \\ r_2 &= a_1 r_1 + a_2, \end{aligned}$$

решая которую получаем значения коэффициентов авторегрессии

$$a_1 = \frac{r_1(1-r_2)}{1-r_1^2}; \quad a_2 = \frac{r_2-r_1^2}{1-r_1^2}.$$

Допустимые значения  $r_{xx}(m)$  для стационарного процесса АР ( $p=2$ ) должны лежать в пределах [1]

$$\begin{aligned} -1 &\leq r_1, r_2 \leq 1, \\ r_1^2 &< \frac{1}{2}(r_2+1). \end{aligned}$$

Дисперсия процесса АР (2) равна [1]  $\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-r_1 a_1 - r_2 a_2}$ .

В общем случае коэффициенты уравнения авторегрессии можно найти из решения уравнения Юла-Уолкера (8) зная последовательность из  $p+1$  отсчетов для коэффициента корреляции значений элементов ряда. Количество выполняемых при этом операций пропорционально  $p^2$ .

Полагая в (7)  $p=0$  и замечая, что  $h(k)=b(k)$  при  $1 \leq k \leq q$ , получаем выражение для определения коэффициентов скользящего среднего

$$r_{xx}(m) = \begin{cases} 0, & m \geq q; \\ \rho_w \sum_{k=m}^q b(k)b'(k-m), & 0 \leq m \leq q; \\ r'_{xx}(-m), & m \leq 0 \end{cases} \quad (9)$$

Из выражения (9) следует, что автокорреляционная последовательность и параметры скользящего среднего связаны нелинейным соотношением типа свертки, что делает их определение достаточно сложным. Относительно просто определяются коэффициенты скользящего среднего (СС) для случая  $q=1$ . В этом случае процесс СС( $q=1$ ) можно представить в виде

$\Delta x(k) = x(k) - x(k-1) = \varepsilon(k) - b_1 \varepsilon(k-1)$ , где  $b_1$  - коэффициент СС первого порядка.

Можно показать [1], что значение искомого коэффициента в этом случае определяется из решения следующего уравнения  $b_1^2 + b_1/r_1 + 1 = 0$ ,

где  $r_1 = \begin{cases} -b_1 / (1 + b_1^2), & k=1; \\ 0, & k>1 \end{cases}$  - значения автокорреляционной функции для СС (1).

Для выполнения стационарности моделируемого процесса значения коэффициентов СС должны удовлетворять условию обратимости, т.е.

$$b_1(b_1)^{-1} = 1, \quad -1 < b_1 < 1.$$

Таким образом, процесс АРСС<sub>(2,1)</sub> может быть представлен выражением

$$x(k) = a_1 x(k-1) + a_2 x(k-2) + \varepsilon(k) - b_1 \varepsilon(k-1).$$

На этапе *идентификации* порядка модели необходимо решить, какое число параметров авторегрессии ( $p$ ) и скользящего среднего ( $q$ ) должно присутствовать в эффективной и *экономной* модели процесса. На практике очень редко бывает, что число параметров  $p$  или  $q$  больше 2.

Встречаются и более сложные модели, например, учитывающие тренд. Основным ограничением при применении описанной смешанной модели является стационарность моделируемого временного ряда.

### ***Модель авторегрессии проинтегрированного скользящего среднего (АРПСС( $p,d,q$ )).***

Дальнейшим развитием модели на случай нестационарного временного ряда является модель АРПСС ( $p,d,q$ ), предложенная Боксом и Дженкинсом [1], которая включает три типа параметров: параметры авторегрессии ( $p$ ), порядок разности ( $d$ ) и параметры скользящего среднего ( $q$ ), которые вычисляются для ряда после взятия разности с лагом  $d$ . Расширение области моделирования в этой модели достигается путем перехода к моделированию не самих значений процесса  $x(k)$ , а разности между значениями ряда до  $d$ -го порядка включительно, обладающей стационарными свойствами, что означает постоянство ее среднего и неизменность во времени соответствующих выборочных дисперсии и автокорреляции. Для того чтобы определить необходимый порядок разности, нужно исследовать график исходного ряда и автокорреллограмму. Сильные изменения уровня (сильные скачки вверх или вниз) обычно требуют взятия несезонной разности первого порядка (лаг= $1$ ), а сильные изменения наклона требуют взятия разности второго порядка.

Процесс оценивания параметров авторегрессии и скользящего среднего проводится по преобразованным данным (подвергнутым применению разностного оператора). До построения прогноза нужно выполнить обратную операцию

(интегрировать данные). Таким образом, прогноз – проинтегрированное значение разности будет сравниваться с соответствующими исходными данными. На интегрирование данных указывает буква *И* в общем названии модели (АРПСС = Авторегрессионное Проинтегрированное Скользящее Среднее).

Большинство встречающихся на практике временных рядов можно с достаточной степенью точности аппроксимировать одной из 5 основных моделей, которые можно идентифицировать по виду автокорреляционной (АКФ) и частной автокорреляционной функции (ЧАКФ) – ненулевому значению АКФ для конкретного  $k \leq p$ .

1. *Один параметр (p)*: АКФ - экспоненциально убывает; ЧАКФ - имеет резко выделяющееся значение для лага 1, нет корреляций на других лагах.
2. *Два параметра авторегрессии (p)*: АКФ имеет форму синусоиды или экспоненциально убывает; ЧАКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах.
3. *Один параметр скользящего среднего (q)*: АКФ имеет резко выделяющееся значение на лаге 1, нет корреляций на других лагах. ЧАКФ экспоненциально убывает.
4. *Два параметра скользящего среднего (q)*: АКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах. ЧАКФ имеет форму синусоиды или экспоненциально убывает.
5. *Один параметр авторегрессии (p) и один параметр скользящего среднего (q)*: АКФ экспоненциально убывает с лага 1; ЧАКФ - экспоненциально убывает с лага 1.

## *Рекуррентный алгоритм оценки параметров временного ряда, оптимальный по критерию наименьших квадратов*

Реализация непрерывного управления качеством обслуживания нагрузки требует определения параметров авторегрессии модели АРСС  $(p, q)$  в темпе текущего времени в отличии от алгоритмов, основанных на решении уравнения Юла-Уолкера [1-3]. Такие возможности открываются при использовании рекуррентного алгоритма наименьших квадратов (РНК), который при поступлении новых текущих данных  $x(N+1)$  позволяет переходить от вектора коэффициентов линейного предсказания  $\vec{a}_{p,N}$  к вектору  $\vec{a}_{p,N+1}$ , не решая уравнение Юла-Уолкера [1].

Для получения алгоритма РНК выражение для ошибки линейного предсказания вперед при использовании выборки размером  $N$  для  $k$ -го временного шага и глубины регрессии  $p$  запишем в следующем виде [1]:

$$e'_{p,N}(n) = \vec{x}_{p,N}^T(n) \vec{a}'_{p,N}(n), \quad (10)$$

где  $\vec{x}_{p,N}^T = (x(n); x(n-1) \dots x(n-p))$  - вектор значений временного ряда размерностью  $p+1$ ;

$\vec{a}'_{p,N}(n) = (1 ; \{a'_{p,N}(n-1) \dots a'_{p,N}(n-p)\})^T$  - вектор значений коэффициентов авторегрессии размерностью  $p+1$ .

Так как суммирование в выражении для ошибки осуществляется с учетом отрицательного знака при коэффициентах авторегрессии, то в результате получаем разность между значением ряда в момент времени  $k$  и взвешенными значениями регрессии на прошлые значения ряда глубиной  $p$ .

Введем понятие суммы экспоненциально взвешенных квадратов ошибок предсказания на всей длине выборки  $\rho'_{p,N} = \sum_{k=1}^N w^{N-k} (e'_{p,N}(n))^2$ .

Можно показать [1], что вектор коэффициентов линейного предсказания  $\vec{a}_{p,N}$ , минимизирующий сумму экспоненциально взвешенных с весом  $w^{N-k}$  квадратов ошибок  $\rho'_{p,N}$ , удовлетворяет решению уравнения

$$R_{p,N} \vec{a}_{p,N} = \begin{pmatrix} \rho'_{p,N} \\ \mathbf{0}_p \end{pmatrix}, \quad (11)$$

где  $R_{p,N} = \begin{bmatrix} r_{p,N}(0,0) & r_{p,N}^T \\ r_{p,N} & R_{p-1,N-1} \end{bmatrix}$  - рекуррентная матрица коэффициентов автокорреляции;

$\vec{a}_{p,N} = [1, a_{1,N} \dots a_{p,N}]^T$  - вектор коэффициентов авторегрессии;

$r_{p,N}(0,0) = \sum_{n=1}^N w^{N-n} (x(n))^2$  - взвешенная дисперсия наблюдаемой последовательности.

Тогда основу базового РНК-алгоритма составляют следующие выражения для векторов коэффициентов предсказания, коэффициентов усиления и дисперсии ошибки фильтрации [1]:

$$\vec{a}_{p,N+1} = \begin{cases} \vec{a}_{p,N} - P_N \vec{x}'_{p-1}(N) (\vec{x}_{p-1}^T(N) \vec{a}_{p,N} + \vec{x}(N+1)) = \\ \quad \vec{i} = \vec{a}_{p,N} - \vec{e}_{p,N}^f(N+1) P_N \vec{x}'_{p-1}(N) = \\ \quad \vec{i} = \vec{a}_{p,N} - \vec{e}_{p,N}^f(N+1) \vec{c}_{p-1,N}; \end{cases} \quad (12)$$

$$\vec{c}_{p-1,N} = P_{N-1} \vec{x}'_{p-1}(N) / (w + \vec{x}_{p-1}^T(N) P_{N-1} \vec{x}'_{p-1}(N));$$

$$P_N = w^{-1} (I - \vec{c}_{p-1,N} \vec{x}_{p-1}^T(N)) P_{N-1},$$

где  $\vec{e}_{p,N}^f(N+1) = \vec{x}^T \vec{a}_{p,N} + \vec{x}(N+1)$  - вектор остаточных ошибок фильтрации, т.к. в отличие от ошибки предсказания здесь используется вектор  $\vec{a}_{p,N}$ , а не  $\vec{a}_{p,N+1}$ ;

$\vec{c}_{p-1,N} = P_N \vec{x}'_{p-1}(N)$  - вектор коэффициентов усиления остаточной ошибки фильтрации ("невязки");

$w$  - взвешивающий множитель, принимающий значение в пределах  $[0, 1]$ ;

$P_{N-1} = R_{p-1,N-1}^{-1}$  - матрица дисперсий ошибок фильтрации;

$R_{p-1,N-1} = \sum_{n=1}^{N-1} \omega^{N-1-n} \vec{x}'_{p-1}(n) \vec{x}_{p-1}^T(n)$  - матрица размерностью  $(p-1 \times p-1)$  взвешенных с весом  $0 < \omega < 1$  вторых моментов процесса на шагах  $p-1$ , усредняемых по выборке объемом  $N-1$ .

Исходные данные для работы фильтра задаются в виде  $\vec{a}_0$  и  $P_{p,0} = \varepsilon I$ , где  $I$  - единичная матрица, а  $\varepsilon$  некоторая положительная величина, обеспечивающая обратимость матрицы  $P_{p,0}$ . Структура алгоритма аналогична структуре известного алгоритма фильтрации Калмана [3].

Существенное уменьшение вычислительных затрат с  $p^2$  операций до  $5p$  может быть достигнуто при использовании быстрых РНК-алгоритмов вычисления

$\vec{a}_{p,N}$ . Ключевым моментом ускорения при этом является введение процедуры обновления значений вектора коэффициентов усиления  $\vec{c}_{p-1,N}$  с использованием лишь векторных операций вместо векторно-матричных [1-3].

### *Методы прогноза временных рядов*

Экспоненциальное сглаживание – как метод прогнозирования многих временных рядов был независимо открыт Броуном и Холтом, а Gardiner (1985) предложил "единую" классификацию методов экспоненциального сглаживания [19,20].

#### *Простое экспоненциальное сглаживание*

Простая модель временного ряда имеет следующий вид:

$$X_t = b + \varepsilon_t, \quad (13)$$

где  $b$  - константа и  $\varepsilon$ (эпсилон) - случайная ошибка. Константа  $b$  относительно стабильна на каждом временном интервале, но может также медленно изменяться со временем. Один из интуитивно ясных способов выделения  $b$  состоит в том, чтобы использовать сглаживание скользящим средним, в котором последним наблюдениям приписываются большие веса, чем предпоследним, предпоследним большие веса, чем пред-предпоследним и т.д. Простое экспоненциальное сглаживание именно так и устроено. Здесь более старым наблюдениям приписываются экспоненциально убывающие веса, при этом, в отличие от скользящего среднего, учитываются *все*



предшествующие наблюдения ряда, а не те, что попали в определенное окно. Формула простого экспоненциального сглаживания имеет следующий вид:

$$S_t = \alpha * X_t + (1 - \alpha) * S_{t-1}. \quad (14)$$

Когда эта формула применяется рекурсивно, то каждое новое сглаженное значение (которое является также прогнозом) вычисляется как взвешенное среднее текущего наблюдения и сглаженного ряда. Очевидно, результат сглаживания зависит от параметра  $\alpha$  (*альфа*). Если  $\alpha$  равно 1, то предыдущие наблюдения полностью игнорируются. Если  $\alpha$  равно 0, то игнорируются текущие наблюдения. Значения  $\alpha$  между 0, 1 дают промежуточные результаты.

Эмпирические исследования [1] показали, что простое экспоненциальное сглаживание дает достаточно точный прогноз.

На практике параметр сглаживания часто ищется с *поиском на сетке*. Возможные значения параметра разбиваются сеткой с определенным шагом. Например, рассматривается сетка значений от  $\alpha = 0.1$  до  $\alpha = 0.9$ , с шагом 0.1. Затем выбирается  $\alpha$ , для которого сумма квадратов (или средних квадратов) остатков (наблюдаемые значения минус прогнозы на шаг вперед) является минимальной. Рекомендуется брать начальное значение  $S_0$ , дающее наилучший прогноз. С другой стороны, влияние выбора уменьшается с длиной ряда и становится не критичным при большом числе наблюдений.

### *Общая модель.*

Основная идея сезонной декомпозиции проста. В общем случае временной ряд типа того, который описан выше, можно представить себе состоящим из четырех различных компонентов: (1) сезонного компонента (обозначается  $S_t$ , где  $t$  обозначает момент времени), (2) тренда ( $T_t$ ), (3) циклического компонента ( $C_t$ ) и (4) случайного (флуктуационного) компонента ( $I_t$ ). Разница между циклическим и сезонным компонентами состоит в том, что последний имеет регулярную (сезонную) периодичность, тогда как циклический фактор имеет более длительный эффект, который к тому же меняется от цикла к циклу. Обычно тренд и циклический компонент объединяют в один *тренд-циклический компонент* ( $TC_t$ ). Конкретные функциональные взаимосвязи между этими компонентами могут иметь самый разный вид. Для случая аддитивной взаимосвязи имеем:

$$X_t = TC_t + S_t + I_t \quad (15)$$

Здесь  $X_t$  обозначает значение временного ряда в момент времени  $t$ .

Качество прогноза определяется временем упреждения и точностью. Время упреждения, в свою очередь, определяется временем запаздывания в принятии решений по изменению плана распределения ресурсов ретранслятора, а точность прогноза определяется вероятностью достижения заданной погрешности прогнозирования.

## ***Марковские процессы и модели.***

Случайный процесс называется **марковским** , если вероятность любого состояния в будущем зависит только от его состояния в настоящем и не зависит от того , когда и каким образом процесс оказался в этом состоянии.

Описывающий поведение системы процесс называется цепью Маркова.

Для того чтобы случайный процесс с непрерывным временем был марковским , необходимо , чтобы интервалы времени между соседними переходами из состояния в состояние были распределены по экспоненциальному закону .

Использование марковских моделей позволяет существенно снизить размерность математического описания процессов при сохранении сведений о вероятностно-временном механизме изменения их состояния.

Рассмотрим два основных класса марковских процессов и соответствующих им моделей: непрерывные по состоянию случайные последовательности на базе стохастических разностных уравнений (СРУ) и дискретные по состоянию последовательности (цепи), описываемые также СРУ.

### ***Марковские модели непрерывных процессов.***

Пусть экспоненциально коррелированный исследуемый управляемый процесс описывается вектором фазовых координат  $x(k+1)$ , подчиняющийся нормальному закону распределения:

$$p(x) = \frac{1}{[(2\pi)^I |D_x|]^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\bar{x})^T D_x^{-1}(x-\bar{x})};$$

где  $\bar{x}$  - вектор математических ожиданий (средних ) процесса  $\bar{x}(k)$ ;

$|D_x| = |\sigma_{ij}|, i=1, I$  - определитель матрицы дисперсии процесса  $\bar{x}(k)$  .

Пусть также существует некоторый возбуждающий процесс  $\vec{v}(k)$ , являющийся независимым по отношению к  $\vec{x}(k)$  и подчиняющийся тоже нормальному распределению.

Тогда теоремы Дж.Дуба определяют возможность представления векторной гауссовской (марковской) последовательности с экспоненциальной функцией корреляции СРУ следующего вида:

$$\vec{x}(k+1) = A(k+1, k)\vec{x}(k) + B(k)\vec{u}(k) + G(k)v(k), (2.)$$

где  $A(k+1, k) = \{\alpha_{ij}(k)\}$  - диагональная матрица состояния размерностью  $I \times J$ , ненулевые элементы которой  $\alpha_{ii}(k)$  представляют собой скорости изменений (иначе спектры флуктуаций) значений процессов  $x_i(k)$ ;

$B(k)$  - матрица эффективности управления  $\vec{u}(k)$ ;

$G(k)$  - матрица диффузии процесса, элементы которой определяются параметрами

шума возбуждения  $g_{ii} = \alpha_i T \left( \frac{\sigma_i^2}{\alpha_i V_{ii}} \right)^{\frac{1}{2}}$ ;

$\sigma_i^2$  - дисперсия  $i$ -го процесса  $x_i(k)$ ;

$\alpha_i = \frac{1}{\tau_{кор_i}}$  - скорость изменения процесса  $x_i(k)$ , величина обратно пропорциональная

его интервалу корреляции;

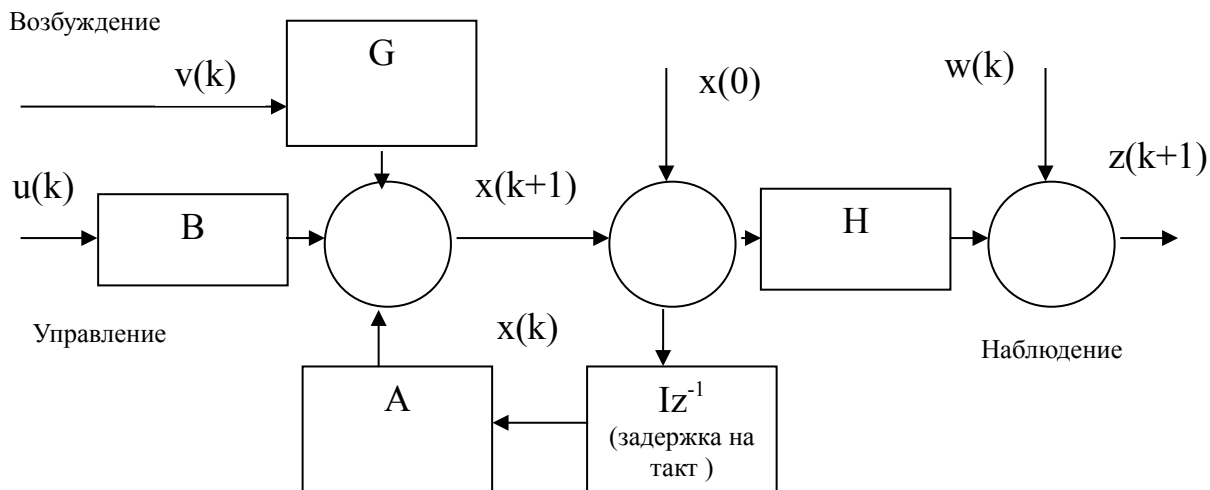
$T$  - период дискретизации по времени процесса;

$v(k)$  - вектор белых возбуждающих последовательностей с нулевым средним и ковариационной матрицей  $M [\vec{v}(k)\vec{v}(s)] = V(k)\delta(k-s)$  ;

$V(k)$  - матрица спектральных плотностей мощности процессов  $v(k)$ , диагональные члены которой характеризуют нормированные дисперсии  $V_{ii}$  возбуждающих  $i$ -х последовательностей;

$\delta(k-s)$  - символ Кронекера - индикатор равенства элементов, формально: функция двух целых переменных, которая равна 1, если они равны, и 0 в противном случае.

Рисунок. 2. Структурная схема формирующего фильтра для марковского процесса



## **Марковские модели дискретных последовательностей.**

Простейшей моделью, описывающей дискретные по состояниям и времени процессы, является цепь Маркова. Пусть дискретному множеству состояний стохастического объекта управления  $x(k) \in X = \{1, \dots, i, \dots, I\}$  в дискретные моменты времени  $k \in T = \{1, \dots, k, \dots, K\}$  соответствует конечное множество решений (управлений)  $u_i(k)$ ,  $x \in U = \{1, \dots, s, \dots, S\}$ . Простейшей моделью, описывающей вероятностно-временной механизм изменения дискретных по состояниям и времени управляемых случайных последовательностей, является простая однородная управляемая цепь Маркова. Цепь задается вектором вероятностей начальных состояний процесса  $\mathbf{P}(0) = \{p_i(0)\}$ , матрицей одношаговых переходных вероятностей  $\mathbf{P}^u(k/k-1) = \{p_{ij}^u(k/k-1)\}$ , а также периодом смены состояния марковской цепи  $(t_k - t_{k-1} = T)$ . Под простой цепью понимается односвязная цепь, удовлетворяющая марковскому свойству, а однородность цепи связана с постоянством значений элементов матрицы одношаговых переходных вероятностей.

Для однородной управляемой марковской цепи вероятность принятия ею  $i$ -го состояния на  $k$ -м шаге определяется следующей рекуррентным уравнением Маркова:

$$P_i(k) = \sum_{j=1}^I p_j(0) p_{ij}^u(k/k-1), \quad i, j = 1, \dots, I, \quad u \in U = \{1, \dots, s, \dots, S\}, \quad (3)$$

где

$$P^{\bar{u}}(k/k-1) = \{p_{ij}^{\bar{u}}(k/k-1)\} = \left\{ \sum_{m=1}^I p_{im}^{(k-n)} p_{mj}^{(n)} \right\} = [P^{\bar{u}}(1/0)]^k, \quad 0 < n < k$$

-матрица одношаговых переходных вероятностей на  $k$ -й шаг, определяемая из уравнения Колмогорова-Чепмена;

$$\sum_{i=1}^I p_{ij} = 1, \sum_i p_i = 1, p_i \geq 0, p_{ij} \geq 0 \text{ - условие нормировки и ограничения.}$$

Здесь вектор начальных вероятностей определяет значения вероятностей принятия процессом  $x(0)$  состояния  $i$  в нулевой момент времени, при этом если процесс в момент  $k-1$  принял состояние  $x(k-1)=i$ , то одношаговая вероятность перехода  $p_{ij}(k/k-1)$  определяет вероятность принятия процессом на следующем шаге  $k$  состояния  $j$ .

Для агрегированной марковской цепи элементы новой укрупненной матрицы одношаговых переходных вероятностей определяются соотношением:

$$p^{(n)}_{i'j'} = p_{ij} / \sum_i p_{ij}, i' = \overline{1, I'}, I' \subset I. \text{ Время смены состояний } T^{(n)} \text{ новой укрупненной цепи}$$

также изменяется и становится большим по сравнению с  $T$ .

Свойство эргодичности определяется наличием в цепи одного класса сообщающихся состояний. Следовательно, при  $k$ , стремящемся к  $\infty$ -ти эргодическая цепь становится стационарной, а в остальное время она находится в переходном состоянии. Для эргодической цепи также может быть определена корреляционная функция

$$K_x(kT) = \sum_{i=1}^I \sum_{j=1}^I x_i x_j p(x_i, x_j, kT) - \left( \sum_{i=1}^I x_i p_i \right)^2,$$

где  $p(x_i, x_j, kT) = p_i p_{ij}(kT)$  - совместная вероятность принятия состояний процессом  $x$  в моменты времени, отстоящие на  $kT$ ;

$p_i$ - вероятность принятия процессом значения  $i$  в начальный момент времени.

Основным недостатком представленного описания является невозможность представления с его помощью выборочных значений (реализации) случайного процесса  $x(k)$ ,  $k = 0, 1, 2, \dots, K$ . Поэтому в работах [2,3] введены специальные индикаторы состояния моделируемых последовательностей:

$$\theta_m(k) = x_m \quad (4)$$

В этом случае выборочные значения процесса  $x(k)$  определяются соотношением следующего вида:

$$x(k) = C^T(k) \theta(k), \quad (5)$$

где  $C^T(k) = [x_m]$  - строка возможных состояний процесса  $x(k)$ ;

$m = 1, \dots, M$  - номер состояния дискретного процесса, для которого выполняется условие  $\{x_1 < x_2 < \dots < x_m < \dots < x_M\}$ .

Наконец, для случая однородной марковской цепи можно записать следующее уравнение состояния:

$$\begin{aligned} \vec{\theta}(k+1) &= F(k+1, k, \vec{u}) \vec{\theta}(k) + \Delta \vec{\theta}(k), \\ \Delta \vec{\theta}(k) &= \Phi(G(k) \vec{v}(k), \vec{\theta}(k), \Lambda), \end{aligned} \quad (6)$$

$$x(k+1) = C^T(k+1) \theta(k+1),$$



где  $F(k+1, k) = [I + T Q^T(k+1, k, \vec{u})]$  – матрица вероятностей перехода процесса из одного состояния в другое на соседних шагах;

$Q(k+1, k, \vec{u}) = [q_{ml}]$  – матрица интенсивностей перехода процесса  $\vec{x}(k)$ ; ;

$q_{mm}(\vec{u}) = -\sum_{m \neq l} q_{ml}(\vec{u})$  – диагональные члены матрицы интенсивностей;

$T_c$  – период смены состояний цепи;

$I$  – единичная матрица соответствующей размерности;

$\Phi(\cdot)$  – функция формирования компенсирующей последовательности на основе случайного выбора на каждом шаге из исходной матрицы добавок  $\Lambda$  ту, которая обеспечит заданные статистические характеристики моделируемого процесса;

$G(k) = \text{diag} \left\{ T_c \sqrt{\frac{2 \sigma_\theta^2 q_{mm}}{V_m}} \right\}$  – матрица диффузии процесса  $x(k)$ ;

$V_m$  – спектральная плотность мощности возбуждающего шума;

$\Delta \theta(k)$  – вектор последовательностей, компенсирующий дробные значения первого слагаемого в выражении (6.);

Таким образом, дискретное состояние объекта управления в любой  $k$ -й момент времени, при марковской природе происходящих в нем процессов, определяется состоянием объекта на предыдущем шаге, матрицей вероятностей перехода и диффузионными свойствами шума возбуждения (случайным значением дискретного приращения, компенсирующего нецелочисленную часть прогноза состояния).

Порядок определения вектора компенсирующих последовательностей  $\vec{v}(k)$  в выражении (6.) подробно рассмотрен в работе[3].

Наряду с уравнениями состояния (2.,6) полная математическая модель случайного процесса обычно содержит уравнение наблюдения за его состоянием следующего вида:

$$\vec{z}(k) = H(k)\vec{x}(k) + \vec{w}(k), \quad (7)$$

где

$\vec{z}(k)$  – процесс наблюдения за процессом  $x(k)$ ;

$H(k)$  - матрица наблюдения, элементы которой определяют способ преобразования измерений в значения процесса  $x(k)$ ;

$\vec{w}(k)$  – вектор непрерывных гауссовских последовательностей со средним, равным нулю и ковариационной матрицей  $M \{ w(k)w(s) \} = R(k)\delta(k - s)$ ;

$R(k)$  – матрица спектральных плотностей мощности процесса  $\vec{w}(k)$ .

Примеры моделей СМО с марковскими моделями подробно рассматривались на практических занятиях.