ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.О. 10 Современные методы оптимизации

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность программы: Автоматизированные системы обработки информации и управления

Квалификация (степень) выпускника: магистр

Нормативный срок обучения 2 года
Форма обучения очная

СОДЕРЖАНИЕ

.

1.1 Лекция 1 (Л-1) (2 ч.)

Тема: Общие сведения о математическом моделировании

1.2.1. Вопросы лекции:

1.Моделирование как метод научного познания.

2. Адекватность и эффективность моделей.

3. Моделирование и подобие в научно-технических исследованиях.

1.2.2 Краткое содержание вопросов:

1.Моделирование как метод научного познания.

В настоящее время нельзя назвать область человеческой деятельности, в которой в той или иной степени не использовались бы методы моделирования. Однако моделирование как специфическое средство и форма научного познания не является изобретением XX века. Достаточно указать на представления Демокрита и Эпикура об атомах, их форме, и способах соединения, об атомных вихрях и ливнях, объяснения физических свойств различных веществ с помощью представления о круглых и гладких или крючковатых частицах, сцепленных между со- бой. Эти представления являются прообразами современных моделей, отражающих ядерно-электронное строение атома вещества.

В настоящее время с моделированием, в тех или иных его проявлениях, явно или неявно связаны и научные исследования в естественных, технических и гуманитарных областях знания, и работы инженеров по проектированию, конструированию, экспериментальным исследованиям и испытаниям новых технических систем и деятельность руководителя по принятию инженерных и управленческих решений. Методологическая основа моделирования заключается в следующем. Все то, на что направлена человеческая деятельность, называется объектом (лат. objectum – предмет). Выра ботка методологии направлена на упорядочение получения и обработки информации об объектах, которые существуют вне нашего сознания и взаимодействуют между собой и внешней средой. Объект – некоторая часть окружающего мира, рассматриваемого человеком как единое целое. Каждый объект имеет имя и обладает параметрами. Параметр – признак или величина, характеризующая какое-либо свойство объекта и принимающая различные значения. В научных исследованиях большую роль играют гипотезы, то есть определенные предсказания, основывающиеся на небольшом количестве опытных данных, наблюдений, догадок. Быстрая и полная проверка гипотез может быть проведена в ходе специально поставленного эксперимента. При формулировании и проверки правильности гипотез большое значение в качестве метода суждений имеет аналогия. Аналогией называют суждение о каком либо частном сходстве двух объектов, причем такое сходство может быть существенным и несущественным. Необходимо отметить, что понятия существенности и несущественности сходства или различия объектов условны и относительны. Существенность сходства (или различия) зависит от уровня абстрагирования и в общем случае определяется конечной целью про- водимого исследования. Современная научная гипотеза создается, как правило, по аналогии с проверенными на практике научными положениями. Таким образом, аналогия связывает гипотезу с экспериментом. Гипотезы и аналогии, отражающие реальный, объективно существующий мир, должны обладать наглядностью или сводится к удобным для исследования логическим схемам. Такие логические схемы, упрощающие рассуждения и логи- ческие построения или позволяющие проводить эксперименты, уточняющие природу явлений, называются моделями. Другими словами модель (лат. modulus - мера) – это объект заместитель объекта-оригинала, обеспечивающий изучение некоторых свойств оригинала. Моделированием называется замещение одного объекта другим с целью получения информации о важнейших свойствах объекта-оригинала с помощью объекта-модели. Таким образом, моделирование может быть определено как представление объекта моделью для получения информации об этом объекте путем проведения экспериментов с его моделью.

Теория замещения одних объектов (оригиналов) другими объектами (моделями) и исследование свойств объектов на их моделях называется теорией моделирования.

2. Адекватность и эффективность моделей.

Если результаты моделирования подтверждаются и могут служить основой для прогнозирования процессов, протекающих в исследуемых объектах, то говорят, что модель адекватна объекту. При этом адекватность модели зависит от 8 цели моделирования и принятых критериев. Каждому материальному объекту соответствует бесчисленное множество в равной мере адекватных, но различных по существу моде- лей, связанных с разными задачами. Так, один и тот же технический объект, например, электрический двигатель, при расчете электрических характеристик, тепловом расчете и оценке надежности представляется тремя различными моделями. При построении моделей и выявлении их места в научном исследовании необходимо, прежде всего, установить цель исследования, сформулировать допущения и выделить из бесконечного множества подсистем и процессов, составляющих систему и происходящие в ней явления, те, которые под- лежат изучению. При создании моделей нельзя охватить все многообразие процессов, составляющих явление: необходимо выделить их него те процессы, которые интересуют ис- следователя в данной постановке задачи.

Абсолютно точная модель, так же как и абсолютное подобие является математической абстракцией. Это означает, что модель и отображаемый ею объект находятся в отношении сходства, а не тождества. Когда модель становиться «слишком точной», она теряет смысл, перестает быть моделью; когда же она несовершенна – она источник ошибки. Точность измерений и вычислений определяет точность проводимых исследований. С наибольшей точностью должны быть исследованы существенные факторы. Слишком точные измерения и вычисления несущественных факторов могут привести к тому, что действительно существенные в данном явлении закономерности выпадут из рассмотрения. Распространенной ошибкой является недостаточное внимание к качеству исходных данных. Самый точный вычислительный метод может быть в значительной мере обесценен, если воспользоваться не вполне верными исходными данными или данными, имеющими низкую точность. Важно, чтобы выбираемый метод решения задачи был рассчитан на использование таких данных, которые можно получить с требуемой достоверностью.

Определить при построении модели математическим путем наилучшее сочетание полноты-точности создаваемой модели, с одной стороны, и простоты, с другой, практически никогда не удается из-за неформализуемости и неоднозначности большей части подлежащих учету при создании моделей факторов. Из вышесказанного следует, что на эффективность и адекватность модели влияют следующие факторы. Первая, практически независящая от исследователя, группа факторов связана с существом и характером решаемой задачи, а также со степенью сложности, изученностью, внутренними свойствами и особенностями объекта моделирования. Пара задача-объект в основном определяет список подлежащих учету переменных объекта, параметры, входящие в модель, число и характер связей между ними, объем необходимой исходной информации, требования к точности данных и ряд других важнейших характеристик модели. В случае математического моделирования решающим фактором эффективности модели оказывается математический аппарат.

Эффективность модели зависит и от такого субъективного фактора, как профессиональные качества и уровень подготовки исследователя-исполнителя. Он должен досконально представлять себе задачу и изучить объект моделирования; владеть аппаратом современной прикладной математики, использовать при решении задачи вычислительные методы и алгоритмы; обладать творческими способностями (воображением, изобретательностью, интуицией

3. Моделирование и подобие в научно-технических исследованиях.

Модель, таким образом, - это естественный или искусственный объект, находящийся в соответствии и изучаемым объектом или, точнее, с какой либо из его сторон. В общетеоретическом смысле моделирование означает осуществление каким-либо способом отображения или воспроизведения действительности для изучения имеющихся в ней объективных закономерностей. Обобщенно моделирование определяется как метод опосредованного познания, при котором изучаемый объект (оригинал) находится в некотором соответствии с другим объектом (моделью), причем объект-модель способен в том или ином отношении замещать оригинал на некоторых стадиях познавательного процесса. Стадии познания, на которых может происходить эта замена, равно как и формы соответствия модели и оригинала, могут быть различны. Исходя из понятия отображения, принятого в теории познания, можно определить два характерных вида моделирования: - моделирование как познавательный процесс, содержащий переработку информации, поступающей из внешнего мира, о происходящих в нем явлениях. В результате этой ин- формации в сознании появляются образы, имеющие определенной сходство с соответствующими объектами. Сумма этих образов позволяет выявлять свойства изучаемых объектов и их взаимодействие. Математическая запись, составленная на основании суммы образов и содержащая описание динамики физических или других (например, экономических) закономерностей, и есть модель; 17 - моделирование как создание некоторой системы – системы-модели, имеющей определенное сходство с системой оригиналом. Две эти материально реализованные системы связаны соотношениями подобия. Отображение одной системы в другой – следствие выявления сложных зависимостей между двумя системами, отраженных в соотношении подобия, а не результат непосредственного изучения поступающей информации. При первом виде моделирование имеет мысленный характер, при втором – материальный, поскольку его реализация требует создания специальных установок, воспроизводящих исследуемую систему. Многообразие исследуемых объектов и процессов, целей и задач моделирования породило множество типов моделей. Выбор аппарата для построения модели зависит, как от природы и свойств моделируемого объекта или процесса, так и от характера решаемой задачи.

Классификация моделей.

1) Классификация моделей по области использования: Учебные модели – используются при обучении. Опытные – это уменьшенные или увеличенные копии проектируемого объекта, используют для исследования и прогнозирования его будущих характеристик, Научно - технические - создаются для исследования процессов и явлений. Игровые – ре- петиция поведения объекта в различных условиях. Имитационные – имитируют поведение объекта в различных ситуациях.

2) Классификация моделей по фактору времени: Статические – модели, описывающие состояние системы в определенный момент времени (единовременный срез ин- формации по данному объекту). Примеры таких моделей: классификация животных, строение молекул, список посаженных деревьев и т.д. 18 Динамические – модели, описывающие процессы изменения и развития системы (изменения объекта во времени). Примеры: описание движения тел, развития организмов, процесс химических реакций.

3) Классификация моделей по отрасли знаний - это классификация по отрасли деятельности человека: математические, биологические, химические, социальные, экономические, исторические и т.д.

4)Классификация по степени определенности: Модель называется детерминированной, если каждому набору входных параметров всегда соответствует единственный набор выходных параметров. В противном случае мо- дель называется недетерминированной (стохастической, вероятностной).

5) Классификация моделей по форме представления:

Материальные – это предметные (физические) мо- дели. Они всегда имеют реальное воплощение. Материальные модели можно разделить на модели копии (например, глобус, карта, некоторые виды наглядных пособий и т.п.) и физические модели. Физическое моделирование – это вос- произведение с помощью модели основных геометрических, физических и функциональных характеристик изучаемого объекта. Определение параметров модели при физическом моделировании и перенос данных моделирования на реальный объект осуществляется на основании теории подобия. Теория подобия позволяет связать параметры натурального объекта и модели с помощью критериев подобия, а на их основе подобрать удобные для моделирования параметры модели. Физическая модель – это установка или устройство, позволяющее проводить исследования путем замены изучаемого физического процесса подобным ему процессом с сохранением его физической природы. Физическое моделирование может протекать в реальном или нереальном масштабе времени, а также может рассматриваться без учета времени.

Абстрактные (нематериальные) модели – не имеют реального воплощения. Их основу составляет информация, это теоретический метод познания окружающей среды. В абстрактных моделях описание объектов и процессов осуществляется на каком-либо языке. В качестве языков моделирования можно использовать естественный язык, язык чертежей, схем, математический язык и т.п. По признаку реализации абстрактные модели бывают: мысленные, вербальные; информационные. Мысленные мо- дели формируются в воображении человека в результате раздумий, умозаключений, иногда в виде некоторого образа. Это модель сопутствует сознательной деятельности чело- века. Вербальные – мысленные модели, выраженные в разговорной форме. Информационные модели – целенаправленно отобранная информация об объекте, которая отражает наиболее существенные для исследователя свойств этого объекта.

Типы информационных моделей:

1) По форме представления: Табличные – объекты и их свойства представлены в виде списка, а их значения размещаются в ячейках прямоугольной формы. Перечень однотипных объектов размещен в первом столбце (или строке), а значения их свойств размещаются в следующих столбцах (или строках). Иерархические – объекты распределены по уровням. Каждый элемент высокого уровня состоит из элементов ниж- него уровня, а элемент нижнего уровня может входить в со- став только одного элемента более высокого уровня. Сетевые – применяют для отражения систем, в которых связи между элементами имеют сложную структуру.

2) По степени формализации: Образные модели выражают свойства оригинала с помощью наглядных чувственных образов, имеющих прообразы среди элементов оригинала или объектов материального мира. Например, в кинетической теории газов частицы газа образно моделируются в виде упругих шаров, воздействующих друг на друга только во время столкновений. Знаковые модели выражают свойства оригинала с помощью условных знаков или символов. К ним относят математические выражения и уравнения, физические и химические формулы и т. п. Образно-знаковые модели обладают признаками образных и знаковых. Образно-знаковые модели можно разделить на геометрические (рисунок, пиктограмма, чертеж, карта, план, объемное изображение), структурные (таблица, граф, схема, диаграмма), словесные (описание естественными язы ками), алгоритмические (пошаговое перечисление, блок- схема).

1.2 Лекция 2-3 (Л-2-3) (4 ч.)

Тема: Общая модель линейного программирования в каноническом и неканоническом представлении. Графический метод решения ЗЛП.

1.2.1. Вопросы лекции:

1. Общая модель линейного программирования в каноническом и неканоническом представлении.

2. Графический метод решения ЗЛП.

1.2.2 Краткое содержание вопросов:

1. Общая модель линейного программирования в каноническом и неканоническом представлении.

Существует огромное количество ЗЛП с различными словесными формулировками. Но, несмотря на различие содержательных ситуаций ряда задач ЛП, экстремальные математические модели, соответствующие им, имеют много общего. Так, в каждой из этих задач требуется максимизировать или минимизировать линейную функцию от нескольких переменных. При этом ограничения, наложенные на совокупность переменных, являются либо линейными уравнениями, либо линейными неравенствами, либо состоят как из линейных уравнений, так и из линейных неравенств. Каждая из этих задач является частным случаем общей задачи ЛП.

Общей задачей ЛП называется задача, которая состоит в определении максимального (минимального) значения функции:

$$F=\sum_{j=1}^{n} c_j x_j \quad (.1)$$

при условиях:

$$\sum_{j=1}^{n} a_{ij}\, x_j \le b_i, \quad i=\overline{1,k} \quad (.2)$$

$$\sum_{j=1}^{n} a_{ij}\, x_j = b_i, \quad i=\overline{k+1,m} \quad (.3)$$

$$x_j \ge 0, \quad j=\overline{1,l}, \quad l \le n \quad (.4)$$

где a_{ij} , b_i, c_j - заданные постоянные величины и k £ m.

Функция (.1) называется ***целевой*** функцией задачи (1) – (4), а условия (2) – (4) – ***ограничениями*** данной задачи.

Различают еще две основные формы задач ЛП в зависимости от наличия ограничений разного типа: стандартную и каноническую.

Стандартной (или симметричной) ЗЛП называется задача, которая состоит в определении максимального значения функции (1) при выполнении условий (2) и (4), где $k = m$, и $l=n$, т.е.

$$F=\sum_{j=1}^{n} c_j x_j \to max$$

$$\sum_{j=1}^{n} a_{ij}\, x_j \le b_i, \quad i=\overline{1,m} \quad (5)$$

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

Канонической (или основной) ЗЛП называется задача, которая состоит в определении максимального значения целевой функции (1) при выполнении условий (3) и (4), где $k=0$, и$l=n$, т.е.

$$F=\sum_{j=1}^{n} c_j x_j \to max$$

$$\sum_{j=1}^{n} a_{ij}\, x_j = b_i, \quad i = \overline{1,m}$$
$$x_j \ge 0, \quad j = \overline{1,n} \qquad (3.6)$$

Стандартная форма модели интересна тем, что большое число прикладных моделей естественным образом сводится к этому виду моделей. Каноническая форма модели важна ввиду того, что основные вычислительные схемы различных алгоритмов решения этих задач разработаны именно для этой формы.

Указанные выше три формы ЗЛП эквивалентны в том смысле, что каждая из них с помощью несложных преобразований может быть приведена к любой из двух остальных. Следовательно, любую ЗЛП можно привести к канонической форме. Поэтому умение решать задачу в канонической форме позволяет решать задачу и в любой другой форме.

Чтобы перейти от одной формы записи ЗЛП к другой нужно уметь:

1) сводить задачу минимизации функции к задаче максимизации;

2) переходить от ограничений – неравенств к ограничениям – равенствам;

3) заменять переменные, которые не подчинены условию неотрицательности, на неотрицательные переменные.

1. В том случае, когда требуется найти *min* функции $F=c_1x_1+c_2x_2+\ldots+c_nx_n$, можно перейти к нахождению максимума функции F_1, умножив коэффициенты при переменных в целевой функции модели на (-1), т.е.:

$F_1 = - F = - c_1x_1 - c_2x_2 - \ldots\ldots - c_nx_n$, т.к. $min\ F = - max\ (- F)$.

2. Ограничение – неравенство исходной задачи ЛП можно преобразовать в ограничение – равенство добавлением к его левой части неотрицательной переменной с соответствующим знаком.

Таким образом, ограничение – неравенство вида

$a_{i1}x_1 + a_{i2}x_2 + \ldots\ldots + a_{in}x_n \le b_i$

преобразуется в ограничение – равенство

$a_{i1}x_1 + a_{i2}x_2 + \ldots\ldots + a_{in}x_n + x_{n+i} = b_i, \quad x_{n+i} \ge 0, \quad i = \overline{1,m}$,

а ограничение – неравенство вида

$a_{i1}x_1 + a_{i2}x_2 + \ldots\ldots + a_{in}x_n \ge b_i$

в ограничение – равенство

$a_{i1}x_1 + a_{i2}x_2 + \ldots\ldots + a_{in}x_n - x_{n+i} = b_i, \quad x_{n+i} \ge 0, \quad i = \overline{1,m}$.

Число вводимых дополнительных неотрицательных переменных при подобных преобразованиях равно числу преобразуемых неравенств.

Вводимые дополнительные переменные имеют вполне определенный экономический смысл. Так, если в ограничениях исходной задачи ЛП отражается расход и наличие производственных ресурсов, то числовое значение дополнительной переменной следует интерпретировать как остаток, или неиспользованную часть, данного ресурса.

3. Если переменная $X_к$ не подчинена условию неотрицательности, то ее следует заменить двумя неотрицательными переменными X_k' и X_k'', приняв $X_k = X_k' - X_k''$. Правомерность такой замены очевидна, т.к. любое число можно представить в виде разности двух неотрицательных чисел.

2. Графический метод решения ЗЛП.

Графический метод довольно прост и нагляден для решения задач линейного программирования с двумя переменными. Он основан на геометрическом представлении допустимых решений и целевой функции задачи.

Каждое из неравенств задачи линейного программирования определяет на координатной плоскости (x_1Ox_2) некоторую полуплоскость, а система неравенств в целом – пересечение соответствующих плоскостей. Множество точек пересечения данных полуплоскостей называется областью допустимых решений (ОДР). ОДР всегда представляет собой выпуклую фигуру, т.е. обладающую следующим свойством: если две точки A и B принадлежат этой фигуре, то и весь отрезок AB принадлежит ей. ОДР графически может быть представлена выпуклым многоугольником, неограниченной выпуклой многоугольной областью, отрезком, лучом, одной точкой. В случае несовместности системы ограничений задачи ОДР является пустым множеством.

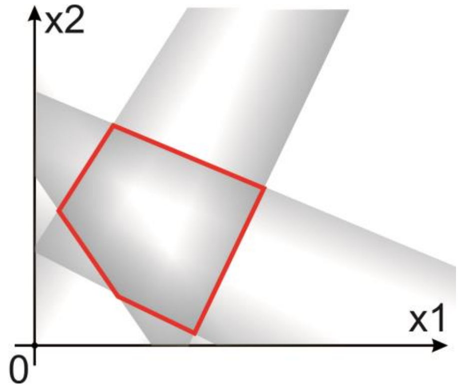
Все вышесказанное относится и к случаю, когда система ограничений включает равенства, поскольку любое равенство $a_{i1}x_1 + a_{i2}x_2 = b_i$ можно представить в виде системы двух неравенств

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 \le b_i \\ a_{i1}x_1 + a_{i2}x_2 \ge b_i \end{cases}$$

Целевая функция $L(x) = c_1x_1 + c_2x_2$ при фиксированном значении $L(x) = L$ определяет на плоскости прямую линию $c_1x_1 + c_2x_2 = L$. Изменяя значения L, мы получим семейство параллельных прямых, называемых линиями уровня.

Это связано с тем, что изменение значения L повлечет изменение лишь длины отрезка, отсекаемого линией уровня на оси Ox_2 (начальная ордината), а угловой коэффициент прямой $tg\alpha = \frac{-c_1}{c_2}$ останется постоянным. Поэтому для решения будет достаточно построить одну из линий уровня, произвольно выбрав значение L.

Вектор $C = (c_1, c_2)$ с координатами из коэффициентов целевой функции при x_1 и x_2 перпендикулярен к каждой из линий уровня. Направление вектора C совпадает с направлением возрастания целевой функции, что является важным моментом для решения задач. Направление убывания целевой функции противоположно направлению вектора .



Суть графического метода заключается в следующем. По направлению (против направления) вектора C в ОДР производится поиск оптимальной точки $x^* = (x_1^*, x_2^*)$. Оптимальной считается точка, через которую проходит линия уровня $L_{max}(L_{min})$, соответствующая наибольшему (наименьшему) значению функции $L(x)$. Оптимальное решение всегда находится на границе ОДР, например, в последней вершине многоугольника ОДР, через которую пройдет целевая прямая, или на всей его стороне.

Рисунок 1 Построение графиков, и выявление общей области решения

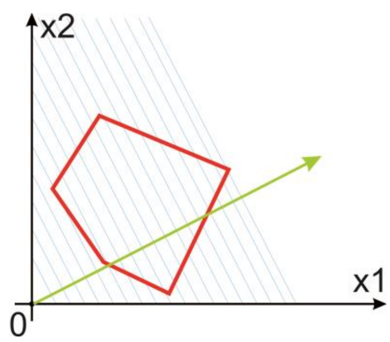


Рис. 2 Построение вектора градиента

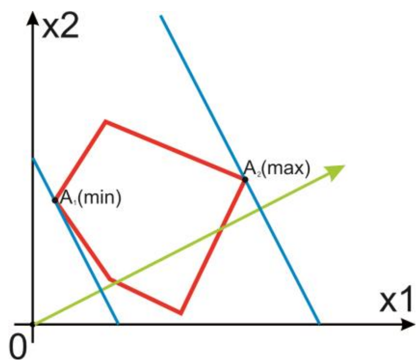


Рис. 3 Определение точек экстремума

Особые случаи

При поиске оптимального решения задач линейного программирования возможны могут возникнуть следующие ситуации, когда:

- Существует бесконечное множество решений (альтернативный оптиум);
- Целевая функция не ограничена;
- Отсутствует общая область допустимых решений.

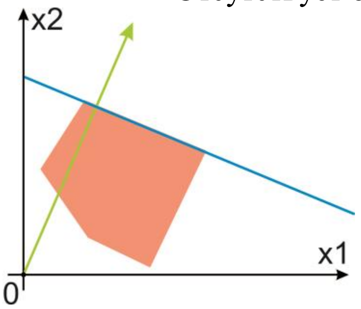


Рис. 4 Бесконечное множество решений (альтернативный оптимум)

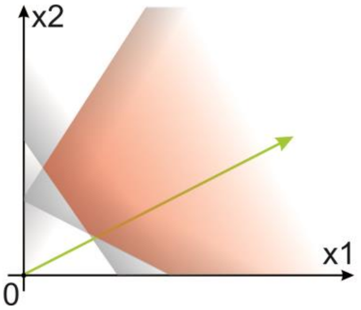


Рис.5 Целевая функция не ограничена

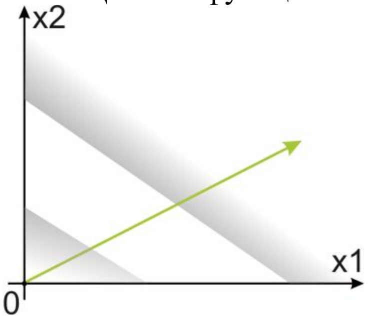


Рис. 6 Отсутствует общая область допустимых решений

Методика решения задач линейного программирования графическим методом.

1. В ограничениях задачи заменить знаки неравенств знаками точных равенств и построить соответствующие прямые.
2. Найти и заштриховать полуплоскости, разрешенные каждым из ограничений-неравенств задачи. Для этого нужно подставить в конкретное неравенство координаты какой-либо точки (например, (0;0)), и проверить истинность полученного неравенства. Если неравенство истинное, то надо заштриховать полуплоскость, содержащую данную точку. Иначе (неравенство ложное) - надо заштриховать полуплоскость, не содержащую данную точку.

Поскольку и должны быть неотрицательными, то их допустимые значения всегда будут находиться выше оси Ox_1и правее оси , т.е. вI-й координатной плоскости. Ограничения-равенства разрешают только те точки, которые лежат на соответствующей прямой. Поэтому необходимо выделить на графике такие прямые.

3. Определить ОДР как часть плоскости, принадлежащую одновременно всем разрешенным областям, и выделить ее. При отсутствии ОДР задача не имеет решений.
4. Если ОДР не пустое множество, то нужно построить целевую прямую, т.е. любую из линий уровня (где L– произвольное число, например, кратное c_1иc_2, т.е. удобное для проведения расчетов). Способ построения аналогичен построению прямых ограничений.

5. Построить вектор $C = (c_1, c_2)$, который начинается в точке (0;0) и заканчивается в точке (c_1, c_2). Если целевая прямая и вектор построены верно, то они будут перпендикулярны.

6. При поиске максимума ЦФ необходимо передвигать целевую прямую в направлении вектора, при поиске минимума ЦФ против направления вектора . Последняя по ходу движения вершина ОДР будет точкой максимума или минимума ЦФ. Если такой точки (точек) не существует, то можно сделать вывод о неограниченности ЦФ на множестве планов сверху (при поиске максимума) или снизу (при поиске минимум).

7. Определить координаты точки max(min) ЦФ и вычислить значение ЦФ $L(x^*)$. Для вычисления координат оптимальной точки X^* необходимо решить систему уравнений прямых, на пересечении которых находится.

1.3 Лекция 4-5 (Л-4-5) (4 ч.)

Тема: Симплекс-метод. Двойственные задачи линейного программирования.

1.3.1. Вопросы лекции:

1. Симплекс-метод.

2. Двойственные задачи линейного программирования.

1.3.2 Краткое содержание вопросов:

1. Симплекс-метод

Симплекс метод является наиболее распространенным вычислительным методом, который может быть применен для решения любых задач ЛП как вручную, так и с помощью ЭВМ.

Этот метод позволяет переходить от одного допустимого решения к другому, причем так, что значения целевой функции непрерывно возрастают. Этот переход возможен, если известен какой-либо опорный план. В этом случае каноническая задача линейного программирования должна содержать единичную подматрицу порядка *M*. Тогда очевиден первоначальный опорный план(неотрицательное базисное решение системы ограничений ЗЛП).

В результате оптимальное решение находят за конечное число шагов. Алгоритм симплекс-метода позволяет также установить является ли задача ЛП разрешимой.

Любая задача линейного программирования с помощью элементарных преобразований может быть приведена к каноническому виду. $F(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\text{mir}$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + õ_{n+1} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + õ_{n+2} = b_2 \\ \ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + õ_{n+ò} = b_m \\ x_i, õ\,(i = 1,n), \end{cases}$$

где F(x) - целевая функция; $x_1, x_2, \dots, x_n$ - базисные переменные; остальные переменные называются свободными.

Задача имеет *m+n* ограничений, среди них *m* ограничений типа равенства и *n* ограничений неотрицательности. По определению крайняя точка удовлетворяет *n* линейно-независимым ограничениям задачи как точным равенствам.

Алгоритм симплекс метода в случае неотрицательности свободных членов

1. Переходим от системы неравенств к равенствам с помощью ослабляющих переменных.
2. Находим единичные вектора условий, которые войдут в базис. Их должно быть столько, сколько уравнений в системе.
3. Составляем симплекс таблицу.

Таблица 1 Общий вид симплекс таблицы

Сб	Б	В	C_1	C_2	…	C_n	C_{n+1}	C_{n+2}	…	C_{n+m}	
			X_1	X_2	…	X_n	X_{n+1}	X_{n+2}	…	X_{n+m}	
	X_{n+1}	b_1	a_{11}	a_{12}	…	a_{1n}	1	0	…	0	
	X_{n+2}	b_2	a_{21}	a_{22}	…	a_{2n}	0	1	…	0	
…	…	…	…	…	…	…	…	…	…	…	
	X_{n+m}	b_m	a_{m1}	a_{m2}	…	a_{mn}	0	0	…	1	
Δ		Z									
L											

В первой строке записываем коэффициенты целевой функции.

Коэффициенты ограничений переносятся в среднюю часть таблицы.

Вносим свободные члены в третий столбец «В».

Во второй столбец записываем базис (единичные вектора).

Находим «Z» как сумму попарных произведений 1-го и 3-го столбцов.

Предпоследнюю строку таблицы L находим как Сумму попарных произведений элементов первого столбца и соответствующего столбца средней части таблицы.

Последнюю строку Δ определяем как разность элементов предпоследней строки и 1-й строки (коэффициентов целевой функции).

Величина «Δ» - критерий оптимальности. Для задачи минимизации оптимальным будет план, для которого все значения строки меньше или равны 0 ($\Delta \le 0$), а для задачи максимизации план в котором все значения больше или равны 0 ($\Delta \ge 0$).

Если план не оптимален, то переходим к заполнению новой симплексной таблице.

Находим разрешающий элемент.

Для задачи максимизации выбираем столбец с максимальным по модулю Δ среди отрицательных. А для задачи минимизации столбец с максимальным по модулю среди положительных. Такой столбец называется «Разрешающим»

Для нахождения «Разрешающей» строки, значения элементов 3-го столбца («B») делим на значения соответствующих элементов Разрешающего столбца. Полученный результат записываем в последний столбец (Q).
Среди полученных значений выбираем наименьшее. Строка, содержащая это число, и будет «Разрешающей».

На пересечение Разрешающего столбца и разрешающей строки получаем «Разрешающий» элемент.

В новой таблице меняем базисную переменную и соответственно поправляем 1-й столбец.

Разрешающую строку делим на Разрешающий элемент и записываем на свои места.

Обнуляем остальные элементы Разрешающего столбца.

Остальные элементы таблицы находим по правилу прямоугольника.

Заполняем последние 2 строки.
В случае оптимальности плана Записываем полученные значения в ответ, а в случае неоптимальности составляем новую симплекс таблицу согласно предыдущим пунктам.

Через конечное число итераций либо будет получено решение задачи линейного программирования, либо будет установлено, что решение неограниченно.

Метод искусственного базиса решения задач линейного программирования

Симплексный метод с искусственным базисом применяется в тех случаях, когда затруднительно найти первоначальный опорный план исходной задачи линейного программирования, записанной в канонической форме.

М-метод заключается в применении правил симплекс-метода к так называемой М-задаче. Она получается из исходной добавлением к левой части системы уравнений в канонической форме исходной задачи линейного программирования таких искусственных единичных векторов с соответствующими неотрицательными искусственными переменными, чтобы вновь полученная матрица содержала систему единичных линейно-независимых векторов. В линейную форму исходной задачи добавляется в случае её максимизации слагаемое, представляющее собой произведение числа (-М) на сумму искусственных переменных, где М - достаточно большое положительное число.

В полученной задаче первоначальный опорный план очевиден. При применении к этой задаче симплекс-метода оценки теперь будут зависеть от числа М. Для сравнения оценок нужно помнить, что М достаточно большое положительное число, поэтому из базиса будут выводиться в первую очередь искусственные переменные.

В процессе решения М-задачи искусственные векторы выходят из базиса, в результате чего будет получен опорный план. В противном случае исходная задача неразрешима.

Постановка задачи линейного программирования

При откорме, каждое животное должно получить не менее 9 ед. белка, 8 ед. протеина, и 10 ед. углеводов. Для составления рациона используется 2 вида корма, состав которых приведён в таблице. Стоимость 1 кг. Первого составляет 4 ден. ед., а второго 6 ден. ед..

Составьте дневной рацион питательности, имеющий минимальную стоимость.

Таблица 2. Состав питательных веществ в кормах

Питательное вещество	Содержание вещества	
	Корм 1	Корм 2
Белок	3	1
Углеводы	1	2
Протеин	1	6

Построение математической модели задачи

Практически во всех науках о природе, живой и неживой, об обществе, построение и использование моделей является мощным орудием познания. Реальные объекты и процессы бывают столь многообразны и сложны, что лучшим способом изучения часто является построение модели, отражающей лишь какую-то часть реальности.

В любом случае модель строится с целью узнать про объект что-либо новое или сохранить об объекте информацию, которая может стать недоступной в будущем.

Как правило, процесс изучения, связанный с использованием моделей и называемый моделированием не заканчивается созданием одной модели. Построив модель и получив с её помощью, какие-либо результаты, соотносят их с реальностью, и если это соотношение даёт неудовлетворительные результаты, то в построенную модель вносят коррективы или даже создают другую модель. В случае достижения хорошего соответствия с реальностью выясняют границы применения модели. Это очень важный вопрос, он решается путём сравнения модели с оригиналом путём сравнения предсказаний, полученных с помощью компьютерной модели. Если это сравнение даёт удовлетворительные результаты, то модель принимают на вооружение, если нет, приходится создавать другую модель.

Математическое моделирование относится к классу знакового моделирования, при этом модели могут создаваться из любых математических объектов, чисел, функций, уравнений, графиков, графов.

Практически во всех науках построение и использование моделей является мощным орудием познания.

Цель задачи заключается в нахождение рациона питания, который бы восполнял минимальный набор питательных веществ. При минимальных денежных затратах.

Пусть x_1 число килограммов корма 1-го вида, а x_2 число килограмм корма 2-го вида необходимых для составления рациона.

Цель задачи: Минимизировать расход, при соблюдение нормы рациона.

Следовательно, $Z(x) = 4 \cdot x_1 + 6 \cdot x_2 \to mir$

Состав питательных веществ в рационе не должен быть меньше необходимой дневной нормы. Отсюда следует система неравенств:

$$\begin{cases} 3 \cdot x_1 + 1 \cdot x_2 \geq 9 \\ 1 \cdot x_1 + 2 \cdot x_2 \geq 1C \\ 1 \cdot x_1 + 6 \cdot x_2 \geq 8 \end{cases}$$

2. Двойственные задачи линейного программирования.

Рассмотрим двойственные задачи в общей форме.

Прямая задача:

$$F = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \mathbf{max},$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1, \\ \ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \leq b_k, \\ a_{k+1,1}x_1 + a_{k+1,2}x_2 + \ldots + a_{k+1,n}x_n = b_{k+1}, \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases}$$

$$x_i \geq 0 \quad \left(j = \overline{1,l};\, l \leq n\right).$$

Двойственная задача:

$$F = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \mathbf{min}$$

$$
\begin{cases}
a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \ge c_1, \\
\ldots \\
a_{1i}y_1 + a_{2i}y_2 + \ldots + a_{mi}y_m \ge c_i, \\
a_{1,i+1}y_1 + a_{2,i+1}y_2 + \ldots + a_{m,i+1}y_m = c_{i+1}, \\
\ldots \\
a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m = c_m,
\end{cases}
$$

$$y_i \ge 0 \quad \left(j = \overline{1,k};\, k \le m\right).$$

Двойственная задача по отношению к прямой составляется следующим образом:

1. Целевая функция исходной задачи задаётся на максимум, а в двойственной – на минимум.

2. Матрицы коэффициентов прямой и двойственной задач получаются друг из друга заменой строк столбцами, а столбцов – строками (операция транспонирования):

$$
A = \begin{pmatrix} a_{11}\ a_{12} \ldots a_{1n} \\ a_{21}\ a_{22} \ldots a_{2n} \\ \ldots \\ a_{m1}\ a_{m2} \ldots a_{mn} \end{pmatrix}, \quad
A^T = \begin{pmatrix} a_{11}\ a_{21} \ldots a_{m1} \\ a_{12}\ a_{22} \ldots a_{m2} \\ \ldots \\ a_{1n}\ a_{2n} \ldots a_{mn} \end{pmatrix}.
$$

3. Число переменных в двойственной задаче равно числу ограничений в исходной задаче (и наоборот).

4. Коэффициентами при неизвестных в целевой функции двойственной задачи являются свободные члены в ограничениях исходной задачи, а правыми частями в ограничениях двойственной задачи являются коэффициенты при неизвестных в целевой функции исходной задачи.

Задача. Составить задачу, двойственную следующей задаче:

$$f = -x_1 + 2x_2 \to \max,$$

$$
\begin{aligned}
&2x_1 - x_2 \ge 1, \\
&-x_1 + 4x_2 \le 24, \\
&x_1 - x_2 \le 3, \\
&x_1 + x_2 \ge 5, \\
&x_i \ge 0.
\end{aligned}
$$

Поскольку прямая задача на максимум, то приведём все неравенства системы ограничений к виду «£» (обе части первого и четвёртого неравенства умножим на –1):

$$
\begin{cases}
-2x_1 + x_2 \le -1, \\
-x_1 + 4x_2 \le 24, \\
x_1 - x_2 \le 3, \\
-x_1 - x_2 \le -5.
\end{cases}
$$

Составим расширенную матрицу системы:

$$
A_1 = \begin{pmatrix} -2 & 1 & -1 \\ -1 & 4 & 24 \\ 1 & -1 & 3 \\ -1 & -1 & -5 \\ -1 & 2 & f \end{pmatrix}.
$$

Транспонированная матрица:

$$A_1^T = \begin{pmatrix} -2 & -1 & 1 & -1 & -1 \\ 1 & 4 & -1 & -1 & 2 \\ -1 & 24 & 3 & -5 & z \end{pmatrix}.$$

Сформулируем двойственную задачу:

$$z = -y_1 + 24y_2 + 3y_3 - 5y_4 \to \min,$$

$$\begin{cases} 2y_1 - y_2 + y_3 - y_4 \geq -1, \\ y_1 + 4y_2 - y_3 - y_4 \geq 2, \end{cases}$$

$$y_j \geq 0.$$

Теоремы двойственности

Сначала сформулируем основное неравенство теории двойственности.

Пусть имеется пара двойственных задач. Для любых допустимых решений $X = (x_1, x_2, \ldots, x_n)$ и $Y = (y_1, y_2, \ldots, y_m)$ прямой и двойственной задач справедливо неравенство $f(X) \leq z(Y)$ или в координатном виде $\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i$.

Теперь сформулируем достаточный признак оптимальности.

Если X^* и Y^* – допустимые решения соответственно прямой и двойственной задач, для которых справедливо равенство $f(X^*) = z(Y^*)$, то X^* – оптимальное решение прямой задачи, а Y^* – двойственной задачи.

Связь между оптимальными решениями двойственных задач устанавливается с помощью теорем двойственности.

Первая теорема двойственности. Если одна из взаимно двойственных задач имеет оптимальное решение, то его имеет и другая, причём оптимальные значения их целевых функций равны: $f_{\max} = z_{\min}$ или $f(X^*) = z(Y^*)$.

Если целевая функция одной из задач не достигает оптимума, то условия другой задачи противоречивы.

Экономический смысл первой теоремы двойственности состоит в следующем.

План производства X^* и набор цен ресурсов Y^* оптимальны тогда и только тогда, когда прибыль (выручка) от продукции, найденная при «внешних» (известных заранее) ценах $c_1, c_2, \ldots, c_n$, равна затратам на ресурсы по «внутренним» (определяемым только из решения задачи) ценам $y_1, y_2, \ldots, y_m$.

Связь между двойственными задачами проявляется не только в равенстве оптимальных значений их целевых функций (первая теорема двойственности).

Пусть даны двойственные задачи.

Прямая:

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

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

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

Двойственная:

$$z = b_1 y_1 + b_2 y_2 + \ldots + b_m y_m \to \min,$$

$$\begin{cases} a_{11} y_1 + a_{21} y_2 + \ldots + a_{m1} y_1 \ge c_1, \\ a_{12} y_1 + a_{22} y_2 + \ldots + a_{m2} y_2 \ge c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n} y_1 + a_{2n} y_2 + \ldots + a_{mn} x_m \ge c_m, \end{cases}$$

$$y_i \ge 0 \ (i = \overline{1,m}).$$

Если каждую из этих задач решать симплекс-методом, то необходимо привести их к каноническому виду, для чего в систему ограничений

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

прямой задачи следует ввести m неотрицательных переменных $x_{n+1}, x_{n+2}, \ldots, x_{n+i}, \ldots, x_{n+m}$, а в систему ограничений

$$\sum_{i=1}^{m} a_{ij} y_i \ge c_j \quad (j = \overline{1,n})$$

двойственной задачи следует ввести n неотрицательных переменных $y_{m+1}, y_{m+2}, \ldots, y_{m+i}, \ldots, y_{m+n}$, где – номер неравенства, в которое введена дополнительная переменная $x_{n+i} \ge 0 \ (y_{m+i} \ge 0)$. Системы ограничений каждой из задач примут соответственно вид:

прямая задача –

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i \quad (i = \overline{1,m}),$$

двойственная задача –

$$\sum_{i=1}^{m} a_{ij} y_i - y_{m+j} = c_j \quad (j = \overline{1,n}).$$

Вторая теорема двойственности. Чтобы допустимые решения x^*, y^* пары двойственных задач были оптимальными, необходимо и достаточно, чтобы выполнялись условия:

$$(y^*)^T (Ax^* - b) = 0,$$

$$(c - (y^*)^T A)x^* = 0.$$

Следствие. Если в оптимальном решении одной из двойственных задач какая-либо переменная не равна нулю, то соответствующее ей ограничение двойственной задачи на оптимальном решении выполняется как равенство, и наоборот, если на оптимальном решении одной из двойственных задач какое-либо ограничение выполняется как строгое неравенство, то соответствующая ему переменная в оптимальном решении двойственной задачи равна нулю.

Если в одной из взаимно двойственных задач нарушается единственность оптимального решения (геометрически это означает, что оно достигается на ребре/грани многоугольника/многогранника), то оптимальное решение двойственной задачи вырожденное.

1.4 Лекция 6-7 (Л-6-7) (4 ч.)

Тема: Постановка задач линейного программирования транспортного типа. Базовая транспортная модель. Методы решения задач транспортного типа.

1.4.1. Вопросы лекции:

1. Постановка задач линейного программирования транспортного типа.

Базовая транспортная модель.

2. Первичное распределение поставок.

3 Методы решения задач транспортного типа.

1.4.2 Краткое содержание вопросов:

1. Постановка задач линейного программирования транспортного типа.

Базовая транспортная модель.

Имеется целый набор специфических, для которых разработаны особые методы решения задач линейного программирования. В качестве примера таких задач мы рассмотрим так называемую транспортную задачу.

Начнем с её содержательной формулировки.

Пусть имеется некоторый однородный продукт, сосредоточенный на m пунктах отправления (складах), так что на i-м складе находится a_i единиц этого продукта.

Этот продукт необходимо доставить в n пунктов назначения (потребления), причем на *j*-й пункт необходимо доставить b_j единиц продукта. Запасы и потребности сбалансированы, то есть $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j = A$, то есть наличие продукта равно потребности в нем.

Пусть стоимость перевозки единицы продукта из *i*-го склада в *j*-й пункт назначения равна c_{ij}. Пусть x_{ij} есть то количество продукта, которое перевозится из *i*-го склада в *j*-й пункт потребления.

Тогда общие транспортные расходы составят величину $\sum_{i,j} c_{ij} x_{ij}$.

Из каждого склада весь продукт должен быть вывезен. Это значит, что должно быть выполнено условие $\sum_{j=1}^{n} x_{ij} = a_i$.

С другой стороны, потребности *j*-го пункта назначения должны быть полностью удовлетворены. Это означает, что $\sum_{i=1}^{m} x_{ij} = b_j$.

Желание минимизировать транспортные расходы приводит нас к следующей задаче:

$$\sum_{i,j} c_{ij} x_{ij} \Rightarrow \min$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i, & i = \overline{1,m}, \\ \sum_{i=1}^{m} x_{ij} = b_j, & j = \overline{1,n}, \\ x_{ij} \geq 0 \end{cases}$$

являющейся типичной задачей линейного программирования.

Определение. Транспортная задача называется открытой транспортной задачей, если условие баланса нарушаются; в случае выполнения условия баланса она называется сбалансированной транспортной задачей.

Однако у этой задачи есть одна очень существенная особенность: в ограничениях перед неизвестными x_{ij} **всегда стоит 1**. И именно это позволяет разработать гораздо более эффективные и простые алгоритмы решения транспортной задачи, чем симплекс-метод.

Сам же симплекс-метод был бы не эффективен по двум причинам:

1. Большая размерность решаемой задачи. Общее число неизвестных величин x_{ij} равно mn, и даже при $n = m = 10$ размерность решаемой задачи уже будет равна 100. Даже ЭВМ будет решать такую задачу симплекс-методом достаточно долго.

2. Опорные планы в транспортной задаче очень часто бывают вырожденными, а наличие вырождения приводит к необходимости несколько модифицировать симплекс-метод.

Приведение открытой транспортной задачи к сбалансированной

1. Превышение запасов над потребностями.

В этом случае вводится «фиктивный» $n+1$ потребитель с потребностями равными абсолютной величине разности между общим количеством запасов и общим количеством требуемых единиц. Стоимость по доставке будет для $n+1$ потребителя равна 0, т.к. поставки фактически нет.

2. Превышение потребностей над запасами.

Вводим «фиктивного» $m+1$ производителя (склад) с потребностями равными абсолютной величине разности между общим количеством запасов и общим количеством требуемых единиц. Стоимость по доставке будет для $m+1$ производителя равна 0, т.к. поставки фактически нет.

Простейшие свойства транспортной задачи

Теорема 1. Для любой транспортной задачи существует план (то есть для любой транспортной задачи допустимая область не пуста).

Теорема 2. Транспортная задача всегда имеет оптимальный план.

Теорема 3. Любой опорный план имеет не более $n+m-1$ положительных компонент.

Следствие. Оптимальный план содержит не более, чем $n+m-1$ перевозку.

Виды, сводящихся к задаче линейного программирования транспортного типа

Пример 1. На пашне производственного участка сельскохозяйственной организации выделено 3 категории эродированных земель, площадь которых составляет:

¬ I категория – 140;

¬ II – 190;

¬ III – 220 га.

Необходимо так разместить культуры на землях различных категорий, чтобы полученный чистый доход в стоимостном выражении от всего производства был максимальным.

Площадь посева сельскохозяйственных культур следующая:

¬ яровые зерновые – 100;

¬ многолетние травы – 200;

¬ кукуруза – 150;

¬ озимые зерновые – 50 га.

Не предполагается возделывать кукурузу на участке с I категорией эродированных земель. Чистый доход с 1 га приведен в табл. 1

Таблица 1. - Чистый доход с 1 га, у.д.е.

Сельскохозяйственные культуры	Категории эродированных земель		
	I	II	III
Яровые зерновые	120	110	150
Многолетние травы	50	90	70
Кукуруза	-	80	60
Озимые зерновые	140	100	130

Транспортная интерпретация задачи:

ресурсы в источниках A_i - это площади культур;

потребности в ресурсах B_j - площади земель различных категорий эродированности;

транспортная прибыль c_{ij} - чистый доход с 1 га;

транспортируемый ресурс x_{ij} - площадь i-й культуры на землях j-й категории;

целевая функция – максимизация общего чистого дохода производственного участка сельскохозяйственной организации.

2. Первичное распределение поставок.

Пусть задана некоторая классическая транспортная задача.

Предположим, что имеется n пунктов потребления (например, промышленных предприятий или типографий) Пj, j $\in$ {1..n}, требующих снабжения некоторым определенным видом сырья. Потребности в сырье каждого предприятия Пj равны соответственно bj условных единиц j $\in$ {1..n}. Кроме того, имеется m складов Ci, i $\in$ {1..m}, на которых хранится требуемое предприятиям сырье. На каждом складе Ci имеется в наличии запас товара в количестве ai соответственно, i $\in$ {1..m}.

Склады удалены от предприятий на некоторые расстояния и связаны с ними некоторыми путями сообщений с различными тарифами на перевозку грузов (в нашем случае сырья). Будем считать, что каждый склад связан с каждым пунктом потребления некоторым единственным маршрутом с неограниченной пропускной способностью. Единица сырья, получаемая предприятием Пj со склада Ci с учетом известных тарифов на перевозки, обходится в cij рублей. Для простоты предположим, что все заявки выполнимы и обеспечивают отсутствие излишек на складах, т.е. сумма всех заявок в точности равна сумме всех имеющихся запасов: $\sum_{j=1}^{n} b_j = \sum_{i=1}^{m} a_i$. Требуется составить план перевозок, т.е. указать с какого склада на какие предприятия, и какое количество сырья нужно направить, чтобы заявки были выполнены, а общие расходы на все перевозки были минимальными.

Составим для данной задачи, как это уже было показано раньше таблицу издержек:

Сток Исток	$П_1$	$П_2$	…	$П_n$	Запасы: $\sum a_i$
C_1	c_{11}	c_{12}	…	c_{1n}	a_1
C_2	c_{21}	c_{22}	…	c_{2n}	a_2
…	…	…	…	…	…
C_m	c_{m1}	c_{m2}	…	c_{mn}	a_m

Заявки: $\sum b_j$	b_1	b_2	…	b_n	$\sum_i a_i = \sum_j b_j$

Очевидно, что не все m + n уравнений системы ограничений транспортной задачи являются независимыми. Действительно. складывая все ограничения по заявкам и все ограничения по запасам, в силу равенства заявок и запасов, получаем доказательство того, что ранг системы ограничений r = m + n – 1. Следовательно, можно разрешить эти уравнения относительно r базисных переменных, выразив их через остальные k свободные.

k = m·n – r = m·n – m – (n – 1) = m·(n – 1) – (n – 1) ⇒ k = (m – 1)·(n – 1).

Мы уже останавливались на факте того, что в задаче линейного программирования оптимальное решение достигается в одной из вершин области допустимых решений, где, по крайней мере, k переменных обращаются в нуль.

Значения x_{ij} количества единиц груза, направляемых из пункта Ci в пункт Пj будем называть перевозками. Любую совокупность значений (x_{ij}) будем называть планом перевозок, или просто планом. Тогда план будем называть допустимым, если он удовлетворяет балансовым условиям: все заявки удовлетворены и, все запасы исчерпаны.

В свою очередь, допустимый план будем называть оптимальным в том случае, если он приводит к наименьшей стоимости всех перевозок.

Методы решения транспортной задачи не требуют манипуляций с симплекс-таблицами, а сводятся к операциям с таблицей, где в определенном порядке (см. примеры таблиц издержек выше) записаны все условия транспортной задачи – транспортной таблицей. Стоимость перевозок (соответствующие издержки из таблицы издержек) c_{ij} будем помещать в правом верхнем углу каждой ячейки, с тем, чтобы в самой ячейке помещать значения соответствующих перевозок (x_{ij}). Ниже приведен пример заполнения транспортной таблицы:

Сток Исток	$П_1$	$П_2$	…	$П_n$	Запасы: $\sum a_i$
C_1	c_{11} x_{11}	c_{12} x_{12}	… …	c_{1n} x_{1n}	a_1
C_2	c_{21} x_{21}	c_{22} x_{22}	… …	c_{2n} x_{2n}	a_2
…	… …	… …	… …	… …	…
C_m x_{m1}	c_{m1} x_{m2}	c_{m2}	… x_{mn}	c_{mn}	a_m
Заявки: $\sum b_j$	b_1	b_2	…	b_n	$\sum_i a_i = \sum_j b_j$

Ячейки таблицы с отличными от нуля перевозками условимся называть базисными, а остальные (пустые – в дальнейшем в транспортную таблицу мы будем заносить только отличные от нуля значения) свободными.

Решение транспортной задачи, как и всякой задачи линейного программирования, начинается с нахождения опорного решения или, в случае транспортной задачи, опорного плана перевозок. В отличие от общего случая задачи линейного программирования, решение транспортной задачи всегда существует. Рассмотрим один из способов построения опорного плана – так называемый метод «северо-западного угла».

Нахождение опорного плана методом «северо-западного угла»

Метод «северо-западного» угла реализует интерактивный поиск опорного решения. Идея метода состоит в следующем. На каждой итерации выбирается такое решение, которое бы с одной стороны учитывало бы результаты предыдущих итераций, а с другой стороны по возможности минимизировало бы количество оставшихся итераций.Применительно к транспортной таблице работу метода «северо-западного угла» можно представить так. Первоначально заполняется самая левая верхняя клетка (северо-западный угол таблицы). Это означает, что мы принудительно посылаем товар со склада 1 потребителю 1. Если количество имеющегося в наличии товара на складе 1 превышает размер запроса потребителя 1, то в северо-западную клетку (1,1) следует поместить значение запроса потребителя 1. В противном случае, то есть в ситуации, при которой склад 1 не в состоянии самостоятельно удовлетворить запрос потребителя 1, в ячейку (1,1) транспортной таблицы следует поместить значение запаса склада 1.

Очевидно, что в случае неравенства (или равенства) запаса на складе 1 запросу потребителя 1 возможны два варианта:

1. Склад 1 полностью удовлетворил запрос потребителя 1, но при этом запас его товаров еще не исчерпан. Тогда на следующей итерации алгоритма мы будем рассматривать соседнюю с востока ячейку, то есть ячейку (1,2) транспортной таблицы.
2. Склад 1 не в состоянии полностью удовлетворить запрос потребителя 1, то есть запас его товаров исчерпан, а запрос потребителя еще не удовлетворен. Тогда на следующей итерации алгоритма мы будем рассматривать соседнюю с юга ячейку, то есть ячейку (2,1) транспортной таблицы.
3. Склад 1 полностью удовлетворил потребности потребителя 1, при этом запас товаров на складе был полностью исчерпан. Тогда на следующей итерации алгоритма мы будем рассматривать соседнюю с юго-востока ячейку, то есть ячейку (2,2) транспортной таблицы.

Описанная процедура обеспечивает выбор ячейки транспортной таблицы для следующей итерации алгоритма поиска опорного решения.

Очевидно, что в силу равенства суммарных заявок (запросов) суммарным запасам на последней итерации алгоритма – ячейка (i,j) – в общем случае окажется, что i-ый склад оставшимся своим запасом полностью удовлетворит оставшийся неудовлетворенным на предыдущих итерациях запрос j-того потребителя.

Рассмотрим работу данного метода на конкретном примере. Пусть задана некоторая транспортная задача и соответствующая ей транспортная таблица:

Сток Исток	B_1	B_2	B_3	B_4	B_5	Запасы:
A_1	10	8	5	6	9	48
A_2	6	7	8	6	5	30
A_3	8	7	10	8	7	27
A_4	7	5	4	6	8	20

Заявки:	18	27	42	12	26	125

Тогда в соответствии с только что описанным методом «северо-западного угла» будем заполнять таблицу перевозками, начиная с северо-западной ячейки (1,1), рассуждая так. Удовлетворим пункт B1 за счет запаса A1, следовательно x11= 18. После этого в пункте A1 осталось 30 единиц груза. Удовлетворим запрос пункта B2 за счет остатка A1, следовательно x12 = 27. оставшиеся 3 единицы груза направим в пункт B3 □ x13 = 3. В составе заявки пункта B3 осталось неудовлетворенным 39 единиц груза. Из них 30 покроем за счет пункта A2, чем его запас будет исчерпан, и еще 9 единиц возьмем из пункта A3, следовательно x23 = 30 и x33 = 9 и т.д. Полученное решение будет не только допустимым, но и опорным.

В результате этих действий получим следующее опорное решение:

Сток Исток	B_1	B_2	B_3	B_4	B_5	асы:
A_1	10 18	8 27	5 3	6		48
A_2	6	7	8 30	6		30
A_3	8	7	10 9	8 12	6	27
A_4	7	5	4	6	20	20
вки:	8	7	2	2	6	25

Остановимся теперь на одной особенности плана перевозок. Речь идет о так называемом «вырожденном» плане, в котором некоторые из базисных перевозок оказываются равными 0.

Забегая вперед, отметим, что для решения транспортной задачи необходимо, чтобы уравнения, формирующие план перевозок, имели базис размерности m + n – 1. В противном случае дальнейшее решение транспортной задачи становится не возможным. Исходя из этого, можно сделать вывод о необходимости строго поддерживать размерность базиса, равную m + n – 1. Тогда в случае получения на некоторой итерации вырожденного плана перевозок необходимо искусственным образом ввести дополнительную базисную переменную. Для этого в любую из свободных клеток транспортной таблицы следует поместить некоторую бесконечно малую величину□ и соответственно скорректировать значения всех соседних базисных клеток. В качестве иллюстрации метода преобразования вырожденного плана, рассмотрим довольно простой пример, в котором вырожденный план перевозок получается при нахождении опорного плана методом северо-западного угла.

Особенностью этого опорного плана является то, что в нем только 6 , а не 8 отличных от нуля перевозок (r = m + n – 1 = 4 + 5 – 1 = 8).

В дальнейшем нам удобно будет всегда иметь в транспортной таблице r базисных клеток, хотя в некоторых из них, может быть, будут стоять нулевые значения перевозок. Для этого можно ничтожно мало изменить запасы или заявки, например, на величину □, а после нахождения оптимального решения положить □ = 0.

Улучшение плана перевозок. Цикл пересчета

Метод «северо-западного» угла позволяет отыскать допустимый план перевозок, который мы назвали опорным планом. Очевидно, что в общем случае опорный план, являясь допустимым, не является оптимальным. Для нахождения оптимального плана перевозок необходимо последовательно, пока это возможно, улучшать опорный план. В этом параграфе мы рассмотрим термины и определения, используемые при улучшении опорного плана, которые нам в дальнейшем потребуются при рассмотрении алгоритмов решения транспортной задачи.

Возвращаясь к нашему исходному примеру, рассмотрим его опорное решение – опорный план перевозок.

Сток Исток	B_1	B_2	B_3	B_4	B_5	Запас
A_1	1 18	8 27	5 3	6	9	48
A_2	6	7	8 30	6	5	30
A_3	8	7	1 9	8 12	7 6	27
A_4	7	5	4	6	8 20	20
Заявки:	18	27	42	12	26	125

Стоимость этого плана равна: L = 18*10 + 27*8 + 3*5 + 30*8 + 9*10 + 12*8 + 6*7 + 20*8 = 1039.

Попробуем улучшить этот план, перенеся, как показано в приведенной выше таблице, 18 единиц из клетки (1,1) в клетку (2,1) и, чтобы не нарушать баланса, перенесем те же 18 единиц из клетки (2,3) в клетку (1,3). В результате переноса мы получили новый план перевозок, который тоже будет допустимым, так как переброску груза с одного маршрута на другой мы осуществляли циклически, заботясь о сохранении баланса заявок и запасов.

Сток Исток	B_1	B_2	B_3	B_4	B_5	Запасы:
A_1	10	8 27	5 21	6	9	48
A_2	6 18	7	8 12	6	5	30
A_3	8	7	10 9	8 12	7 6	27
A_4	7	5	4	6	8 20	20
Заявки:	18	27	42	12	26	125

Таким образом, мы получили новый допустимый план, стоимость которого равна:
L = 18*6 + 27*8 + 21*5 + … = 913.

Очевидно, что полученный план перевозок по стоимости предпочтительнее первоначального опорного плана.

Из приведенной таблицы видно, что за счет циклической перестановки грузоперевозок объемом 18 единиц по маршрутам (1,1)□, (2,1)+, (2,3)□, (1,3)+ удалось понизить стоимость плана перевозок на 126 условных единиц стоимости.

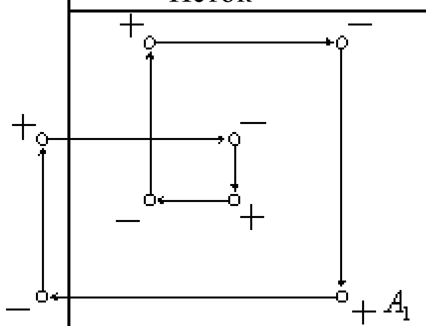
Здесь следует обратить внимание на следующее равенство:

1039 – 913 = –126 = 18□(□10 + 6 – 8 + 5) = 18□(□7).

Рассмотрев на практическом примере принципы, лежащие в основе методов улучшения планов перевозок, формально определим некоторые из использованных нами при этом понятий.

Итак, циклом в транспортной таблице назовем несколько клеток, соединенных замкнутой ломаной линией, которая в каждой клетке совершает поворот на угол, равный 90о. Условимся помечать символами (+) те вершины ломаной линии, образующей цикл, в которых объемы перевозок увеличивается, а символом (□) те вершины цикла, в которых они уменьшаются.

Очевидно, что прямоугольник представляет собой наиболее простой случай такой замкнутой ломаной линии. В таблице, расположенной ниже, представлен более сложный пример возможного цикла:

Сток Исток B_1	B_2	B_3	B_4	B_5		асы:
 A_1						18
A_2						30
A_3						27
A_4						20
Заявки:						25

Цикл с отмеченными вершинами будем называть «означенным». Перенести какое-то количество единиц груза по означенному циклу – это означает увеличить объемы перевозок в положительных вершинах (вершинах, помеченных символом «+») на это количество единиц и одновременно с этим уменьшить перевозки на то же количество в отрицательных вершинах (вершинах, помеченных символом «–»).

Очевидно что, при переносе любого числа единиц по циклу равновесие между запасами и заявками не меняется.

Назовем ценой (стоимостью) цикла– алгебраическую сумму стоимостей, стоящих в вершинах цикла, с учетом знака этих вершин, например:

$$a_1 = c_{21} - c_{23} + c_{43} - c_{41},$$

$$a_2 = c_{34} - c_{35} + c_{55} - c_{54} + c_{14} - c_{16}.$$

3 Методы решения задач транспортного типа.

Методы решения транспортной задачи закрытого и открытого типов.

Распределительный метод решения транспортной задачи

Данный метод состоит в последовательном улучшении опорного плана перевозок путем отыскания на каждом шаге выгодных циклов переноса грузов. Опорный план для данного метода (как и для других методов решения транспортной задачи методом потенциалов) можно сформировать, применяя метод «северо-западного» угла.

Более подробно рассмотрим теперь процесс формирования очередного цикла переноса на каждом новом шаге алгоритма.

Очевидно, что при перемещении x единиц груза по некоторому циклу с ценой g стоимость перевозок изменяется на величину x×g.

Тогда, для улучшения текущего плана перевозок имеет смысл перемещать перевозки только по тем циклам, цена которых отрицательна.

Если циклов с отрицательной ценой в таблице больше не осталось, это означает, что оптимальный план достигнут.

При улучшении плана циклическими переносами пользуются приемом, заимствованным из симплекс-метода: на каждом шаге (цикле) заменяют одну свободную переменную на базисную, т.е. заполняют одну клетку и взамен того освобождают одну из базисных клеток.

Можно доказать, что для любой свободной клетки транспортной таблицы всегда существует цикл (и притом единственный), одна из вершин которого лежит в этой клетке, а все остальные в базисных клетках. Если цена такого цикла, с плюсом в свободной клетке, отрицательна, то план можно улучшить. Количество единиц груза (x), которые можно переместить, определяется минимальным значением перевозок, стоящих в отрицательных вершинах цикла.

Теорема о платежах транспортной задачи

Для заданной совокупности платежей (α_i, β_j) суммарная псевдостоимость плана перевозок, равная величине:

$$\widetilde{L} = \sum_{i=1}^{m}\sum_{j=1}^{n}\widetilde{c}_{ij}x_{ij}$$

.

при любом допустимом плане перевозок (x_{ij}) не зависит от конкретного плана перевозок, то есть сохраняет одно и то же постоянное значение.

Теорема об оптимальном плане транспортной задачи

Если для всех базисных клеток некоторого допустимого плана перевозок псевдостоимость совпадает со стоимостью перевозки по соответствующему маршруту, то есть $(x_{ij} > 0), \widetilde{c}_{ij} = c_{ij}$, а для всех свободных клеток псевдостоимость не превышает стоимости перевозки, то есть $(x_{ij} = 0), \widetilde{c}_{ij} \le c_{ij}$, то рассматриваемый допустимый план перевозок является оптимальным.

Нетрудно догадаться, что эта теорема справедлива и для вырожденного плана, в котором некоторые из базисных переменных равны нулю. Действительно то, что в базисных клетках перевозки строго положительны, для доказательства несущественно: достаточно, чтобы они были неотрицательны.

Таким образом, нами доказано, что признаком оптимальности допустимого плана перевозок (x_{ij}) является выполнение двух условий:

$$\widetilde{c}_{ij} = c_{ij} \quad \forall(i, j): x_{ij} \in B \qquad (a)$$

$$\widetilde{c}_{ij} \le c_{ij} \quad \forall(i, j): x_{ij} \notin B \qquad (б)$$

План перевозок, обладающий таким свойством, называется потенциальным, а соответствующие ему платежи (α_i, β_j) - потенциалами пунктов A_i и B_j.

Оптимальный план перевозок можно построить методом последовательных приближений, задаваясь сначала какой-то произвольной системой платежей, удовлетворяющих условию (а). Затем следует улучшить план, одновременно меняя систему платежей так, чтобы они приближались к потенциалам. При улучшении плана перевозок нам помогает следующее свойство платежей и псевдостоимостей:

Для любой системы платежей (α_i, β_j), удовлетворяющей условию (а), каждая свободная клетка имеет цену цикла, равную разности между стоимостью и псевдостоимостью в этой клетке транспортной таблицы:

$\gamma_{ij} = c_{ij} - \tilde{c}_{ij}$.

Таким образом, можно предположить следующий алгоритм решения классической транспортной задачи методом потенциалов:

1. Взять любой опорный план перевозок, в котором отмечены $m+n-1$ базисные клетки.
2. Определить для этого плана платежи (α_i, β_j) исходя из требования, чтобы в любой базисной клетке псевдостоимости были равны стоимостям:

$\alpha_i + \beta_j = c_{ij}$.

Очевидно, что один из платежей можно назначить произвольно, например, положить равным нулю, так как количество уравнений вида $\alpha_i + \beta_j = c_{ij}$ всего равно величине $m+n-1$, а число неизвестных $m+n$.

3. Подсчитать псевдостоимости $\tilde{c}_{ij} = \alpha_i + \beta_j$ для всех свободных клеток. Если окажется, что выполняется условие $\tilde{c}_{ij} \le c_{ij}$ (условие (а)), то рассматриваемый план перевозок является оптимальным.
4. Если условие (б) не выполняется, то есть $\tilde{c}_{ij} > c_{ij}$ хотя бы в одной свободной клетке, то следует улучшить план путем переноса перевозок по циклу любой свободной клетки с отрицательной ценой $(\tilde{c}_{ij} > c_{ij})$.
5. Перейти к пункту 2.

Важно заметить, что понятие цикла в методе потенциалов ничем не отличается от такого же понятия в распределительном методе. Таким образом, мы не будем здесь повторно рассматривать вопросы, связанные с правилами формирования и использования циклов в транспортной таблице.

Рассмотренным выше понятиям платежей за перевозку и псевдостоимостей можно дать следующую экономическую интерпретацию.

Пусть поставщики A_i и потребители B_j действуют как единая экономическая система, а платежи (α_i, β_j) - реальные платежи, которые участники системы A_i и B_j вынуждены платить за перевозку единицы груза «перевозчику».

Перевозка единицы груза из пункта A_i в пункт B_j объективно стоит c_{ij} условных единиц. В то же время соответствующий поставщик и потребитель вместе платят за эту же перевозку сумму в $\tilde{c}_{ij} = \alpha_i + \beta_j$ условных единиц.

Тогда с точки зрения вносимых за перевозку груза платежей оптимальным будет такой план перевозок, при котором пункты отправления и назначения A_i и B_j не переплачивают «перевозчику» ничего сверх объективной стоимости перевозок, то есть выполняется условие $\tilde{c}_{ij} \le c_{ij}$.

Проиллюстрируем описанный выше алгоритм решения классической транспортной задачи методом потенциалов на нескольких примерах. Здесь, как и при рассмотрении распределительного метода, мы рассмотрим два примера решения транспортной задачи.

Для того, чтобы дать читателю возможность сравнить решение задачи различными методами и понять преимущества и недостатки этих методов по сравнению друг с другом, примеры данного раздела будут идентичны уже рассмотренным нами примерам.

1.5 Лекция 8 (Л-8) (2 ч.)

Тема: Марковские процессы, их приложения к решению инженерных задач.

1.5.1. Вопросы лекции:

1. Основные понятия теории марковских процессов. Поток СС, простейший поток, его свойства.

2. Классификация марковских процессов.

3. Основные понятия теории систем массового обслуживания.

1.4.2 Краткое содержание вопросов:

1. Основные понятия теории марковских процессов. Поток СС, простейший поток, его свойства.

Поток событий. Простейший поток и его свойства

При рассмотрении процессов, протекающих в системе с дискретными состояниями и непрерывным временем, часто бывает удобно представить себе процесс так, как будто переходы системы из состояния в состояние происходят под действием каких-то потоков событий. Потоком событий называется последовательность однородных событий, следующих одно за другим в какие-то, вообще говоря, случайные моменты времени. (Поток вызовов на телефонной станции; поток неисправностей (сбоев) ЭВМ; поток грузовых составов, поступающих на станцию; поток посетителей; поток выстрелов, направленных на цель). Будем изображать поток событий последовательностью точек на оси времени *ot*. Положение каждой точки на оси случайно. Поток событий называется ***регулярным***, если события следуют одно за другим через строго определенные промежутки времени (редко встречается на практике). Рассмотрим специального типа потоки, для этого введем ряд определений. 1. Поток событий называется ***стационарным***, если вероятность попадания того или иного числа событий на участок времени длиной τ зависит только от длины участка и не зависит от того, где именно на оси ot расположен этот участок (однородность по времени) – вероятностные характеристики такого потока не должны меняться от времени. В частности, так называемая интенсивность (или плотность) потока событий (среднее число событий в единицу времени) постоянна.

2. Поток событий называется **потоком без последствия**, если для любых непересекающихся участков времени число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой (или другие, если рассматривается больше двух участков). Отсутствие последствия в потоке означает,

что события, образующие поток, появляются в последовательные моменты времени независимо друг от друга.

3. Поток событий называется ***ординарным***, если вероятность попадания на элементарный участок двух или более событий пренебрежительно мала по сравнению с вероятностью попадания одного события (события в потоке приходят поодиночке, а не парами, тройками и т. д.).

Поток событий, обладающий всеми тремя свойствами, называется ***простейшим*** (*или стационарным пуассоновским*). Нестационарный пуассоновский поток обладает только свойствами 2 и 3. Пуассоновский поток событий (как стационарный, так и нестационарный) тесно связан с известным распределением Пуассона. А именно, число событий потока, попадающих на любой участок, распределено по закону Пуассона. Поясним это подробнее.

Рассмотрим на оси *ot*, где наблюдается поток событий, некоторый участок длины t, начинающийся в момент *t0* и заканчивающийся в момент *t0*+t. Нетрудно доказать (доказательство дается во всех курсах теории вероятности), что вероятность попадания на этот участок ровно m событий выражается формулой:

$$P_m = \frac{a^m}{m!}e^{-a} \quad (m=0,1\ldots),$$

где *a* – среднее число событий, приходящееся на участок t.

Для стационарного (простейшего) пуассоновского потока *a=lt*, т. е. не зависит от того, где на оси *ot* взят участок t. Для нестационарного пуассоновского потока величина *a* выражается формулой

$$a = \int_{t_0}^{t_0+\tau} \lambda(t)dt$$

и значит, зависит от того, в какой точке t_0 начинается участок t.

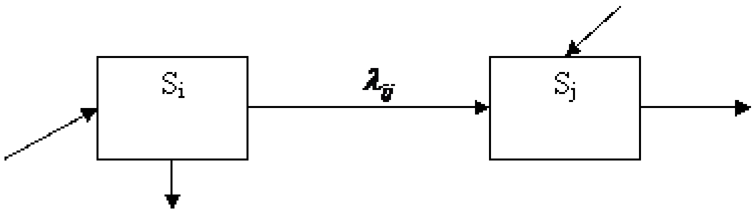
Рассмотрим на оси *ot* простейший поток событий с постоянной интенсивностью l. Нас будет интересовать интервал времени T между событиями в этом потоке. Пусть l - интенсивность (среднее число событий в 1 времени) потока. Плотность распределения *f(t)* случайной величины *T* (интервал времени между соседними событиями в потоке) *f(t)=le-lt (t>0)*. Закон распределения с такой плотностью называется показательным (экспоненциальным). Найдем численные значения случайной величины *T*: математическое ожидание (среднее значение) $m_t = \int_0^{\infty} t f(t)dt = \frac{1}{\lambda}$ и дисперсию $D_t = \int_0^{\infty} t^2 f(t)dt - m_t^2 = \frac{1}{\lambda^2}$.

Промежуток времени *T* между соседними событиями в простейшем потоке распределен по показательному закону; его среднее значение и среднее квадратичное отклонение равны $\frac{1}{\lambda}$, где l - интенсивность потока. Для такого потока вероятность появления на элементарном участке времени Δt ровно одного события потока выражается как $P_1(\Delta t) \approx \lambda\Delta t$. Эту вероятность мы будем называть «элементом вероятности появления события».

Для нестационарного пуассоновского потока закон распределения промежутка T уже не будет показательным. Вид этого закона будет зависеть, во- первых, от того, где на оси *ot* расположено первое из событий, во-вторых, от вида зависимости $\lambda(t)$. Однако, если $\lambda(t)$ меняется сравнительно медленно и его изменение за время между двумя событиями невелико, то закон распределения промежутка времени между событиями можно приближенно считать показательным, полагая в этой формуле величину λ равной среднему значению $\lambda(t)$ на том участке, который нас интересует.

Пуассоновские потоки событий и непрерывные марковские цепи

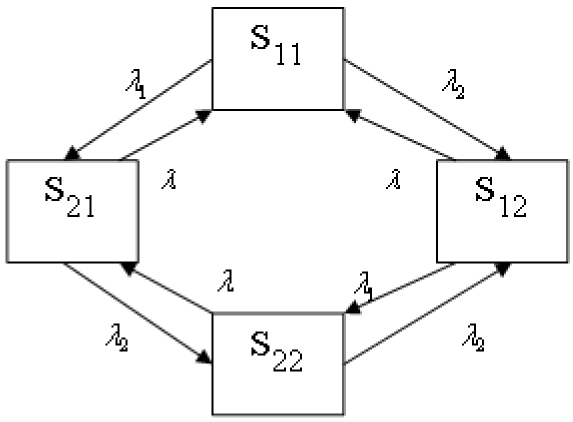
Рассмотрим некоторую физическую систему *S={S1,S2,…Sn}*, которая переходит из состояния в состояние под влиянием каких-то случайных событий (вызовы, отказы, выстрелы). Будем себе это представлять так, будто события, переводящие систему из состояния в состояние, представляют собой какие-то потоки событий.



Пусть система S в момент времени t находится в состоянии Si и может перейти из него в состояние Sj под влиянием какого-то пуассоновского потока событий с интенсивностью lij: как только появляется первое событие этого потока, система мгновенно переходит из Si в Sj . Как мы знаем, вероятность этого перехода за элементарный промежуток времени Δt (элемент вероятности перехода) равна, отсюда вытекает, что плотность вероятности перехода lij в непрерывной цепи Маркова представляет собой не что иное, как интенсивность потока событий, переводящих систему по соответствующей стрелке. Если все потоки событий, переводящие систему S из состояния в состояние пуассоновские, то процесс, протекающий в системе, будет марковским.

Проставим интенсивности пуассоновских потоков (плотности вероятностей переходов) на графе состояний системы у соответствующих стрелок. Получим размеченный граф состояний. На его основе можно написать уравнения Колмогорова и вычислить вероятности состояний.

Пример. Техническая система S состоит из двух узлов I и II, каждый из которых независимо от другого может отказывать. Поток отказов первого узла пуассоновский с интенсивностью lI, второго также пуассоновский с интенсивностью lII. Каждый узел сразу после отказа начинает ремонтироваться (восстанавливаться). Поток восстановлений (окончаний ремонта узла) для обоих узлов – пуассоновский с интенсивностью l. Составить граф состояний системы и написать уравнение Колмогорова. Состояния системы: S_{11} - оба узла исправны; S_{21} – первый узел ремонтируется, второй исправен; S_{12}, S_{22}.



$$\begin{cases} \frac{dp_{11}}{dt} = -(\lambda_I + \lambda_{II})p_{11} + \lambda\, p_{21} + \lambda p_{12} \\ \frac{dp_{21}}{dt} = -(\lambda + \lambda_{II})p_{21} + \lambda_I p_{11} + \lambda\, p_{22} \\ \frac{dp_{12}}{dt} = -(\lambda + \lambda_I)p_{12} + \lambda_{II} p_{11} + \lambda p_{22} \\ \frac{dp_{22}}{dt} = -2\lambda p_{22} + \lambda_{II} p_{21} + \lambda_I p_{12} \end{cases}$$

$$t=0\ \ p_{11}=1\ \ p_{21}=p_{22}=p_{12}=0$$

$$p_{11}+p_{12}+p_{21}+p_{22}=1.$$

2. Классификация марковских процессов.

Аппарат теории марковских процессов с дискретными состояниями и цепей Маркова широко используют в теории систем, в исследовании операций и других прикладных дисциплинах. Это обусловлено многими причинами, среди которых отметим следующие:

1) многие реальные технические системы имеют конечные множества возможных состояний, а их поведение в процессе функционирования адекватно моделируется марковскими процессами,

2) теория марковских процессов с дискретными состояниями и цепей Маркова разработана настолько глубоко, что позволяет решать широкий класс прикладных задач.

Марковские процессы. Представление случайных процессов графом состояний

Рассмотрим физическую систему *S*, в которой протекает случайный процесс с дискретными состояниями: $s_1, s_2, \ldots, s_i$, (1) число которых конечно (или счетно). Состояния $s_1, s_2, \ldots$ могут быть качественными (т. е. описываться словами) или же каждое из них характеризуется случайной величиной (либо случайным вектором).

Прежде всего, рассмотрим множество состояний (1) с точки зрения его структуры - возможности системы *S* переходить из состояния s_j в данное состояние s_i - непосредственно или через другие состояния. Для этого удобно пользоваться наглядной схемой, так называемым графом состояний. Здесь и далее мы будем отчасти пользоваться терминологией теории графов. Имеется две основные разновидности графов: неориентированные и ориентированные.

Неориентированный граф - совокупность точек (вершин графа) с соединяющими некоторые из них отрезками (ребрами графа).

Ориентированный граф - это совокупность точек (вершин) с соединяющими некоторые из них ориентированными отрезками (стрелками).

При изложении теории случайных процессов с дискретными состояниями мы будем пользоваться только ориентированными графами. Вершины графа будут соответствовать состояниям системы. Вершину будем изображать прямоугольником, в который вписано обозначение состояния; стрелка, ведущая из вершины s_j в вершину s_i , будет обозначать возможность перехода системы *S* из состояния s_j в состояние s_i - непосредственно, минуя другие состояния. Стрелки графа могут изображаться не только прямолинейными, но и криволинейными отрезками (рис. 1). Сам граф системы *S* будем обозначать буквой *G*.

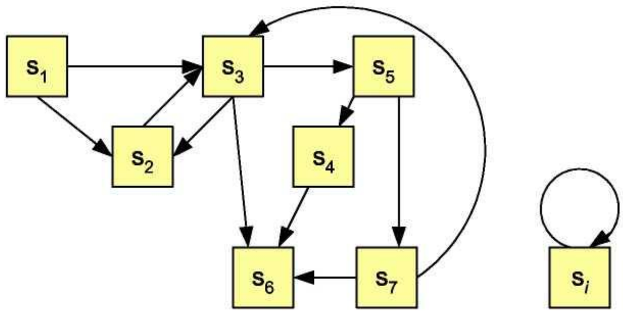


Рисунок 1 – Пример графа состояний

Переход по стрелке, ведущей из состояния s_i в него же, означает задержку системы в состоянии s_i. «Обратные стрелки» можно на графе не проставлять, так как все расчеты можно вести и без них.

Проведем некоторую необходимую для дальнейшего классификацию состояний. Состояние s_i называется источником, если система S может выйти из этого состояния, но попасть в него обратно уже не может, т. е. на графе G состояний в состояние s_i не ведет ни одна стрелка. На рисунке 1 состояние s_1 является источником.

Состояние s_i называется концевым (или поглощающим), если система S может попасть в это состояние, но выйти из него уже не может. Для графа состояний это означает, что из состояния s_i не ведет ни одна стрелка (для графа, изображенного на рисунке 1, состояние s_6 поглощающее).

Если система S может непосредственно перейти из состояния s_i в состояние s_j то состояние s_j - называется соседним по отношению к состоянию s_i.

Состояние s_i называется транзитивным, если система S может войти в это состояние и выйти из него, т. е. на графе состояний есть хотя бы одна стрелка, ведущая в s_i и хотя бы одна стрелка, ведущая из s_i. На рисунке 1 все состояния, кроме s_1 и s_6, являются транзитивными.

Для полноты картины можно рассматривать также и «изолированные» состояния. Состояние s_i называется изолированным, если из него нельзя попасть ни в одно из других состояний и в него нельзя попасть ни из какого другого состояния.

Наряду с отдельными состояниями системы S в ряде задач практически бывает нужно рассматривать подмножества ее состояний.

Обозначим W множество всех состояний системы S (конечное или бесконечное, но счетное) и рассмотрим его подмножество $V \subset W$. Подмножество V называется замкнутым (концевым), если система S, попав в одно (или находясь в одном) из состояний $s_i \in V$, не может выйти из этого подмножества состояний. Концевое подмножество состояний может включать в себя поглощающее состояние, а может и не включать.

Подмножество состояний $V \subset W$ называется связным или эргодическим, если из любого состояния, входящего в него, можно попасть в любое другое состояние, принадлежащее этому подмножеству. Эргодическим может быть и все множество *W* состояний системы *S*. В эргодическом множестве состояний нет ни источников, ни поглощающих состояний.

Подмножество состояний *V* называется транзитивным, если система *S* может войти в это подмножество и выйти из него, т. е. из любого состояния $s_i \subset V$ можно (за то или другое число перескоков) выйти из этого подмножества.

Случайный процесс, протекающий в системе *S*, можно трактовать как процесс блуждания системы по множеству состояний *W*. Если подмножество $V \subset W$ является концевым, то, попав в него, система будет продолжать блуждание уже по этому подмножеству состояний *V*. Если все множество эргодично, то блуждание будет происходить по всем его состояниям.

На практике очень часто встречаются системы, состояния которых образуют цепь (рисунок 2), в которой каждое состояние s_i (кроме двух крайних s_0 и s_n) связано прямой и обратной связью с двумя соседними s_{i-1}, s_{i+1}, а каждое из двух крайних связано прямой и обратной связью только с одним соседним.

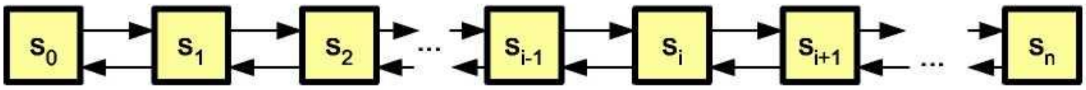


Рисунок 2 - Схема процесса гибели и размножения

Такая схема случайного процесса называется схемой гибели и размножения, а сам процесс — процессом гибели и размножения.

Если на графе состояний системы *S* стрелки, ведущие справа налево, отсутствуют, то говорят о процессе «чистого размножения», в противоположном случае — о процессе «чистой гибели».

Процесс гибели и размножения может в некоторых случаях иметь не конечное число состояний: $s_1, s_2, \ldots, s_i, \ldots, s_n$, а бесконечное (счетное): $s_1, s_2, \ldots, s_i, \ldots$.

При анализе случайных процессов, протекающих в системах с дискретными состояниями, важную роль играют вероятности состояний.

Обозначим *S(t)* состояние системы *S* в момент *t*. Вероятностью *i*-го состояния в момент *t* называется вероятность события, состоящего в том, что в момент *t* система *S* будет в состоянии s_i. Обозначим ее $p_i(t)$:

$$p_i(t) = P\{S(t) = s_i\}, \qquad (2)$$

где *S(t)* - случайное состояние системы *S* в момент *t*. Очевидно, что для системы с дискретными состояниями $s_1, s_2, \ldots, s_i, \ldots,$ в любой момент *t* сумма вероятностей состояний равна единице:

$$\sum_i p_i(t) = 1, \qquad (3)$$

как сумма вероятностей полной группы несовместных событий.

В ряде задач практики нас интересует так называемый установившийся или стационарный режим работы системы, который в ней устанавливается, когда от начала процесса прошло достаточно большое время *t*. Например, процесс изменения напряжения в сети питания технического устройства, пройдя сразу после включения через ряд колебаний, по прошествии времени, устанавливается. Аналогично этому и в некоторых случайных процессах по прошествии достаточно большого времени *t* устанавливается стационарный режим, во время которого состояния системы хотя и меняются случайным образом, но их вероятности $p_i(t)$ $(i = 1, 2, \ldots)$ остаются постоянными. Обозначим эти постоянные вероятности p_i.

$$p_i = \lim p_i(t) \qquad (4)$$

Вероятности $p_i (i = 1, 2, \ldots)$, если они существуют, называются финальными (предельными) вероятностями состояний. Финальную вероятность p_i можно истолковать как среднюю долю времени, которую в стационарном режиме проводит система *S* в состоянии s_i. В дальнейшем будет показано, при каких условиях финальные вероятности существуют и какими они могут быть для разных состояний и подмножеств состояний.

Введем очень важное для дальнейшего понятие марковского случайного процесса.

Случайный процесс, протекающий в системе *S* с дискретными состояниями $s_1, s_2, \ldots, s_i, \ldots,$ называется марковским, если для любого момента времени t_0 вероятность каждого из состояний системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, когда и как она пришла в это состояние; т. е. не зависит от ее поведения в прошлом (при $t < t_0$).

Не надо понимать марковское свойство случайного процесса как полную независимость «будущего» от «прошлого»; в общем случае «будущее» зависит от «настоящего», т. е. вероятности $p_i(t)$ при $t > t_0$ зависят от того, в каком состоянии s_i находится система в настоящем (при $t = t_0$); само же это «настоящее» зависит от «прошлого», от того, как вела себя система *S* при $t < t_0$. Это можно сформулировать следующим образом: для марковского случайного процесса «будущее» зависит от «прошлого» только через «настоящее» (рисунок 3). При фиксированном «настоящем» условные вероятности всех состояний системы в «будущем» не зависят от предыстории процесса, т. е. от того, когда и как система S к моменту t_0 пришла в состояние

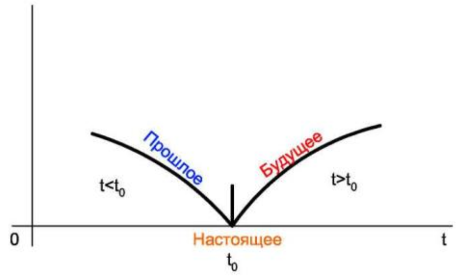


Рисунок 3 – Схема марковского свойства случайного процесса

«Настоящее» может быть задано не одним каким-то состоянием s_i, а целым подмножеством состояний $V \subset W$, где W - множество всех возможных состояний системы.

Подчеркнем также, что «настоящее» может быть задано не только одним состоянием системы S в момент t_0; в него при желании можно включить и те элементы из «прошлого», от которых, при заданном «настоящем», зависит будущее. Например, вероятности состояний в «будущем» могут зависеть не только от состояния s_i системы в настоящем, но и от того, из какого состояния s_i система перешла к моменту t_0 в состояние s_i; в этом случае настоящее характеризуется не только состоянием s_i , в которое система перешла к моменту t_0, но и состоянием s_j, из которого она перешла в s_i. Вводя в состав параметров, характеризующих настоящее состояние системы, те параметры из прошлого, от которых зависит будущее, можно, как говорится, «марковизировать» многие немарковские случайные процессы, но, как правило, это приводит к сильному усложнению математического аппарата.

Марковские случайные процессы с дискретными состояниями и дискретным временем

Пусть имеется система S с дискретными состояниями $s_1, s_2, ..., s_i, ..., s_n$. Предположим, что случайные переходы («перескоки») системы из состояния в состояние могут происходить только в определенные моменты времени $t_0, t_1, t_2, ...$. Эти моменты мы будем называть шагами процесса; $t_0 = 0$ - его началом. Сам процесс представляет собой случайное блуждание системы S по состояниям. После первого шага система может оказаться в одном (и только в одном) из своих возможных состояний: $s_1^{(1)}, s_2^{(1)}, ..., s_i^{(1)}, ..., s_n^{(1)}$; на втором шаге - $s_1^{(2)}, s_2^{(2)}, ..., s_i^{(2)}, ..., s_n^{(2)}$, на k-м шаге $s_1^{(k)}, s_2^{(k)}, ..., s_i^{(k)}, ..., s_n^{(k)}$ (число состояний в общем случае может быть бесконечным, но счетным. Здесь же для простоты ограничимся конечным числом n состояний).

Предположим, что граф состояний системы S имеет вид, представленный на рисунке 4. Процесс блуждания системы S по состояниям можно представить как последовательность или «цепь» событий, состоящих в том, что в начальный момент

$t_0=0$ система находится в одном из состояний (например, в состоянии $s_1^{(0)}$), в момент первого шага перешла из него скачком в состояние $s_5^{(1)}$, из которого на втором шаге перешла в $s_3^{(2)}$, на третьем шаге перешла в $s_2^{(3)}$ и т. д. «Траектория» системы, блуждающей по состояниям s_1, s_5, s_3, s_2 показана на рисунке 4 жирными линиями. На каких-то шагах система может задерживаться в том или другом из своих состояний, $s_i^{(k)} = s_i^{(k+1)}$ (это показано «возвратной стрелкой» на рисунке 4) или же вернуться в него после ряда шагов.

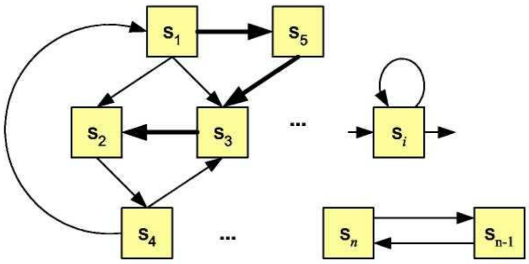


Рисунок 4 – Граф состояний системы S

«Траектория» блуждания системы по графу состояний, изображенная на рисунке 4 жирными линиями, представляет собой не что иное, как реализацию случайного процесса, полученную в результате одного опыта. При повторении опыта, естественно, реализации в общем случае не совпадают.

Рассмотрим общий случай. Пусть происходит случайный процесс в системе *S* с дискретными состояниями $s_1, s_2, \ldots, s_i, \ldots, s_n$, которые она может принимать в последовательности шагов с номерами *0, 1, 2, …, k, …*.

Случайный процесс представляет собой последовательность событий вида $\{S(k) = s_i\}$ $(i = 1,2,\ldots \ldots n;\ k = 0,1,2,\ldots)$. Наиболее важной ее характеристикой являются вероятности состояний системы

$$\mathbf{P}\{S(k)= s_i\} \qquad (i = 1,2,\ldots\ldots n;\ k = 0,1,2,\ldots), \tag{5}$$

где $P\{S(k)=s_i\}$ - вероятность того, что на *k*-м шаге система *S* будет находиться в состоянии s_i.

Распределение вероятностей (5) представляет собой не что иное, как одномерный закон распределения случайного процесса *S(t)*, протекающего в системе *S* с «качественными» дискретными состояниями и дискретным временем $t_0, t_1, t_2, \ldots, t_k$.

Процесс, протекающий в такой системе *S*, называется марковским процессом с дискретными состояниями и дискретным временем (или, короче, марковской цепью), если выполняется условие: для любого фиксированного момента времени (любого шага k_0) условные вероятности состояний системы в будущем (при $k > k_0$) зависят только от состояния системы в настоящем (при $k = k_0$) и не зависят от того, когда (на каком шаге, при $k < k_0$) и откуда система пришла в это состояние. Марковская цепь представляет собой разновидность марковского процесса, в котором будущее зависит от прошлого только через настоящее.

Цепь, в которой условные вероятности состояний в будущем зависят только от состояния на данном, последнем, шаге и не зависят от предыдущих, иногда называют простой цепью Маркова, в отличие от такой, где будущее зависит от состояний системы не только в настоящем на данном шаге, но и от ее состояний на нескольких предыдущих шагах; такую цепь называют сложной цепью Маркова. Сам А. А. Марков рассматривал сложные цепи, построенные на материале буквенных последовательностей, взятых из текста пушкинского «Евгения Онегина».

Если в качестве системы, в которой происходит случайный процесс, рассмотреть букву, входящую в текст, которой могут быть: а, б, в, ... щ, ъ, ы, ь, э, ю, я, «пробел», то сразу ясно, что вероятность последующей буквы быть той или другой зависит от того, какова была предыдущая (например, последовательности букв «яы» или «эь» в русском языке исключены); не так очевидно, но все же ясно, что эта вероятность зависит не только от предыдущей буквы, но и от других, ей предшествовавших (например, последовательность букв «ттт» в русском языке если не исключена, то практически невозможна, тогда как последовательность «тт» встречается довольно часто). Мы в данном элементарном изложении будем рассматривать только простые цепи Маркова и вычислять для них вероятности состояний.

Из определения марковской цепи следует, что для нее вероятность перехода системы S в состояние s_i на $(k+1)$-м шаге зависит только от того, в каком состоянии s_i находилась система на предыдущем k-м шаге и не зависит от того, как она вела себя до этого k-го шага.

Основной задачей исследования марковской цепи является нахождение безусловных вероятностей нахождения системы S на любом k-м шаге в состоянии s_i; обозначим эту вероятность $p_i(k)$:

$$p_i(k) = P\{S(k) = s_i\} \ (i = 1,2,......n; k = 0,1,2,...). \qquad (6)$$

Для нахождения этих вероятностей необходимо знать условные вероятности перехода системы S на k-м шаге в состояние s_i, если известно, что на предыдущем $(k - 1)$-м шаге она была в состоянии s_i. Обозначим эту вероятность

$$p_{ij}(k) = P\{S(k) = s_i \mid S(k-1) = s_j\} \quad (i, j = 1,2,......n; k = 0,1,2,...). \qquad (7)$$

Вероятности $p_{ij}(k)$ называются переходными вероятностями марковской цепи на k-м шаге. Вероятность $p_{ij}(k)$ есть вероятность того, что на k-м шаге система задержится (останется) в состоянии s_i.

Переходные вероятности $p_{ij}(k)$ можно записать в виде квадратной таблицы (матрицы) размерности n x n:

$$\|p_{ij}(k)\| = \left\| \begin{matrix} p_{11}(k) & p_{12}(k) \ ... \ p_{1j}(k) \ ... \ p_{1n}(k) \\ p_{21}(k) & p_{22}(k) \ ... \ p_{2j}(k) \ ... \ p_{2n}(k) \\ \multicolumn{2}{l}{\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ p_{i1}(k) & p_{i2}(k) \ ... \ p_{ij}(k) \ ... \ p_{in}(k) \\ \multicolumn{2}{l}{\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ p_{n1}(k) & p_{n2}(k) \ ... \ p_{nj}(k) \ ... \ p_{nn}(k) \end{matrix} \right\| \quad (k = 0,1,2,...). \qquad (8)$$

По главной диагонали матрицы (8) стоят вероятности задержки системы в данном состоянии s_j $(j = 1,..., n)$ на k-м шаге.

$$p_{11}(k), p_{22}(k),..., p_{ii}(k),..., p_{nn}(k). \qquad (9)$$

Так как на каждом шаге система S может находиться только в одном из взаимно исключающих состояний, то для любой k-й строки матрицы (8) сумма всех стоящих в ней вероятностей равна единице:

$$\sum_{j=1}^{n} p_{ij}(k) = 1. \qquad (10)$$

Матрица, обладающая таким свойством, называется стохастической. Естественно, что все элементы стохастической матрицы отвечают условию $0 \le p_{ij}(k) \le 1$. В силу условия (10) можно в матрице (8) не задавать вероятности задержки, а получать их как дополнения до единицы всех остальных членов строки:

$$p_{ii}(k) = 1 - \sum_{j=1}^{n} p_{ij}(k). \qquad (11)$$

Чтобы найти безусловные вероятности $p_i(k)$, недостаточно знать матрицу переходных вероятностей (8); нужно еще знать начальное распределение вероятностей, т. е. вероятности состояний $p_i(0)$, соответствующие началу процесса - моменту $t_0 = 0$:

$$p_1(0), p_2(0), \ldots, p_i(0), \ldots, p_n(0), \qquad (12)$$

в сумме образующие единицу:

$$\sum_{i=1}^{n} p_i(0) = 1 \qquad (13)$$

Если известно, что в начальный момент система S находится во вполне определенном состоянии s_i, то вероятность $p_i(0)$ этого состояния в формуле (13) равна единице, а все остальные - нулю:

$$p_i(0), p_1(0) = p_2(0) = \ldots = p_{i-1}(0) = p_{i+1}(0) = \ldots = p_i(0) = 0. \qquad (14)$$

Цепь Маркова называется однородной, если переходные вероятности $p_{ij}(k)$ не зависят от номера шага k: $p_{ij}(k) = p_{ij}$. Матрица переходных вероятностей для однородной цепи Маркова имеет вид:

$$\|p_{ij}\| = \left\| \begin{matrix} p_{11} \quad p_{12} \cdots p_{1j} \cdots p_{1n} \\ p_{21} \quad p_{22} \cdots p_{2j} \cdots p_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p_{i1} \quad p_{i2} \cdots p_{ij} \cdots p_{in} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p_{n1} \quad p_{n2} \cdots p_{nj} \cdots p_{nn} \end{matrix} \right\| \qquad (14)$$

При выводе формул для вероятностей состояний, в целях простоты записи, будем рассматривать только однородные цепи Маркова (в случае, когда цепь неоднородна, можно все переходные вероятности в формулах просто положить зависящими от номера шага k).

При нахождении вероятностей состояний марковской цепи на k-м шаге $p_i(k)$ $(k = 1,2,...)$ удобно бывает пользоваться так называемым размеченным графом состояний системы S, где возле каждой стрелки, ведущей из состояния s_i в состояние s_j, проставлена переходная вероятность p_{ij}; вероятности задержки на размеченном графе не проставляются, а просто получаются дополнением до единицы суммы вероятностей, стоящих у всех стрелок, ведущих из данного состояния s_i.

Теперь покажем, как найти для однородной цепи Маркова безусловную вероятность нахождения системы S на k-м шаге в состоянии s_j ($j=1, 2, ..., n$)

$$p_j(k) = \mathbf{P}\{S(k) = s_j\}, \qquad (15)$$

если задана матрица переходных вероятностей $||p_{ij}||$ (или, что равнозначно, размеченный граф состояний) и начальное распределение вероятностей

$$p_i(0) \quad (i = 1,2,...,n).$$

$$\sum_{j=1}^{n} p_{ij}(k) = 1 \qquad . \qquad (16)$$

Сделаем гипотезу, состоящую в том, что в начальный момент ($k=0$) система находилась в состоянии s_i. Вероятность этой гипотезы известна из (16) и равна $p_i(0)=P\{S(0)=s_i)$. В предположении, что эта гипотеза имеет место, условная вероятность того, что система S на первом шаге будет в состоянии s_j, равна переходной вероятности $p_{ij}(k) = \mathbf{P}\{S(1) = s_j \mid S(0) = s_i\}$.

По формуле полной вероятности получим:

$$p_j(1) = \sum_{i=1}^{n} P\{S(1) = s_j | S(0) = s_i\} \cdot P\{S(0) = s_i\} = \sum_{i=1}^{n} p_{ij}\, p_i(0), \quad (j = 1,...,n) \qquad (17)$$

Таким образом, мы нашли распределение вероятностей системы S на первом шаге. Теперь у нас есть все необходимое для того, чтобы найти распределение вероятностей на втором шаге, которое для цепи Маркова зависит только от распределения вероятностей на первом шаге и матрицы переходных вероятностей.

Опять сделаем гипотезу, состоящую в том, что на первом шаге система находится в состоянии s_i вероятность этой гипотезы нам уже известна и равна $p_i(1) = P\{S(1) = s_i\}$. При этой гипотезе условная вероятность того, что на втором шаге система S будет в состоянии s_i, равна:

$$p_{ij}(k) = P\{S(2) = s_j | S(1) = s_i\}$$

По формуле полной вероятности находим

$$p_j(2)=\sum_{i=1}^{n} p_i(1)p_{ij},(j=1,2,...,n) \qquad (18)$$

Таким образом, мы выразили распределение вероятностей (18) на втором шаге через распределение вероятностей на первом шаге и матрицу $\|p_{ij}\|$. Переходя таким же способом от $k = 2$ к $k = 3$ и т. д., получим рекуррентную формулу:

$$p_j(k)=\sum_{i=1}^{n} p_i(k-1)p_{ij},(k=1,2,...,n;j=1,2,...,n) \qquad (19)$$

При некоторых условиях в цепи Маркова с возрастанием k (номера шага) устанавливается стационарный режим, в котором система S продолжает блуждать по состояниям, но вероятности этих состояний уже от номера шага не зависят. Такие вероятности называются предельными (или финальными) вероятностями цепи Маркова.

Например, если рассматривать ЭВМ в двух состояниях: s_1 - исправна, s_2 - не исправна, то имеет место следующая динамика изменения вероятностей (при начальных условиях):

$$p_1(0)=1,p_2(0)=0:\ p_1(1)=0,7;p_1(2)=0,61;p_1(3)=0,583;p_1(4)=0,5749\ .$$

Ниже мы покажем, что в этом случае $p_1=\lim_{k\to\infty} p_1(k)=0,4/(0,4+0,3)=0,5714$. Таким образом, в рассматриваемой системе стационарный режим наступит практически через четыре шага.

Можно убедиться в том, что в этом примере финальные вероятности не зависят от начальных условий.

Сформулируем условия существования стационарного режима для системы S с конечным числом состояний n, в которой протекает марковский случайный процесс с дискретными состояниями и дискретным временем (цепь Маркова):

1. Множество всех состояний W системы S должно быть эргодическим.

2. Цепь Маркова должна быть однородной:

$$p_{ij}(k)=p_{ij} \qquad (20)$$

3. Цепь Маркова должна быть «достаточно хорошо перемешиваемой» (не должна быть «циклической»).

Цепи Маркова, отвечающие этим условиям, будем называть эргодическими цепями Маркова.

3. Основные понятия теории систем массового обслуживания.

Марковский случайный процесс с дискретными состояниями и непрерывным временем характерен для ***систем массового обслуживания*** (СМО).

Понятие систем массового обслуживания

При решении задач оптимизации управления производством, информационными сетями, транспортными системами часто возникает ряд однотипных задач:

- оценка пропускной способности каналов связи, системы автомобильных и железных дорог и т. п.;
- оценка эффективности работы предприятия, компьютерной сети;
- определение количества кналов связи и транспортных путей сообщения и др.

Все эти задачи однотипны в том смысле, что в них присутствует массовый спрос на обслуживание. В удовлетворении этого спроса участвует определенная совокупность элементов, образующая ***системумассового обслуживания*** **(СМО)** (рис. 1).

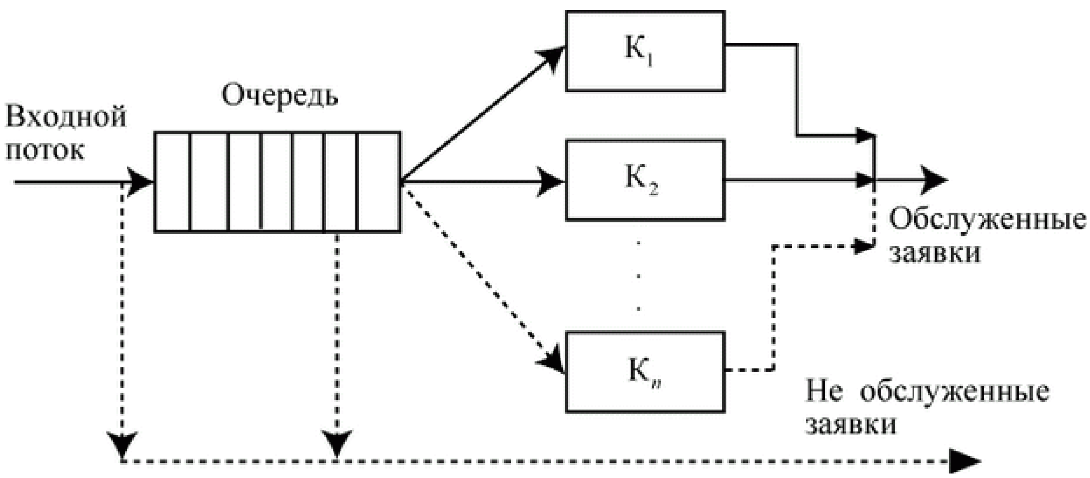


Рис. 1. Система массового обслуживания

Элементами СМО являются:

- входной (входящий) поток требований (заявок) на обслуживание;
- приборы (каналы) обслуживания;
- очередь заявок, ожидающих обслуживания;
- выходной (выходящий) поток обслуженных заявок;
- поток не обслуженных заявок;
- очередь свободных каналов (для многоканальных СМО).

Входящий поток - это совокупность заявок на обслуживание. Часто заявка отождествляется с ее носителем. Например, поток неисправной радиоаппаратуры, поступающий в мастерскую объединения, и представляет собой поток заявок - требований на обслуживание в данной СМО.

Как правило, на практике имеют дело с так называемыми рекуррентными потоками, потоками, обладающими свойствами:

- стационарности;
- ординарности;
- ограниченного последействия.

Первые два свойства мы определили ранее. Что касается ограниченного последействия, то оно заключается в том, что интервалы между поступающими заявками являются независимыми случайными величинами.

Рекуррентных потоков много. Каждый закон распределения интервалов порождает свой рекуррентный поток. Рекуррентные потоки иначе называют ***потоками Пальма***.

Простейший стационарный поток - пуассоновский поток с полным отсутствием последействия. У него случайные интервалы между заявками имеют экспоненциальное распределение:

$F(t) = 1 - e^{-\lambda t}, f(t) = \lambda e^{-\lambda t},$

здесь λ - интенсивность потока.

Название потока - пуассоновский - происходит от того, что для этого потока вероятность $P_k(\Delta t)$ появления k заявок за интервал Δt определяется законом Пуассона:

$$P_k(\Delta t) = \frac{(\lambda \cdot \Delta t)^k}{k!} e^{-\lambda \cdot \Delta t}$$

Именно такой поток предполагают проектировщики при разработке СМО. Вызвано это тремя причинами.

Во-первых, поток этого типа в теории массового обслуживания аналогичен нормальному закону распределения в теории вероятностей в том смысле, что к простейшему потоку приводит предельный переход для потока, являющегося суммой потоков с произвольными характеристиками при бесконечном увеличении слагаемых и уменьшении их интенсивности. То есть сумма произвольных независимых (без преобладания) потоков с интенсивностями λ_i является простейшим потоком с интенсивностью $\lambda = \sum_i \lambda_i$

Во-вторых, если обслуживающие каналы (приборы) рассчитаны на простейший поток заявок, то обслуживание других типов потоков (с той же интенсивностью) будет обеспечено с не меньшей эффективностью.

В-третьих, именно такой поток определяет марковский процесс в системе и, следовательно, простоту математического анализа системы. При других потоках анализ функционирования СМО сложен.

Часто встречаются системы, у которых поток входных заявок зависит от количества заявок, находящихся в обслуживании. Такие СМО называют **замкнутыми** (иначе - **разомкнутыми**). Например, работа мастерской связи объединения может быть представлена моделью замкнутой СМО. Пусть эта мастерская предназначена для обслуживания радиостанций, которых в объединении m. Каждая из них имеет интенсивность отказов λ. Входной поток отказавшей аппаратуры будет иметь интенсивность λ_p:

$$\lambda_p = \lambda(m - n),$$

где n - количество радиостанций, уже находящихся в мастерской на ремонте.

Заявки могут иметь разные права на начало обслуживания. В этом случае говорят, что заявки **неоднородные**. Преимущества одних потоков заявок перед другими задаются шкалой приоритетов.

Важной характеристикой входного потока является **коэффициент вариации**: $\nu = \frac{\sigma}{\overline{\tau}_{\text{инт}}}$, где $\overline{\tau}_{\text{инт}}$ - математическое ожидание длины интервала; σ - среднеквадратическое отклонение случайной величины (длины интервала) $\tau_{\text{инт}}$.

Для простейшего потока $\left(\sigma = \frac{1}{\lambda}, \tau_{\text{инт}} = \frac{1}{\lambda}\right) : \nu = 1$

Для большинства реальных потоков $0 \le \nu \le 1$.

При $\nu = 0$ поток регулярный, детерминированный.

Коэффициент вариации - характеристика, отражающая степень неравномерности поступления заявок.

Каналы (приборы) обслуживания. В СМО могут быть один или несколько обслуживающих приборов (каналов). Согласно с этим СМО называют одноканальными или многоканальными.

Многоканальные СМО могут состоять из однотипных или разнотипных приборов. Обслуживающими приборами могут быть:

- линии связи;
- мастера ремонтных мастерских, продавцы, кассиры;
- маршрутизаторы в компьютерных сетях;
- транспортные средства;
- платежные терминалы;
- серверы, и др.

Основная характеристика канала - время обслуживания. Как правило, время обслуживания - величина случайная.

Обычно практики полагают, что время обслуживания имеет экспоненциальный закон распределения:

$F(t) = 1 - e^{-\mu t}, f(t) = e^{-\mu t}$, где μ - интенсивность обслуживания, $\mu = \frac{1}{\overline{\tau}_{\text{обсл}}}$;

$\overline{\tau}_{\text{обсл}}$ - математическое ожидание времени обслуживания.

То есть процесс обслуживания - марковский, а это, как теперь нам известно, дает существенные удобства в численно-математическом моделировании.

Кроме экспоненциального встречаются k -распределение Эрланга, гиперэкспоненциальное, треугольное и некоторые другие. Это нас не должно смущать, так как показано, что значение критериев эффективности СМО мало зависят от вида закона распределения вероятностей времени обслуживания.

При исследовании СМО выпадает из рассмотрения сущность обслуживания, качество обслуживания.

Каналы могут быть **абсолютно надежными**, то есть не выходить из строя. Вернее, так может быть принято при исследовании. Каналы могут обладать **конечной надежностью**. В этом случае модель СМО значительно сложнее.

Очередь заявок. В силу случайного характера потоков заявок и обслуживания пришедшая заявка может застать канал (каналы) занятым обслуживанием предыдущей заявки. В этом случае она либо покинет СМО не обслуженной, либо останется в системе, ожидая начало своего обслуживания. В соответствии с этим различают:

- СМО с отказами;

- СМО с ожиданием.

СМО с ожиданием - характеризуются наличием очередей. Очередь может иметь ограниченную или неограниченную емкость: $1 \leq L \prec \infty.$

Исследователя обычно интересуют такие статистические характеристики, связанные с пребыванием заявок в очереди:

- среднее количество заявок в очереди за интервал исследования;
- среднее время пребывания (ожидания) заявки в очереди. **СМО с ограниченной емкостью очереди** относят к СМО смешанного типа.

СМО смешанного типа - такие СМО, в которых заявки имеют ограниченное время пребывания в очереди независимо от ее емкости.

Выходящий поток - это поток обслуженных заявок, покидающих СМО.

Встречаются случаи, когда заявки проходят через несколько СМО: транзитная связь, производственный конвейер и т. п. В этом случае выходящий поток является входящим для следующей СМО.

Многофазные СМО, сети СМО - совокупность последовательно связанных между собой СМО

Входящий поток первой СМО, пройдя через последующие СМО, искажается и это затрудняет моделирование. Однако, следует иметь в виду, что при простейшем входном потоке и экспоненциальном обслуживании (то есть в марковских системах) выходной поток тоже простейший. Если время обслуживания имеет не экспоненциальное распределение, то выходящий поток не только не простейший, но и не рекуррентный.

Заметим, что интервалы между заявками выходящего потока, это не то же самое, что интервалы обслуживания. Ведь может оказаться, что после окончания очередного обслуживания СМО какое-то время простаивает из-за отсутствия заявок. В этом случае интервал выходящего потока состоит из времени незанятости СМО и интервала обслуживания первой, пришедшей после простоя, заявки.

В системах с отказами есть **поток необслуженных заявок**. Если в СМО с отказами поступает рекуррентный поток, а обслуживание - экспоненциальное, то и поток необслуженных заявок - рекуррентный.

Очереди свободных каналов

В многоканальных СМО могут образовываться очереди свободных каналов. Количество свободных каналов - величина случайная. Исследователя могут интересовать различные характеристики этой случайной величины. Обычно это среднее число каналов, занятых обслуживанием за интервал исследования.

Таким образом, по признакам, влияющим на функционирование, СМО может принадлежать к одному из типов в соответствии с приводимой классификацией (рис. 2).

Рис. 2 Классификация СМО

Для обозначения простых (однофазных) СМО используется символика, предложенная Кендаллом:

$$A/B/n/m.$$

A - входящий поток заявок: $A = GI$ - рекуррентный поток; $A = \mathrm{M}$ - простейший поток с показательным законом распределения вероятностей; $A = D$ - регулярный или детерминированный поток (с постоянными интервалами между моментами поступления заявок).

B - случайная длительность обслуживания: $B = G$ или $B = GI$ - рекуррентное обслуживание с одной и той же функцией распределения $B(t)$ для разных каналов; $B = \mathrm{M}$ - показательное обслуживание; $B = D$ - регулярное обслуживание.

n - количество обслуживающих каналов. Если $n > 1$, то система называется многоканальной.

m - количество мест для ожидания заявок в очереди. Если $m = 0$, то СМО с потерями (без ожидания); $m = \infty$ - система с неограниченным ожиданием; $0 < m < \infty$ - система с ограниченным числом мест для ожидания.

МАТЕРИАЛЫ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1 Практическое занятие № 1 (2часа).

Тема: Общие сведения о математическом моделировании.

2.1.1 Задание для работы:

Обсудить основные понятия теории математического моделирования, провести их анализ.

2.1.2 Краткое описание проводимого занятия:

Вопросы для обсуждения:

1. Понятие модели, ее эффективность, адекватность

2.Классификация моделей

3. Основные характеристики и параметры моделей

2.1.3 Результаты и выводы:
В результате проведенного занятия студенты:
- ознакомились с основными понятиями математического моделирования;
- ознакомились с классификацией моделей по различным основаниям;
- проанализировали виды моделей, рассмотрели основные параметры

2.2 Практическое занятие № 2-3 (4 часа).

Тема: Общая модель линейного программирования в каноническом и неканоническом представлении. Графический метод решения ЗЛП.

2.2.1 Задание для работы:
Ознакомится с основными моделями линейного программирования, научиться применять к решению ЗЛП графический метод.

2.2.2 Краткое описание проводимого занятия:
Вопросы для обсуждения:
1. Модели математического программирования
2.ЗЛП, их модели
3. Графический метод решения ЗЛП.

2.2.3 Результаты и выводы:
В результате проведенного занятия студенты:
- ознакомились с основными понятиями математического программирования;
- ознакомились с классификацией моделей линейного программирования;
- научились применять графический метод к решению ЗЛП

2.3 Практическое занятие № 4-5 (4 часа).

Тема: Симплекс-метод. Двойственные задачи линейного программирования.
.
2.3.1 Задание для работы:
Ознакомится с основными принципами и алгоритмами симплекс- метода, научиться применять теоремы двойственности к решению ЗЛП.

2.3.2 Краткое описание проводимого занятия:
Вопросы для обсуждения:
1. Алгоритм симплекс метода
2.Способы реализации (методы решения) симплекс-таблиц
3. Применение двойственности к решению и интерпретации ЗЛП

2.3.3 Результаты и выводы:
В результате проведенного занятия студенты:
- ознакомились с основным алгоритмом симплекс метода;
- ознакомились с методами работы с симплекс-таблицами;
- научились применять теоремы двойственности.

2.4 Практическое занятие № 6-8 (6 часа).

Тема: Постановка задач линейного программирования транспортного типа. Базовая транспортная модель. Методы решения задач транспортного типа

2.4.1 Задание для работы:
Ознакомится с основными принципами и алгоритмами решения транспортной задачи.

2.4.2 Краткое описание проводимого занятия:
Вопросы для обсуждения:
1. Постановка транспортной задачи
2.Первичное распределение в ТЗ
3. Методы решения ТЗ

2.4.3 Результаты и выводы:
В результате проведенного занятия студенты:
- ознакомились с основными методами первичного распределения поставок в транспортной задаче;
- научились работать с циклами;
- научились применять методы оптимизации распределения поставок.