МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ДЕПАРТАМЕНТ НАУЧНО-ТЕХНОЛОГИЧЕСКОЙ ПОЛИТИКИ И ОБРАЗОВАНИЯ
ФЕДЕРАЛЬОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
Азово-Черноморская государственная агроинженерная академия

Математика: исследование операций

Учебное пособие

Д.В. Степовой, Л.В. Кравченко

Зерноград 2011

УДК 519.8

Исследование операций. Учебное пособие. – Зерноград: ФГОУ ВПО АЧГАА, 2011. – 116с.

Учебное пособие содержит в себе основные методы исследования операций и написано в соответствии с новым федеральным государственным образовательным стандартом высшего профессионального образования третьего поколения. Пособие предназначено для студентов очного и заочного форм обучения по направлению подготовки 190700 «Технология транспортных процессов», получающих степень бакалавра.

Рецензенты:

кафедра высшей математики ТТИ ЮФУ (зав.кафедрой к.ф.-м.н., доцент А.В. Никитина),

проректор по информатизации и электронному обучению ЮФУ, д.ф.-м.н. А.Н. Карапетянц.

Содержание

Введение

Деятельность отдельных людей и коллективов, как правило, связана с принятием таких решений, которые позволили бы получить некие оптимальные результаты: достичь максимальной прибыли предприятия, затратить минимум средств на кормление скота, получить максимальный урожай, рассчитать оптимальный поток обслуживаемых автомобилей на станции технического обслуживания и т.п. Но в каждой конкретной ситуации приходится считаться с реальными условиями задачи. Предприятие не может получить максимальной прибыли без учета конъюнктуры рынка, стоимости сырья, рабочей силы и целого ряда других факторов. В процессе решения такого рода задач появились методы их решения, которые в совокупности объединяются под названием «Исследование операций».

Под операцией понимают совокупность действий, направленных на достижение определенных целей.

Исследование операций включает в себя следующие методы (см. Рис.1):

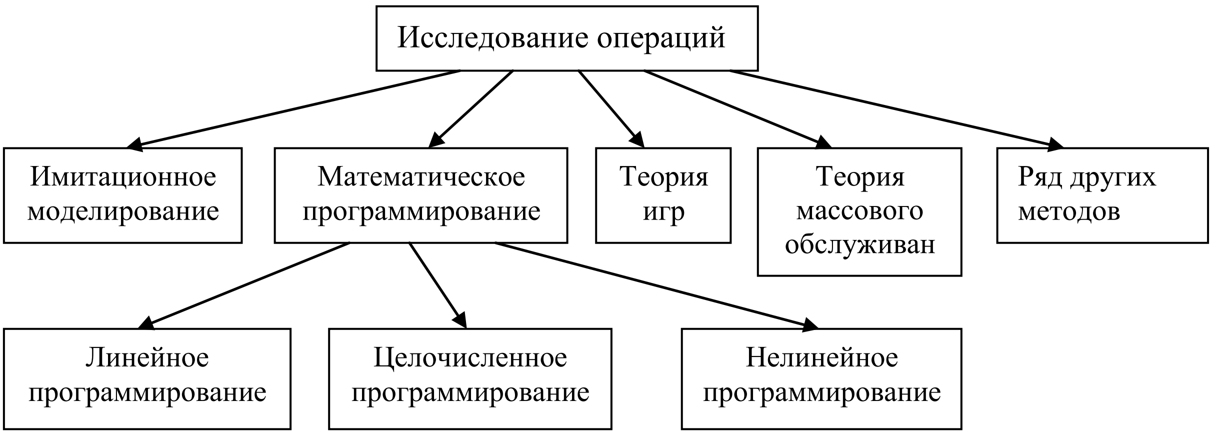


Рис. 1. Методы исследования операций

Основным методам исследования операций и посвящено данное учебное пособие. Для желающих изучить соответствующие разделы более глубоко, мы рекомендуем обратиться к литературе, приводимой в конце учебного пособия.

ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Общая задача линейного программирования

Задачи математического и линейного программирования

Исследование различных процессов начинается с их моделирования, т.е. отражения реального процесса через математические соотношения. При этом составляются уравнения или неравенства, которые связывают различные показатели (переменные) исследуемого процесса, образуя систему ограничений. В этих соотношениях выделяются такие переменные, меняя которые можно получит оптимальное значение основного показателя данной системы. Соответствующие методы, позволяющие решать такие задачи, объединяются под общим названием «математическое программирование» или «математические методы исследования операций».

Математическое программирование включает в себя следующие разделы математики: линейное программирование, нелинейное программирование, целочисленное программирование, динамическое программирования.

Математическое программирование – это раздел высшей математики, посвященный решению задач, связанных с нахождением экстремумов функций нескольких переменных при наличии ограничений на переменные.

Методами математического программирования решаются задачи о распределении ресурсов, планировании выпуска продукции, ценообразовании, транспортные задачи и т.д.

Математическое программирование возникло в 30 – годы XX века. Венгерский математик Б.Эгервари в 1931 решил задачу, называемую «задачей проблемы выбора». Американский ученый Г.У. Кун обобщил этот метод, после чего он стал называться «венгерским» методом. В 1939 году российский ученый Л.В. Канторович разработал метод разрешающих множителей решения задач линейного программирования. Большой вклад в развитие математического программирования внесли американские ученые. В 1947 году американский ученый Дж.Данциг описал один из основных методов решения задач линейного программирования, получивший название «симплексный».

Построение математической модели задачи включает следующие этапы: 1) выбор переменных задачи; 2) составление системы ограничений; 3) выбор целевой функции.

Переменными задачи называются величины $x_1, x_2,...,x_n$, которые полностью характеризуют процесс. Их обычно записывают в виде вектора

$X=(x_1, x_2,...,x_m)$.

Система ограничений – это система линейных уравнений и неравенств, которым удовлетворяют переменные задачи и которые следуют из ограниченности ресурсов или физических переменных, например положительности переменных.

Целевой функцией называют функцию переменных задачи, которая характеризует качество выполненной задачи, и экстремум которой требуется найти.

Общая задача математического программирования формулируется следующим образом: найти экстремум целевой функции

$$Z(X)=f(x_1,x_2,...,x_n)\to\max\ (\min) \tag{1.1}$$

при условии, что эти переменные удовлетворяют системе ограничений

$$\begin{cases} \varphi_i(x_1,x_2,...,x_n)=0, & i=1,2,...,l, \\ \varphi_i(x_1,x_2,...,x_n)>(<)0, & i=l+1,l+2,...,m \end{cases} \tag{1.2}$$

Если целевая функция (1.1) и система ограничений (1.2) линейны, то задача математического программирования называется задачей линейного программирования.

Итак, задача линейного программирования будет поставлена, если:

1) указана целевая функция;

2) записана система ограничений (часто в системе ограничений особо выделить неравенства вида $x_j \geq 0$);

3) определено, к какому типу (максимизации или минимизации) принадлежит данная задача. Задачу максимизации всегда можно свести к задаче минимизации, поменяв знаки у коэффициентов целевой функции.

В общем случае задача линейного программирования может быть записана в таком виде:

$$Z(X)=c_1x_1+c_2x_2+\ldots+c_nx_n\to\max(\min), \tag{1.3}$$

$$\begin{cases} a_{11}x_1+a_{12}x_2+...+a_{1n}x_n=b_1, \\ .... \\ a_{l1}x_1+a_{l2}x_2+...+a_{ln}x_n=b_l, \\ a_{(l+1)1}x_1+a_{(l+1)2}x_2+...+a_{(l+1)n}x_n\leq b_{l+1}, \\ .... \\ a_{m1}x_1+a_{m2}x_2+...+a_{mn}x_n\leq b_m, \end{cases} \tag{1.4}$$

$$x_j\geq 0,\ j=1,2,\ldots,t,\quad t\leq n \tag{1.5}$$

Данная запись означает следующее: найти экстремум целевой функции (1.3) переменных $X=(x_1,\ x_2,...,x_m)$ при условии, что эти переменные удовлетворяют системе ограничений (1.4) и условиям неотрицательности (1.5).

Допустимым решением (планом) задачи линейного программирования называется любой n – мерный вектор $X=(x_1,\ x_2,...,x_m)$, удовлетворяющий системе ограничений и условиям неотрицательности его координат.

Множество допустимых решений (планов) задачи образует область допустимых решений (ОДР).

Оптимальным решением (планом) задачи линейного программирования называется такое допустимое решение (план) задачи, при котором целевая функция достигает экстремума.

Свойства решений задач линейного программирования. Элементы геометрии выпуклых множеств

Множество A называется подмножеством B, если все элементы множества A являются одновременно элементами множества B.

Множество точек называется *выпуклым*, если оно вместе с любыми двумя своими точками содержит весь отрезок, соединяющий эти точки.

Примерами выпуклых множеств являются круг, сектор, многоугольники и т.д.

Пересечение (общая часть) любого числа выпуклых множеств есть выпуклое множество.

Среди точек выпуклого множества можно выделить внутренние, граничные и угловые точки.

Точка множества называется *внутренней,* если в некоторой ее окрестности содержатся только точки данного множества.

Точка множества называется *граничной*, если в любой ее окрестности содержатся как точки, принадлежащие данному множеству, так и точки, не принадлежащие ему.

Точка множества называется *угловой* (*или крайней*), если она не является внутренней ни для какого отрезка, целиком принадлежащего данному множеству.

На рисунке 1.1 приведены примеры различных точек: внутренней (точка M), граничной (точка N) и угловых (точки A, B, C, D, E). Точка A – угловая, так как для любого отрезка, целиком принадлежащего многоугольнику (например, отрезка AP), она не является внутренней; точка A – внутренняя для отрезка KL, но этот отрезок не принадлежит целиком многоугольнику.

Для выпуклого множества угловые точки всегда совпадают с вершинами многоугольника (многогранника), в то же время для невыпуклого множества это не обязательно. Множество может иметь любое число угловых точек: одну, две и т.д., а также бесконечное число угловых точек.

Множество точек называется *замкнутым*, если включает все свои граничные точки. Множество точек называется *ограниченным*, если существует шар (круг) радиуса конечной длины с центром в любой точке множества, который полностью содержит в себе данное множество; в противном случае множество называется *неограниченным*.

Выпуклое замкнутое множество точек пространства (плоскости), имеющее конечное число угловых точек, называется *выпуклым* *многогранником (многоугольником)*, если оно ограниченное, и *выпуклой многогранной (многоугольной) областью*, если оно неограниченное.

Рассмотрим *геометрический смысл* решения неравенств, уравнений и их систем.

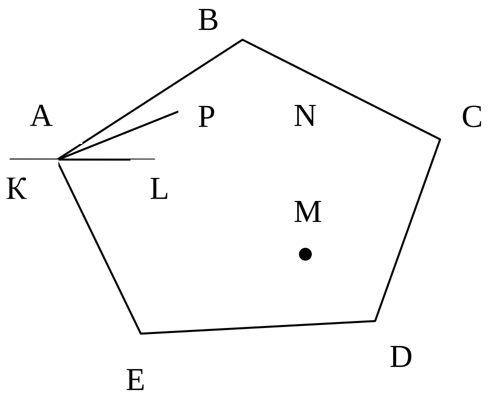


Рис. 1.1. Точки многогранника

Множество решений неравенства с двумя переменными

$$a_{11}x_1 + a_{12}x_2 \le b_1 \qquad (1.6)$$

является одной из двух полуплоскостей, на которые вся плоскость делится прямой $a_{11}x_1 + a_{12}x_2 = b_1$, включая и эту прямую, а другая полуплоскость с той же прямой есть множество решений неравенства

$$a_{11}x_1 + a_{12}x_2 \ge b_1 \qquad (1.7)$$

Множество всех допустимых решений совместной системы m линейных с n переменными $(m < n)$ является выпуклым многогранником (выпуклой многогранной областью) в n – мерном пространстве.

Обобщением понятия *отрезка* для нескольких точек является их выпуклая линейная комбинация.

Точка X называется *выпуклой линейной комбинацией точек* $X_1, X_2 ... X_n$, если выполняются условия:

$$X = \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_n X_n \qquad (1.8)$$

$$\alpha_j \ge 0 \quad (j = 1,2..n), \sum_{j=1}^{n} \alpha_j = 1$$

Очевидно, что при $n = 2$ выпуклой линейной комбинацией двух точек является соединяющий их отрезок. Поэтому *множество точек является **выпуклым**, если оно вместе с любыми своими двумя точками содержит их произвольную выпуклую комбинацию.*

Выпуклый n - мерный многогранник является выпуклой линейной комбинацией своих угловых точек.

Из утверждения следует, что выпуклый многогранник порождается своими угловыми точками или вершинами: отрезок – двумя точками, треугольник – тремя, тетраэдр – четырьмя и т.д. В то же время выпуклая многогранная область, являясь неограниченным множеством, не определяется однозначно своими угловыми точками: любую ее точку нельзя представить в виде выпуклой линейной комбинации угловых точек.

Свойства задачи линейного программирования.

1. Множество всех допустимых решений системы ограничений задачи линейного программирования является выпуклым, т.е. представляет выпуклый многогранник или выпуклую многогранную область.

В дальнейшем будем называть их одним термином – многогранником решений.

2. Если задача линейного программирования имеет оптимальное решение, то линейная функция принимает максимальное (минимальное) значение в одной из угловых точек многогранника решений. Если линейная функция принимает максимальное значение более чем в одной угловой точке, то она принимает его в любой точке, являющейся выпуклой линейной комбинацией этих точек.

Данное утверждение является фундаментальным, так как оно указывает принципиальный путь решения задач линейного программирования. Действительно, вместо исследования бесконечного множества допустимых решений для нахождения среди них искомого оптимального решения необходимо исследовать лишь конечное число угловых точек многогранника решений.

3. Каждому допустимому базисному решению задачи линейного программирования соответствует угловая точка многогранника решений, и наоборот, каждой угловой точке многогранника решений соответствует допустимое базисное решение.

Следствие: если задача линейного программирования имеет оптимальное решение, то оно совпадает, по крайней мере, с одним из ее допустимых базисных решений.

Это означает, что оптимум линейной функции задачи линейного программирования следует искать среди конечного числа ее допустимых базисных решений.

Так как мы решаем задачу на экстремум, то возникает вопрос: можно ли использовать классические методы исследования на экстремум функции многих переменных. Применим необходимое условие экстремума функции, которое состоит в том, что частные производные функции многих переменных или равны нулю, или не существуют. В данном случае $\frac{\partial Z}{\partial x_i} = c_i = 0, \quad i = 1,2,\ldots,n$. Но если все $c_i = 0$, то и $Z = 0$, т.е. экстремум функции не находится. Связано это с тем, что производную можно использовать для определения экстремума только во внутренних точках области решений, а в данном случае экстремум, находится на границах области. Следовательно, необходимо разработать специальные методы поиска экстремума.

Математические модели простейших задач

Задача об использовании ресурсов (задача планирования производства).

Для изготовления нескольких видов продукции $P_1, P_2,...,P_n$ используют m видов ресурсов $S_1, S_2,...,S_m$. Это могут быть различные материалы, электроэнергия, полуфабрикаты и т.д. Объем каждого вида ресурсов ограничен и известен $(b_1, b_2,..b_m)$. Известно также a_{ij} $(i=1..m;\ j=1..n)$ - количество каждого i - го вида ресурса, расходуемого на производство j - го вида продукции. Кроме того, известна прибыль, получаемая от реализации единицы каждого вида продукции $(c_1, c_2,...,c_n)$. Условия задачи можно представить в виде табл.1.

Таблица 1.

вид ресурсов	объем ресурсов	a_{ij}			
		P_1	P_2	…	P_n
S_1	b_1	a_{11}	a_{12}	…	a_{1n}
S_2	b_2	a_{21}	a_{22}	…	a_{2n}
…	…	…	…	…	…
S_m	b_m	a_{m1}	a_{m2}	…	a_{mn}
прибыль		c_1	c_2	…	c_n

Необходимо составить такой план производства продукции, при котором прибыль от ее реализации будет максимальной.

Решение. Пусть x_j $(j=1..n)$ – количество каждого вида продукции, которое необходимо произвести. Для первого ресурса имеет место неравенство – ограничение $a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1$.

Аналогичные неравенства будут и для остальных видов ресурсов. Следует учитывать также, что все значения $x_j \ge 0,\ j=1..n$. Общая прибыль, получаемая от реализации всей продукции, может быть представлена как функция $Z(X) = c_1x_1 + c_2x_2 + ... + c_nx_n$. Необходимо эту функцию максимизировать. Таким образом, математическая модель задачи использования ресурсов запишется в виде

$$Z(X) = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max,$$

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

$$x_j \ge 0,\ j=1..n$$

В более компактной форме целевую функцию и систему ограничений можно записать следующим образом

$$Z(X)=\sum_{j=1}^{n} c_j x_j \to \max,$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i=1..m,$$

$$x_j \ge 0,\ j=1..n.$$

Задача легко обобщается на случай выпуска n видов продукции с использованием m видов ресурсов. Обозначим через x_j $(j=1..n)$ – число единиц продукции P_j, запланированной к производству; a_{ij} - число единиц ресурса S_i, затрачиваемого на изготовление единицы продукции P_j (числа a_{ij} часто называют *технологическими коэффициентами*), b_i $(i=1..m)$ – запас ресурсов S_i, c_j - прибыль от реализации единицы продукции P_j. Тогда математическая модель задачи об использовании ресурсов в общей постановке примет вид: *найти такой план* $X=(x_1, x_2, \ldots, x_n)$ *выпуска продукции, удовлетворяющий системе*

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

и условию $x_j \ge 0$, $j=1..n$, *при котором функция* $Z(X)=c_1x_1 + c_2x_2 + \ldots + c_nx_n$ *принимает максимальное значение.*

<u>Задача о составлении рациона питания (задача о диете, задача о смесях).</u>

Требуется составить дневной рацион питания животного на основе имеющихся видов кормов так, чтобы общая стоимость использованных кормов была минимальной. При этом животное не должно получать менее определенного количества питательных веществ, например, таких, как жиры, углеводы, белки, витамины и т.п. Каждый вид корма содержит разную комбинацию этих веществ. Известна цена единицы каждого корма.

Пусть имеются n различных кормов (продуктов) $P_1, P_2, \ldots, P_n$ и перечень из m необходимых питательных веществ $S_1, S_2, \ldots, S_m$. $S_1, S_2, \ldots, S_m$. Обозначим через a_{ij} содержание (в весовых единицах) i - го питательного вещества в единице j - го корма, а через b_i минимальную суточную потребность животного в i - м питательном веществе. Условия задачи можно представить в виде табл. 2.

Необходимо составить дневной рацион, имеющий минимальную стоимость, в котором содержание каждого вида питательных веществ было бы не менее установленного предела.

Решение. Обозначим через x_j $(j=1..n)$ количество каждого вида корма в ежедневном рационе, $x_j \geq 0$. Для первого вида питательного вещества неравенство – ограничение примет вид $a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \geq b_1$.

Таблица 2.

питательное вещество	суточная потребность	a_{ij}			
		P_1	P_2	$\ldots$	P_n
S_1	b_1	a_{11}	a_{12}	$\ldots$	a_{1n}
S_2	b_2	a_{21}	a_{22}	$\ldots$	a_{2n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
S_m	b_m	a_{m1}	a_{m2}	$\ldots$	a_{mn}
стоимость 1 кг корма		c_1	c_2	$\ldots$	c_n

Аналогично запишутся неравенства и для остальных питательных веществ. Общие затраты на весь рацион питания животного можно найти на основе линейной функции $Z(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$. Эту функцию нужно минимизировать. Математическая модель задачи составления рациона питания имеет вид $Z(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \min$

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

$$x_j \geq 0,\ j=1..n$$

Каноническая форма задачи линейного программирования

В общем случае задача линейного программирования записывается так, что ограничения являются как уравнения, так и неравенства, а переменные могут быть как неотрицательными, так и произвольно изменяющимися. В случае, когда все ограничения являются уравнениями и все переменные удовлетворяют условию неотрицательности, задачу линейного программирования называют *канонической*. Она может быть представлена в координатной, векторной или матричной форме записи.

1. Каноническая задача линейного программирования в *координатной* записи имеет вид $Z(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\min)$

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

$$x_j \geq 0,\ j=1..n$$

Данную задачу можно записать в следующем виде

$$Z(X)=\sum_{j=1}^{n} c_j x_j \to \max(\min),$$

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

$$x_j \ge 0,\ j=1..n.$$

2. Каноническая задача линейного программирования в *векторной* записи имеет вид

$$Z(X)=C\cdot X \to \max(\min)$$
$$A_1x_1 + A_2x_2 + \ldots + A_nx_n = A_0$$
$$X \ge \theta$$

В данном случае введены векторы $X=(x_1,x_2,\ldots,x_n)$, $C=(c_1,c_2,\ldots,c_n)$,

$$A_j=\begin{pmatrix} a_{1j}\\ a_{2j}\\ \ldots\\ a_{mj}\end{pmatrix},\ j=1..n,\qquad A_0=\begin{pmatrix} b_1\\ b_2\\ \ldots\\ b_m\end{pmatrix},\qquad \theta=(0,0,\ldots,0).$$

Здесь $C\cdot X$ - скалярное произведение векторов C и X.

3. Каноническая задача линейного программирования в *матричной* записи имеет вид

$$Z(X)=CX \to \max(\min)$$
$$AX=A_0,\quad X\ge\theta$$

где

$$A=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \ldots & \ldots & \ldots & \ldots\\ a_{m1} & a_{m2} & \ldots & a_{mn}\end{pmatrix},\ X=\begin{pmatrix} x_1\\ x_2\\ \ldots\\ x_n\end{pmatrix},\ A_0=\begin{pmatrix} b_1\\ b_2\\ \ldots\\ b_m\end{pmatrix}.$$

Здесь **A** – матрица коэффициентов системы уравнений, X - матрица-столбец правых частей системы ограничений.

Нередко используются задачи линейного программирования, называемые *симметричными*, которые в *матричной* записи имеют вид

$$Z(X)=CX \to \max,$$
$$AX \le A_0,\ X\ge\theta$$

или $Z(X)=CX \to \min$

$$AX \ge A_0,\ X \ge \theta.$$

Приведение общей задачи линейного программирования к канонической форме

В большинстве методов решения задач линейного программирования предполагается, что система ограничений состоит из уравнений и естественных условий неотрицательности переменных. Однако при составлении математических моделей задач ограничения в основном формируются в системы неравенств, поэтому необходимо уметь переходить от системы неравенств к системе уравнений. Это может быть сделано следующим образом.

Возьмем линейное неравенство

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n \le b$$

и прибавим к его левой части некоторую величину x_{n+1}, такую, чтобы неравенство превратилось в равенство

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n + x_{n+1} = b,$$

где $$x_{n+1} = b - a_1x_1 - a_2x_2 - \ldots - a_nx_n.$$

Неотрицательная переменная $x_{n+1} \ge 0$ называется *дополнительной переменной*.

Следующая теорема дает основание для возможности такого преобразования.

Теорема 1. Каждому решению $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ неравенства $a_1x_1 + a_2x_2 + \ldots + a_nx_n \le b$ соответствует единственное решение $\overline{\beta} = (\beta_1, \beta_2, \ldots, \beta_n, \beta_{n+1})$ уравнения $a_1x_1 + a_2x_2 + \ldots + a_nx_n + x_{n+1} = b$.

Если в задаче использования ресурсов в левую часть каждого уравнения системы ограничений

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i,\ i = 1..m$$

добавить положительную переменную x_{n+i}, то получится система уравнений – ограничений

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i,\ i = 1..m.$$

В задаче составления рациона питания система ограничений – неравенств имеет вид

$$\sum_{j=1}^{n} a_{ij}x_j \ge b_i,\ i = 1..m.$$

В этом случае система уравнений – ограничений получится, если в левой части каждого неравенства вычесть соответствующую неотрицательную дополнительную переменную

$$\sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i,\ i = 1..m.$$

Полученная таким образом система уравнений – ограничений вместе с условиями неотрицательности переменных, т.е. $x_j \geq 0,\ j = 1..(n+m)$ и целевой функцией является канонической формой записи задачи линейного программирования.

Дополнительные переменные вводятся в целевую функцию с нулевыми коэффициентами и поэтому не влияют на ее значение.

В случае, когда задача имеет произвольно изменяющиеся переменные, любую такую переменную $x_j \geq 0,\ j = 1..n.$ заменяют разностью двух неотрицательных переменных, т.е. $x_j = x_j^{'} - x_j^{''}$, где $x_j^{'} \geq 0$ и $x_j^{''} \geq 0$.

Иногда возникает также необходимость перейти в задаче от нахождения минимума к нахождению максимума и наоборот. Для этого достаточно изменить знаки всех коэффициентов целевой функции на противоположные, а в остальном задачу оставить без изменения. Оптимальные решения полученных таким образом задач на максимум и минимум совпадают, а значения целевых функции при оптимальных решениях отличаются только знаком.

Графический метод решения задач линейного программирования

Задача с двумя переменными

Пусть требуется найти максимальное значение функции

$$Z(X) = c_1x_1 + c_2x_2 \tag{1.9}$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1, \\ a_{21}x_1 + a_{22}x_2 \leq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m \end{cases} \tag{1.10}$$

$$x_1 \geq 0,\ x_2 \geq 0$$

Допустим, что система ограничений (1.10) совместна, т.е. имеет решение, а многоугольник ее решений (ОДР) ограничен.

Каждое из неравенств (1.10) определяет полуплоскость с границей $a_{i1}x_1 + a_{12}x_2 = b_i,\ i = 1..m$ или $x_1 = 0$, $x_2 = 0$. Представим этот многоугольник на плоскости Ox_1x_2 (Рис.1.2).

Линейная функция (1.9) при фиксированных значениях $Z(X)$ является уравнением прямой линии $c_1x_1 + c_2x_2 = const$.

Изобразим прямую, соответствующую линейной функции, при $Z(X) = 0$. Эта прямая пройдет через начало координат. Другим значениям $Z(X)$ будут соответствовать прямые, параллельные друг другу. Прямая, уравнение которой получено из целевой функции задачи при равенстве ее постоянной величине, называется *линией уровня*.

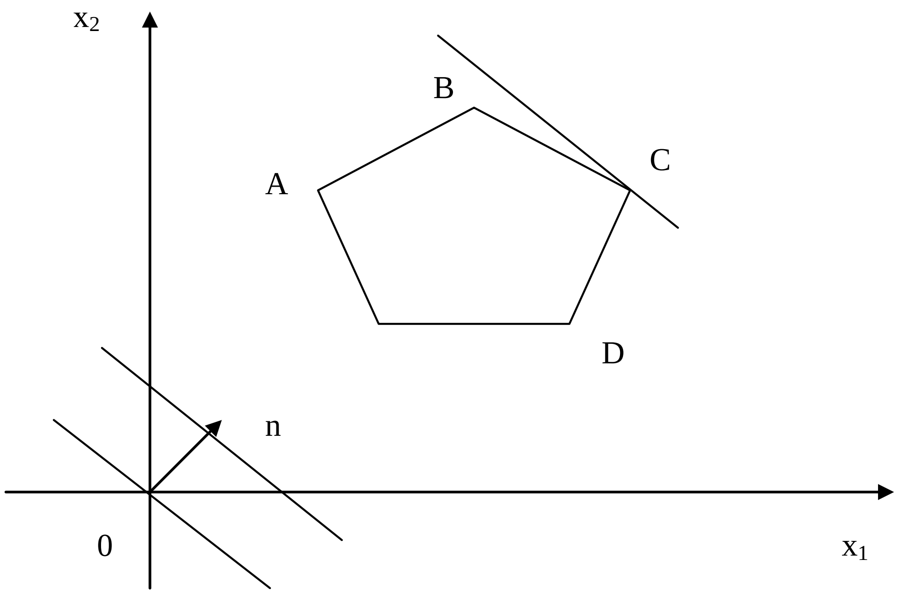


Рис. 1.2 Линии уровня

Коэффициенты при переменных в линейном уравнении являются координатами нормального вектора к соответствующей прямой или плоскости. Следовательно, нормальный вектор линий уровня $\bar{n}$ имеет координаты c_1 и c_2, т.е. $\bar{n} = (c_1, c_2)$.

Если перемещать линию уровня параллельно ее начальному положению в направлении вектора $\bar{n}$, то для данного случая (Рис.1.2) последней точкой, в которой линия уровня коснется ОДР, окажется точка С.

Линия уровня, имеющая общие точки с ОДР и расположенная так, что ОДР целиком находится в одной из полуплоскостей, называется *опорной прямой.*

Теорема 1. Значения целевой функции в точках линии уровня увеличиваются, если линию уровня перемещать параллельно начальному положению в направлении нормали, и убывают при перемещении в противоположном направлении.

Алгоритм решения задачи линейного программирования с двумя переменными графическим методом таков:

1.Строится область допустимых решений.

2.Строится вектор $\bar{n} = (c_1, c_2)$ с точкой приложения в начале координат.

3.Перпендикулярно вектору $\bar{n}$ проводится одна из линий уровня, например, линия уровня, соответствующая уравнению $c_1 x_1 + c_2 x_2 = 0$.

4.Линия уровня перемещается до положения опорной прямой. На этой прямой и будет находиться максимум или минимум функции.

В зависимости от вида ОДР и целевой функции $Z(X)$ задача может иметь единственное решение (Рис.1.3а), бесконечное множество решений (рис.1. 3б) или не иметь ни одного оптимального решения (рис.1.3в).

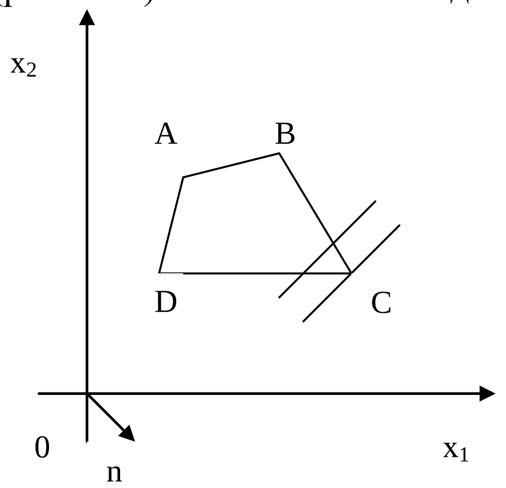


Рис. 1.3а. Единственность решения

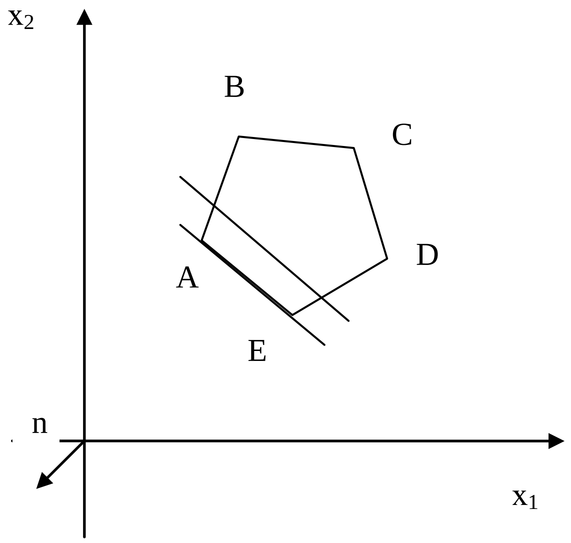


Рис. 1.3б. Бесконечное множество решений

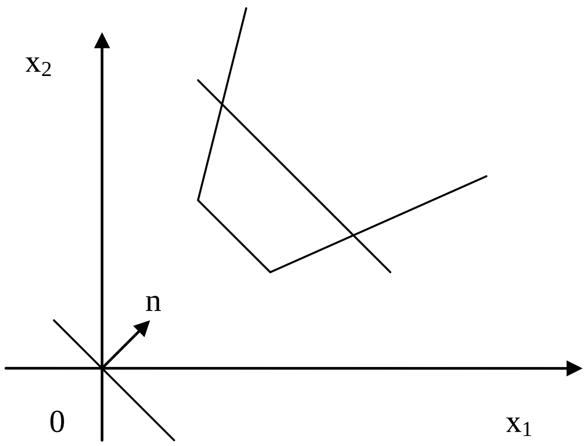


Рис. 1.3в. Задача, не имеющая решения

Пример1.1. Решить задачу линейного программирования графическим методом:

$$Z(X) = 2x_1 + 4x_2 \to \max$$

$$\begin{cases} -2x_1 + 3x_2 \le 12, \\ x_1 + x_2 \le 9, \\ 3x_1 - 2x_2 \le 12, \end{cases}$$

$$x_1 \ge 0,\ x_2 \ge 0$$

Решение. Изобразим на плоскости систему координат Ox_1x_2 и построим граничные прямые области допустимых решений (номера прямых соответствуют их порядковому номеру в системе).

$$\begin{cases} -2x_1 + 3x_2 = 12, \\ x_1 + x_2 = 9, \\ 3x_1 - 2x_2 = 12, \end{cases}$$

$$x_1 = 0,\ x_2 = 0$$

Область допустимых решений определяется многоугольником OABCD (Рис. 1.4). Для линий уровня $2x_1 + 4x_2 = c$ (c – const) строим нормальный вектор $\overline{n} = (2;4)$. Перпендикулярно вектору $\overline{n}$ построим одну из линий уровня (в нашем случае она проходит через начало координат). Так как задача на максимум, то перемещаем линию уровня в направлении вектора $\overline{n}$ до опорной прямой. Опорной прямой является прямая, проходящая через точку пересечения граничных прямых L_1 и L_2, т.е. через точку $B = L_1 \cap L_2$.

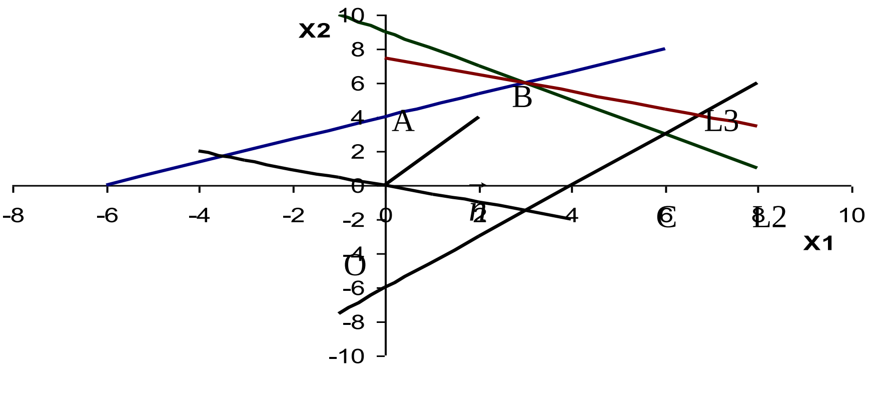


Рис. 1.4 Область допустимых решений

Для определения координат точки B решаем систему уравнений

$$\begin{cases} -2x_1 + 3x_2 = 12, \\ x_1 + x_2 = 9 \end{cases}$$

L1

Получаем $x_1 = 3, x_2 = 6$. Это и будет оптимальное значение данной задачи, которому соответствует максимальное значение целевой функции $\max Z(B) = 2 \cdot 3 + 4 \cdot 6 = 30$.

Пример 1.2. Предприятие располагает тремя видами сырья и может выпускать одну и ту же продукцию двумя способами. При этом за 1 ч работы первым способом выпускается 20 единиц продукции, а вторым способом – 30 единиц. Количество сырья (кг) того или иного вида, расходуемого за 1 ч при различных способах производства, и запасы сырья (кг) приведены в таблице.
Таблица 3.

Способ производства	Сырье		
	1	2	3
первый	10	20	15
второй	20	10	15
Запасы сырья	100	100	90

Требуется найти план производства, при котором будет выпущено наибольшее количество продукции.

Решение. Обозначим через x_1 и x_2 время использования (ч) соответственно первого и второго способов производства. Получаем задачу линейного программирования

$$Z(X) = 20x_1 + 30x_2 \to \max$$

$$\begin{cases} 10x_1 + 20x_2 \le 100 \\ 20x_1 + 10x_2 \le 100 \\ 15x_1 + 15x_2 \le 90 \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0,$$

которую можно решить графическим способом.

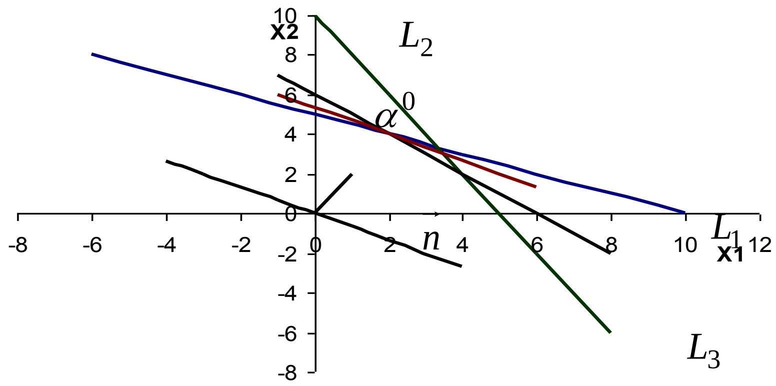


Рис. 1.5. Область допустимых решений

На рисунке 1.5 изображены ОДР и оптимальное значение этой задачи α^0. Любая точка из области допустимых решений является планом работы предприятия, для реализации которого хватит имеющихся запасов сырья. Оптимальное решение α^0 – это план из ОДР, при котором будет выпущено наибольшее количество продукции.

Очевидно, что α^0 является точкой пересечения прямых L_1 и L_3. Решая систему уравнений (номера прямых соответствуют их порядковому номеру в системе), получаем $x_1 = 2, x_2 = 4$.

Таким образом, для производства наибольшего количества продукции при имеющихся запасах сырья необходимо 2 ч применять первый способ производства и 4 ч – второй способ. При этом будет изготовлено $Z(\alpha^0) = 20 \cdot 2 + 30 \cdot 4 = 160$ единиц продукции.

Опорное решение задачи линейного программирования, его взаимосвязь с угловыми точками

Рассмотрим систему ограничений некоторой конкретной задачи

$$\begin{cases} x_1 \qquad - 2x_3 + 5x_4 = 7, \\ \qquad x_2 + 3x_3 + 4x_4 = 15, \end{cases}$$

$x_j \geq 0,\ j = 1,2,3,4$

В *векторной* записи эта система примет вид:

$A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 = A_0$,

где

$$A_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ A_3 = \begin{pmatrix} -2 \\ 3 \end{pmatrix},\ A_4 = \begin{pmatrix} 5 \\ 4 \end{pmatrix},\ A_0 = \begin{pmatrix} 7 \\ 15 \end{pmatrix}.$$

Векторы A_1, A_2, A_3, A_4 называются *векторами условий.*

Данная система имеет бесконечное множество допустимых решений, например $X_1 = (4,8,1,1)$, $X_2 = (9,12,1,0)$, $X_3 = (7,15,0,0)$. Решение X_3 является *базисным.*

Положительным координатам допустимых решений ставят в соответствие векторы условий. Для решений X_1 и X_2 это векторы A_1, A_2, A_3, A_4 и A_1, A_2, A_3 соответственно. Эти системы векторов линейно зависимы, так как число входящих в них векторов (4 и 3) больше размерности векторов (2). Таких решений бесконечное множество. Для базисного решения X_3 единичные векторы A_1, A_2, соответствующие положительным координатам, линейно- независимы.

Любая система уравнений имеет конечное число базисных решений, равное C_n^r, где n - число неизвестных, r - ранг системы векторов условий. Базисные решения, координаты которых удовлетворяют условию неотрицательности, являются *опорными.* Они являются угловыми точками области допустимых решений задачи.

Опорным решением задачи линейного программирования называется такое допустимое решение $X = (x_{10}, x_{20}, \ldots, x_{m0}, 0, \ldots, 0)$, для которого векторы условий, соответствующие положительным координатам $A_1, A_2, \ldots, A_m$, линейно-независимы.

Число отличных от нуля координат опорного решения не может быть больше ранга r системы векторов условий (т.е. числа линейно-независимых уравнений системы ограничений). Будем считать, что система ограничений состоит из линейно-независимых уравнений, т.е. $m = r$. Если число отличных от нуля координат опорного решения равно m, то решение называется *невырожденным*, в противном случае – *вырожденным*.

Базисом опорного решения называется базис системы векторов условий задачи, в состав которого входят векторы, соответствующие отличным от нуля координатам опорного решения.

Теорема. Любое опорное решение является угловой точкой области допустимых решений.

Теорема. Любая угловая точка области допустимых решений является опорным решением.

СИМПЛЕКСНЫЙ МЕТОД РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Симплексный метод – метод последовательного улучшения решения (плана) и нахождения оптимального решения (плана). Он заключается в том, что вначале находится любое допустимое базисное решение, соответствующее одной их угловых точек многогранника решений, а затем это решение целенаправленно улучшается, переходя к новому базисному решению в соседней угловой точке, при котором значение целевой функции больше (меньше) и т.д., пока не будет получено оптимальное решение.

Метод называется симплексным, так как области допустимых решений задач, которые рассматривались на начальном этапе развития метода, имели простейший $(simple)$ вид (рис. 1.6).

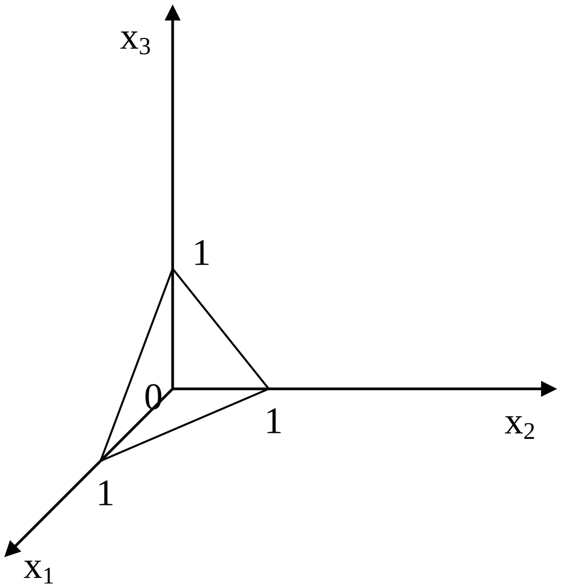


Рис. 1.6 Трехмерный симплекс

Симплексный метод является универсальным методом, которым можно решить любую задачу линейного программирования, приведенную к канонической форме.

Пример. Для изготовления двух видов продукции P_1 и P_2 используют четыре вида ресурсов S_1, S_2, S_3 и S_4. Запасы ресурсов, число единиц ресурсов, затрачиваемых на изготовление единицы продукции, приведены в таблице 4 (цифры условные).

Таблица 4.

Вид ресурса	Запас ресурса	Число единиц ресурсов, затрачиваемых на изготовление единицы продукции	
		P_1	P_2
S_1	18	1	3
S_2	16	2	1

S_3	5	-	1
S_4	21	3	-

Прибыль, получаемая от единицы продукции P_1 и P_2 соответственно 2 и 3 руб.

Необходимо составить такой план производства продукции, при котором прибыль от ее реализации будет максимальной.

$$Z(X) = 2x_1 + 3x_2 \to \max$$

при ограничениях
$$\begin{cases} x_1 + 3x_2 \le 18, \\ 2x_1 + x_2 \le 16, \\ x_2 \le 5, \\ 3x_1 \le 21, \end{cases} \qquad (1.11)$$

$$x_1 \ge 0, x_2 \ge 0$$

Решение. Симплексный метод применяется к задаче в канонической форме, в которой система ограничений есть система уравнений.

Введем в каждое уравнение системы (1.11) по одной из дополнительных неотрицательных переменных x_3, x_4, x_5, x_6 со знаком «плюс», так как все неравенства системы ограничений вида «$\le$» .
Получим следующую систему ограничений:

$$\begin{cases} x_1 + 3x_2 + x_3 = 18, \\ 2x_1 + x_2 + x_4 = 16, \\ x_2 + x_5 = 5, \\ 3x_1 + x_6 = 21, \end{cases} \qquad (1.12)$$

Любое решение системы (1.12) будет состоять из четырех основных и двух неосновных переменных. В качестве основных переменных. В качестве основных переменных на первом шаге берем дополнительные переменные x_3, x_4, x_5, x_6 (они образуют единичную матрицу, что гарантирует, что определитель из коэффициентов при них (базисный минор) будет отличен от нуля).

I шаг. Основные переменные x_3, x_4, x_5, x_6; неосновные переменные x_1, x_2. Выразим основные переменные через неосновные:

$$\begin{cases} x_3 = 18 - x_1 - 3x_2, \\ x_4 = 16 - 2x_1 - x_2, \\ x_5 = 5 - x_2, \\ x_6 = 21 - 3x_1 \end{cases}$$

Полагая неосновные переменные равными нулю, т.е. $x_1 = 0$, $x_2 = 0$, получим первое базисное решение, которое является допустимым и соответствует вершине O(0;0) многогранника *OABCDE* на рис. 7. При этом $Z_1 = Z(X_1) = 2 \cdot 0 + 3 \cdot 0 = 0$. От этого первоначального решения нужно перейти к другому, при котором значение линейной функции Z увеличится.

Этого можно достичь изменением базисного решения, т.е. переводом одной из неосновных переменных в основные, а одной из основных в неосновные. Геометрически это означает переход к соседней вершине A или E (Рис. 1.7).

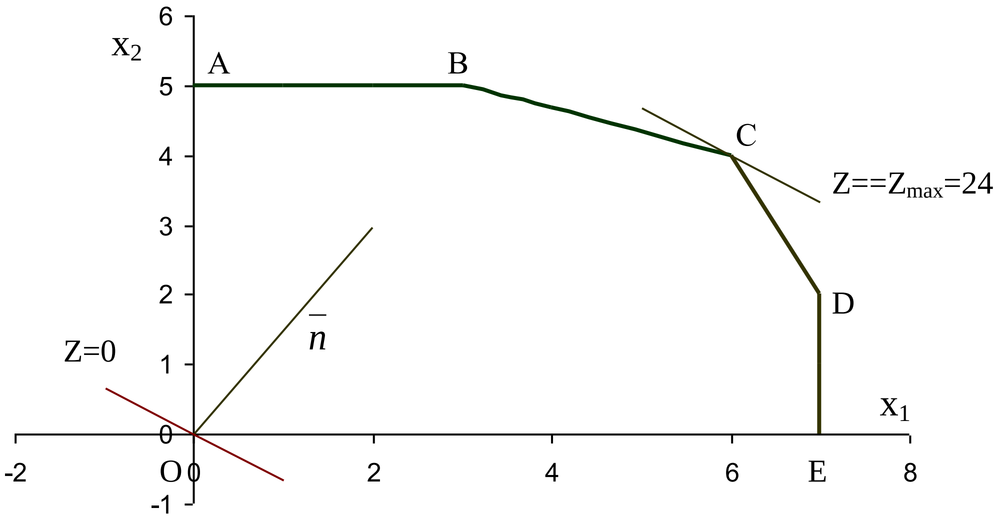


Рис. 1.7. Многогранник допустимых решений

Увеличить значение линейной функции Z можно переводом в основные (в общем случае неравные нулю) переменную, которая входит в выражение линейной функции с *положительным* коэффициентом, например x_2 (или x_1). На ее место из основных в неосновные переводится переменная, которая первая обратится в нуль при росте x_2. Уравнение, из которого находится эта переменная, называется *разрешающим.*

Разрешающее уравнение, а значит и переменная, переводимая из основных в неосновные, находится по минимуму *оценочных отношений*, каждое из которых определяет ограничение данного уравнения по росту переменной x_2:

$$x_2 = \min\left\{\frac{18}{3}; \frac{16}{1}; \frac{5}{1}; \infty\right\} = 5$$

Эти отношения определяются отношением свободных членов уравнений к коэффициентам при переводимой в основные переменной x_2 при условии, что эти числа имеют разные знаки. В случае одинаковых знаков или отсутствия в уравнении переменной, переводимой в основные, соответствующее оценочное отношение считается равным ∞ (четвертое уравнение), что означает, что данное уравнение не ограничивает рост переменной x_2, переводимой в основные. При $x_2 = 5$ переменная x_5 обращается в нуль и переходит в неосновные, т.е. разрешающее уравнение третье.

II шаг. Основные переменные x_2, x_3, x_4, x_6; неосновные переменные x_1, x_5. Выразим новые основные переменные через новые неосновные, начиная с разрешающего уравнения (его используем при записи выражения для x_2):

$$\begin{cases} x_2 = 5 - x_5, \\ x_3 = 18 - x_1 - 3(5 - x_5), \\ x_4 = 16 - 2x_1 - (5 - x_5), \\ x_6 = 21 - 3x_1 \end{cases} \quad \text{или} \quad \begin{cases} x_2 = 5 - x_5, \\ x_3 = 3 - x_1 + 3x_5, \\ x_4 = 11 - 2x_1 + x_5, \\ x_6 = 21 - 3x_1. \end{cases}$$

Второе базисное решение $X_2 = (0;5;3;11;0;21)$ соответствует вершине $A(0;5)$ на рисунке 1.7.

Выразив линейную функцию через неосновные переменные на этом шаге, получим

$$Z(X) = 2x_1 + 3x_2 = 2x_1 + 3(5 - x_5) = 15 + 2x_1 - 3x_5$$

Значение линейной функции $Z_2 = Z(X_2) = 15$.

Переводим в основные переменную x_1, которая входит в выражение функции с положительным коэффициентом. При этом

$$x_1 = \min\left\{\infty; \frac{3}{1}; \frac{11}{2}; \frac{21}{3}\right\} = 3$$

Второе уравнение является разрешающим, переменная x_3 переходит в неосновные.

III шаг. Основные переменные x_1, x_2, x_4, x_6; неосновные переменные x_3, x_5. Выражаем новые неосновные переменные и линейную функцию через новые неосновные, начиная с разрешающего уравнения (его используем при записи выражения для x_1). После преобразований получаем

$$\begin{cases} x_1 = 3 - x_3 + 3x_5, \\ x_2 = 5 - x_5, \\ x_4 = 5 + 2x_3 - 5x_5, \\ x_6 = 12 + 3x_3 - 9x_5, \end{cases}$$

$$Z = 21 - 2x_3 + 3x_5.$$

Третье базисное решение $X_3 = (3;5;0;5;0;12)$ соответствует вершине $B(3;5)$ на рисунке 1.7; $Z_3 = Z(X_3) = 21$.

IV шаг. Основные переменные x_1, x_2, x_5, x_6; неосновные переменные x_3, x_4. После преобразований получим:

$$\begin{cases} x_1 = 6 + \frac{1}{5}x_3 - \frac{3}{5}x_4, \\ x_2 = 4 - \frac{2}{5}x_3 + \frac{1}{5}x_4, \\ x_5 = 1 + \frac{2}{5}x_3 - \frac{1}{5}x_4, \\ x_6 = 3 - \frac{3}{5}x_3 + \frac{9}{5}x_4, \end{cases}$$

$$Z = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_5$$

Четвертое базисное решение $X_4 = (6;4;0;0;1;3)$ соответствует вершине $C(6;4)$ на рисунке 1.7; $Z_4 = Z(X_4) = 24$.

Решение X_4 является оптимальным (обозначаем X^*), так как выполнен *критерий оптимальности* решения задачи на отыскание *максимума* линейной функции – *отсутствие в выражении линейной функции положительных коэффициентов при неосновных переменных.*

Прибыль предприятия принимает максимальное значение 24 руб. при реализации 6 ед. продукции $P_1(x_1 = 6)$ и 4 ед. продукции $P_2(x_2 = 4)$. Дополнительные переменные x_3, x_4, x_5, x_6 показывают разницу между запасами ресурсов каждого вида и их потреблением, т.е. остатки ресурсов. При оптимальном плане производства $x_3 = x_4 = 0$ остатки ресурсов S_1 и S_2 равны нулю, а остатки ресурсов S_3 и S_4 равны соответственно 1 и 3 ед.

При определении *минимума* линейной функции Z на каждом шаге решения задачи симплексным методом необходимо переводить в основные те переменные, которые входят в выражение линейной функции с *отрицательным* коэффициентом. В этом случае *критерий оптимальности* решения задачи на *отыскание минимума* линейной функции – *отсутствие в выражении линейной функции отрицательных коэффициентов при неосновных переменных.*

Следующий пример показывает, что не всегда первое базисное решение является допустимым.

Пример. Решить симплексным методом задачу

$$Z = -x_1 + 2x_2 \to \max$$

при ограничениях $$\begin{cases} 2x_1 - x_2 \ge 1, \\ -x_1 + 4x_2 \le 24, \\ x_1 - x_2 \le 3, \end{cases}$$

$$x_1 \ge 0,\ x_2 \ge 0$$

Решение. Введя в каждое неравенство по одной из дополнительных неотрицательных переменных (со знаком «минус» в неравенствах вида «$\ge$» и со знаком «плюс» - вида «$\le$») получаем следующую систему ограничений:

$$\begin{cases} 2x_1 - x_2 - x_3 = 1, \\ -x_1 + 4x_2 + x_4 = 24, \\ x_1 - x_2 + x_5 = 3. \end{cases}$$

I шаг. Основные переменные x_3, x_4, x_5; неосновные переменные x_1, x_2.

$$\begin{cases} x_3 = -1 + 2x_1 - x_2, \\ x_4 = 24 + x_1 - 4x_2, \\ x_5 = 3 - x_1 + x_2. \end{cases}$$

Первое базисное решение $X_1 = (0;0;-1;24;3)$ - недопустимое. Геометрически ему соответствует точка $O(0;0)$, расположенная вне многогранника решений (Рис. 1.8).

Для получения улучшенного решения (например, не содержащая отрицательных компонент) необходимо перевести в основные переменную x_1, которая входит x_1 с *положительным* коэффициентом в первое уравнение, имеющее *отрицательный* свободный член

$$x_1 = \min\left\{\frac{1}{2}; \infty; \frac{3}{1}\right\} = \frac{1}{2}$$

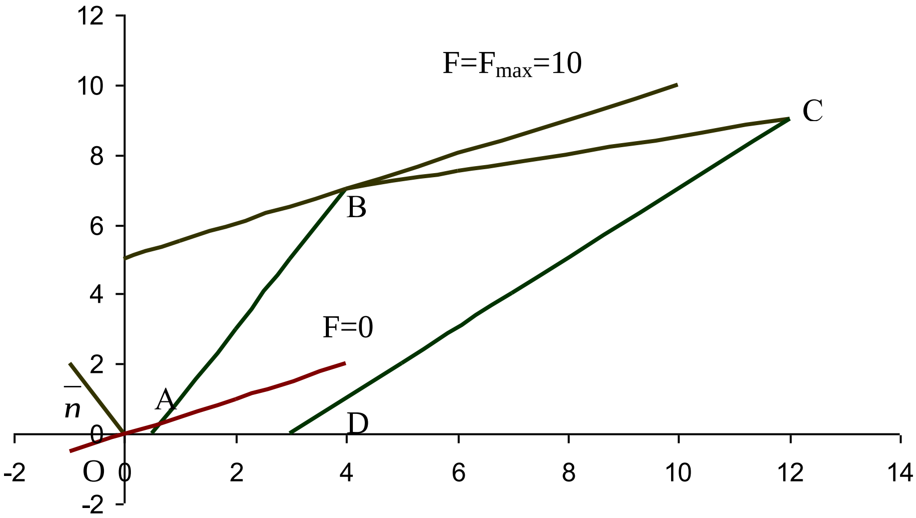


Рис. 1.8. Область допустимых решений

При $x_1 = \dfrac{1}{2}$ первая из основных переменных x_3 станет равной нулю $(x_3 = 0)$ и переходит в неосновные переменные.

II шаг. Основные переменные x_1, x_4, x_5; неосновные переменные: x_2, x_3.

После преобразований получаем

$$\begin{cases} x_1 = \frac{1}{2} + \frac{1}{2}x_2 + \frac{1}{2}x_3, \\ x_4 = \frac{49}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3, \\ x_5 = \frac{5}{2} - \frac{1}{2}x_2 - \frac{1}{2}x_3. \end{cases}$$

Второе базисное решение $X_2 = \left(\frac{1}{2};0;0;\frac{49}{2};\frac{5}{2}\right)$ - допустимое решение, геометрически ему соответствует точка $A\left(\frac{1}{2};0\right)$ (Рис. 1.8) – вершина многогранника решений.

После получения допустимого базисного решения находится выражение линейной функции через новые неосновные переменные:

$$Z = -x_1 + 2x_2 = -\left(\frac{1}{2} + \frac{1}{2}x_2 + \frac{1}{2}x_3\right) + 2x_2 = -\frac{1}{2} + \frac{3}{2}x_2 + \frac{1}{2}x_3$$

и $Z_2 = Z(X_2) = -\frac{1}{2}$.

Далее задача решается до получения оптимального решения $X^* = X_3 = (4;7;0;0;6)$, при котором $Z_{\max} = Z(X_3) = 10$ (на Рис. 1.8 оно соответствует точке $B(4;7)$).

Указанные выше примеры проясняют суть симплексного метода. В ряде случаев бывает удобнее использовать симплексные таблицы, причем в случае недопустимого базисного решения можно применять так называемый *M – метод* или *метод искусственного базиса.*

Метод искусственного базиса

Метод искусственного базиса применяется для решения задач линейного программирования в случае, когда задача не имеет начального опорного решения с базисом из единичных векторов.

Согласно данному методу для задачи линейного программирования составляется так называемая *расширенная* задача, которая решается симплексным методом. На основе результатов решения расширенной задачи либо находится оптимальное решение исходной задачи, либо устанавливается причина его отсутствия.

Пусть имеется каноническая задача линейного программирования

$$Z(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\min)$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{1.13}$$

$$x_j \ge 0 \quad \forall j$$

Без ограничения общности можно считать, что правые части уравнений системы ограничений неотрицательны, т.е. $b_i \ge 0 \quad \forall i$.

В дальнейшем для краткости записи при доказательствах используется компактная запись этой задачи:

$$Z(X) = \sum_{j=1}^{n} c_j x_j \to \max(\min)$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1..m \tag{1.14}$$

$$x_j \ge 0 \quad \forall j, b_i \ge 0 \quad \forall i$$

Для исходной задачи составляется расширенная задача. При этом используются искусственные переменные.

Искусственными переменными называются неотрицательные переменные, которые вводятся в ограничения-равенства для получения начального опорного решения с базисом из единичных векторов. Каждая искусственная переменная вводится в левую часть одного из уравнений системы ограничений с коэффициентом +1 и в целевую функцию в задаче на максимум с коэффициентом $-M$, а в задаче на минимум с коэффициентом $+M$. Число M сколь угодно большое по сравнению с единицей $(M >> 1)$.

В общем случае расширенная задача на максимум имеет вид

$$\overline{Z}(\overline{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n - Mx_{n+1} - Mx_{n+2} - \ldots - Mx_{n+m} \to \max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + x_{n+m} = b_m, \end{cases} \tag{1.15}$$

$$x_j \ge 0,\ j = 1..n+m; b_i \ge 0, i = 1..m$$

или в компактной записи

$$\overline{Z}(\overline{X}) = \sum_{j=1}^{n} c_j x_j - \sum_{i=1}^{m} Mx_{n+i} \to \max,$$

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+1} = b_i, \quad i = 1..m \tag{1.16}$$

$$x_j \geq 0 \quad \forall j, b_i \geq 0 \quad \forall i$$

Данная задача имеет начальное опорное решение $\overline{X_1} = (0;0;\ldots;0;b_1;b_2;\ldots;b_m)$ с базисом $\overline{B_1} = (A_{n+1};A_{n+2};\ldots;A_{n+m})$.

Здесь и дальше для расширенной задачи отмечаются чертой сверху следующие величины: целевая функция $\overline{Z}(\overline{X})$, допустимое решение $\overline{X}$, опорные решения $\overline{X_i}$, базисы опорных решений $\overline{B_i}$, область допустимых решений $\overline{G}$.

Для обоснования метода приведем две леммы и три теоремы.

Лемма 1.1. Любому допустимому решению $X = (x_1, x_2, \ldots, x_n)$ исходной задачи линейного программирования соответствует допустимое решение расширенной задачи $\overline{X} = (x_1, x_2, \ldots, x_n, 0, \ldots, 0)$ и, наоборот, любому допустимому решению расширенной задачи $\overline{X} = (x_1, x_2, \ldots, x_n, 0, \ldots, 0)$ соответствует допустимое решение исходной задачи $X = (x_1, x_2, \ldots, x_n)$. При этом значения целевых функций задач на соответствующих решениях совпадают, т.е. $Z(X) = \overline{Z}(\overline{X})$.

Лемма 1.2. Значение целевой функции расширенной задачи на максимум (минимум) на любом допустимом решении $\overline{X} = (x_1, x_2, \ldots, x_n, 0, \ldots, 0)$, у которого все искусственные переменные равны нулю, больше (меньше) значения целевой функции на любом допустимом решении $\overline{X_l} = (l_1, l_2, \ldots, l_n, l_{n+1}, \ldots, l_{n+m})$, у которого хотя бы одна искусственная переменная отлична от нуля.

Теорема 1.1 (признак оптимальности решения). Если расширенная задача линейного программирования имеет оптимальное решение $\overline{X^*} = (x_1^*, x_2^*, \ldots, x_n^*, 0, \ldots, 0)$, у которого все искусственные переменные равны нулю, то исходная задача имеет оптимальное решение $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$, которое получается из $\overline{X^*}$ отбрасыванием этих нулевых искусственных переменных.

Теорема 1.2 (признак отсутствия решения ввиду несовместности системы ограничений). Если расширенная задача имеет оптимальное решение, у которого хотя бы одна искусственная переменная отлична от нуля, то исходная задача не имеет решения ввиду несовместности системы ограничений.

Теорема 1.3 (признак отсутствия решения ввиду неограниченности целевой функции). Если расширенная задача не имеет решения ввиду неограниченности целевой функции, то исходная задача также не имеет решения по той же причине.

Особенности алгоритма метода искусственного базиса

Алгоритм метода искусственного базиса имеет следующие особенности.

1. Ввиду того, что начальное опорное решение расширенной задачи содержит искусственные переменные, входящие в целевую функцию с коэффициентом *–M* (в задаче на максимум) или *+M* (в задаче на минимум), оценки разложений векторов условий $\Delta_k = C_б X_k - c_k$ (где $C_б$ - коэффициенты целевой функции при базисных переменных, c_k - коэффициент целевой функции при переменной x_k, $X_k = (x_{1k}, x_{2k}, ..., x_{mk})$ - вектор коэффициентов разложения вектора по базису опорного решения) состоят из двух слагаемых $\Delta_k^/$ и $\Delta_k^{//}(M)$, одно из которых $\Delta_k^/$ не зависит от *M*, а другое $\Delta_k^{//}(M)$ зависит от *M*. Так как *M* сколь угодно велико по сравнению с единицей $(M >> 1)$, то на первом этапе расчета для нахождения векторов, вводимых в базис, используются только слагаемые $\Delta_k^{//}(M)$.

2. Векторы, соответствующие искусственным переменным, которые выводятся из базиса опорного решения, исключаются из рассмотрения.

3. После того, как все векторы, соответствующие искусственным переменным, исключаются из базиса, расчет продолжается обычным симплексным методом с использованием $\Delta_k^/$, не зависящих от *M*.

4. Переход от решения расширенной задачи к решению исходной задачи осуществляется с использованием теорем 1.1-1.3.

Пример. Решить задачу методом искусственного базиса задачу линейного программирования

$$Z(X) = 3x_1 + 2x_2 + x_3 - 8x_4 \to \max,$$

$$\begin{cases} 3x_1 + 3x_2 + 4x_3 - 7x_4 = 10, \\ 2x_1 + x_2 + x_3 - 2x_4 = 2, \end{cases}$$

$$x_j \ge 0, \quad j = 1..4$$

Решение. Составляем расширенную задачу. В левые части уравнений ограничений вводим неотрицательные искусственные переменные с коэффициентом $+1$ (всегда). В первое уравнение вводим переменную x_5, во второе – переменную x_6. Данная задача – задача на нахождение максимума, поэтому x_5 и x_6 в целевую функцию вводятся с коэффициентом $-M$. Получаем

$$\overline{Z}(\overline{X}) = 3x_1 + 2x_2 + x_3 - 8x_4 - Mx_5 - Mx_6 \to \max,$$

$$\begin{cases} 3x_1 + 3x_2 + 4x_3 - 8x_4 + x_5 = 10, \\ 2x_1 + x_2 + x_3 - 2x_4 + x_6 = 2, \end{cases}$$

$$x_j \ge 0, \quad j = 1..6$$

Задача имеет начальное опорное решение $\overline{X_1}=(0;0;0;0;10;2)$ с базисом $\overline{B_1}=(A_5;A_6)$. Вычисляем оценки векторов условий по базису опорного решения и значение целевой функции на опорном решении (c_1, c_2, c_3, c_4 - коэффициенты целевой функции при соответствующих переменных):

$$\Delta_1=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}3\\2\end{pmatrix}-c_1=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}3\\2\end{pmatrix}-3=-5M-3,$$

$$\Delta_2=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}3\\1\end{pmatrix}-c_2=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}3\\1\end{pmatrix}-2=-4M-2,$$

$$\Delta_3=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}4\\1\end{pmatrix}-c_3=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}4\\1\end{pmatrix}-1=-5M-1,$$

$$\Delta_4=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}-7\\-2\end{pmatrix}-c_4=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}-7\\-2\end{pmatrix}-(-8)=9M+8,$$

$$\Delta_5=\Delta_6=0,\ \overline{Z}(\overline{X_1})=\begin{pmatrix}-M\\-M\end{pmatrix}\cdot\begin{pmatrix}10\\2\end{pmatrix}=-12M.$$

Записываем исходные данные в симплексную таблицу. При этом оценки Δ_k и $\overline{Z}(\overline{X_1})$ для удобства вычислений записываем в две строки: в первую – слагаемые $\Delta_k^{/}$, не зависящие от M, во вторую – слагаемые $\Delta_k^{//}(M)$, зависящие от M. Значения $\Delta_k^{//}(M)$ удобно указывать без M, имея в виду, что оно там присутствует.

Таблица 5.

			3	2	↓1	-8	-M	-M			
B	Сб	A_0	A_1	A_2	A_3	A_4	A_5	A_6	Θ_1	Θ_2	Θ_3
A_5	-M	10	3	3	4	-7	1	0	10/3	10/3	5/2
← A_6	-M	2	2	1	$\underline{1}$	-2	0	1	1	2	2
$\Delta_k^{/}$		0	-3	-2	-1	8	0	0			
$\Delta_k^{//}(M)$		-12	-5	-4	-5	9	0	0			

Начальное опорное решение не является оптимальным, так как в задаче на максимум имеются отрицательные значения (критерий оптимальности решения задачи на отыскание максимума). Выбираем номер вектора A_k, вводимого в базис опорного решения, и вектора A_l, выводимого из базиса. Для этого вычисляем приращения целевой функции ΔZ_k при введении в базис каждого из векторов с отрицательной оценкой и находим максимум этого приращения. При этом слагаемыми оценок $\Delta_k^{/}$ (без М) пренебрегаем до тех пор, пока хотя бы одно слагаемое $\Delta_k^{//}(M)$ (с М) отлично от нуля. В связи с этим строка со слагаемыми оценок $\Delta_k^{/}$ может отсутствовать в таблице до тех пор, пока присутствует строка $\Delta_k^{//}(M)$. Находим

$\max\limits_{k=1,2,3}\{-1\cdot(-5),-2\cdot(-4),-2\cdot(-5)\}=\max\limits_{k=1,2,3}\{5,8,10\}=10$ при $k=3$. В столбце «A_3» (таблица 4) за разрешающий элемент выбираем коэффициент 1 во второй строке и выполняем преобразование Жордана.

Вектор A_6, выводимый из базиса, исключаем из рассмотрения (вычеркиваем). Получаем опорное решение $\overline{X_2}=(0;0;2;0;2;0)$ с базисом $\overline{B_2}=(A_3;A_5)$ (таблица 6). Решение не является оптимальным, так как имеется отрицательная оценка $\Delta_4^{//}(M)=-1$.

Таблица 6.

B	$C_б$	A_0	A_1	A_2	A_3	A_4	A_5	A_6	Θ_4
A_5	-M	2	-5	-1	0	1	1		2
A_3	1	2	2	1	1	-2	0		-
$\Delta_k^/$		2	-1	-1	0	6	0		
$\Delta_k^{//}(M)$		-2	5	1	0	-1	0		

В столбце «A_4» единственный положительный элемент принимаем за разрешающий и переходим к новому опорному решению $\overline{X_3}=(0;0;6;2;0;0)$ с базисом $\overline{B_3}=(A_3;A_4)$ (таблица 7).

Таблица 7.

B	$C_б$	A_0	A_1	A_2	A_3	A_4	A_5	A_6
A_4	-8	2	-5	-1	0	1		
A_3	1	6	-8	-1	1	0		
Δ_k		-10	29	5	0	0		

Данное опорное решение является единственным оптимальным решением расширенной задачи, так как в задаче на максимум оценки для всех векторов, не входящих в базис, положительны. По признаку оптимальности решения исходная задача также имеет оптимальное решение, которое получается из оптимального решения расширенной задачи отбрасыванием нулевых искусственных переменных, т.е. $X^*=(0;0;6;2)$.

Ответ: $\max Z(X)=-10$ при $X^*=(0;0;6;2)$.

Теория двойственности

Виды математических моделей

Любой задаче линейного программирования (исходной, или прямой) можно поставить в соответствие другую задачу, которая называется *двойственной* или *сопряженной*. Обе эти задачи образуют пару двойственных (или сопряженных) задач линейного программирования. Каждая из задач является двойственной к другой задаче рассматриваемой пары.

Составим двойственную задачу к задаче использования сырья.

Имеется m видов сырья в количестве $b_1, b_2, ..., b_m$, которые используются для изготовления n видов продукции. Известно: a_{ij} - расход i - го вида сырья на единицу j - й продукции; c_j - прибыль при реализации единицы j -го вида продукции.

Составить план выпуска продукции, обеспечивающий максимальную прибыль.

Математическая модель данной задачи имеет вид:

$$Z(X) = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max \tag{1.17}$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m, \end{cases} \left| \begin{array}{l} y_1 \\ y_2 \\ \dots \\ y_m \end{array} \right. \tag{1.18}$$

$$x_j \ge 0, \quad j = 1..n \tag{1.19}$$

Здесь x_j, $j = 1..n$ - объем производства j -го вида продукции.

Предположим, что второй производитель хочет перекупить сырье. Составим двойственную задачу, решение которой позволит определить условия продажи сырья. Введем вектор оценок (цен) видов сырья $Y = (y_1, y_2, ..., y_m)$. Затраты на приобретение i - го вида сырья в количестве b_i равны $b_i y_i$. Второму производителю выгодно минимизировать суммарные затраты на приобретение всех видов сырья, поэтому целевая функция имеет вид

$$F(Y) = b_1y_1 + b_2y_2 + ... + b_my_m \to \min \tag{1.20}$$

Первому производителю невыгодно продавать сырье, если суммарная стоимость всех видов сырья, расходуемых на каждое изделие j - ой продукции, т.е. $a_{1j}y_1 + a_{2j}y_2 + ... + a_{mj}y_m$, меньше прибыли c_j, получаемой при реализации этого изделия. Система ограничений задачи имеет вид

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \le c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \le c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \le c_n, \end{cases} \qquad (1.21)$$

Оценки видов сырья должны удовлетворять условиям неотрицательности

$$y_i \ge 0, \quad i = 1..m \qquad (1.22)$$

Таким образом, связь исходной и двойственной задач состоит в том, что коэффициенты c_j целевой функции исходной задачи являются свободными членами системы ограничений двойственной задачи, свободные члены b_i системы ограничений исходной задачи служат коэффициентами целевой функции двойственной задачи, а матрица коэффициентов системы ограничений двойственной задачи является транспонированной матрицей коэффициентов системы ограничений исходной задачи.

Рассмотренная пара задач относится к симметричным парам двойственных задач. В теории двойственности используются четыре пары двойственных задач (приведем их в *матричной* форме записи):

Исходная задача	Двойственная задача	
	Симметричные пары	
1. $Z(X) = CX \to \max$,	$F(Y) = YA_0 \to \min$,	
$AX \le A_0$,	$YA \ge C$,	(1.23)
$X \ge \Theta$;	$Y \ge \Theta$;	
2. $Z(X) = CX \to \min$,	$F(Y) = YA_0 \to \max$,	
$AX \ge A_0$,	$YA \le C$,	(1.24)
$X \ge \Theta$;	$Y \ge \Theta$.	
	Несимметричные пары	
3. $Z(X) = CX \to \max$,	$F(Y) = YA_0 \to \min$,	
$AX = A_0$,	$YA \ge C$.	(1.25)
$X \ge \Theta$;		
4. $Z(X) = CX \to \min$,	$F(Y) = YA_0 \to \max$	
$AX = A_0$,	$YA \le C$.	(1.26)
$X \ge \Theta$;		

Здесь $C = (c_1, c_2, \ldots, c_n)$, $Y = (y_1, y_2, \ldots, y_m)$, $\Theta = \min\limits_i \left\{ \dfrac{b_i}{a_{ik}} \right\}$,

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & a_{mn} \end{pmatrix},\ A_0=\begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix},\ X=\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

Общие правила составления двойственных задач

При составлении двойственных задач используют следующие правила.

Правило 1. Во всех ограничениях исходной задачи свободные члены должны находиться в правой части, а члены с неизвестными – в левой.

Правило 2. Ограничения – неравенства исходной задачи должны быть записаны так, чтобы знаки неравенств были направлены в одну сторону.

Правило 3. Если знаки неравенств в ограничениях исходной задачи «≤», то целевая функция $Z(X)=c_0+c_1x_1+c_2x_2+\ldots+c_nx_n$ должна максимизироваться, а если «≥», то минимизироваться.

Правило 4. Каждому ограничению исходной задачи соответствует неизвестной в двойственной задаче; при этом неизвестное, отвечающее ограничению – неравенству, должно удовлетворять условию неотрицательности, а неизвестное, отвечающее ограничению – равенству, может быть любого знака.

Правило 5. Целевая функция двойственной задачи имеет вид

$$F(Y)=c_0+b_1y_1+b_2y_2+\ldots+b_my_m,$$

где c_0 - свободный член целевой функции $Z(X)$ исходной задачи; $b_1, b_2, \ldots, b_m$ - свободные члены в ограничениях исходной задачи, при этом b_i - свободный член именно того ограничения, которому соответствует неизвестная y_i: $y_1, y_2, \ldots, y_m$ - неизвестные в двойственной задаче.

Правило 6. Целевая функция $F(Y)$ двойственной задачи должна оптимизироваться противоположным по сравнению с $Z(X)$ образом, т.е. если $Z(X)\to\max$, то $F(Y)\to\min$, и если $Z(X)\to\min$, то $F(Y)\to\max$.

Правило 7. Каждому неизвестному x_j, $j=1,2,\ldots,n$ исходной задачи соответствует ограничение в двойственной задаче. Совокупность этих n ограничений (вместе с условиями неотрицательности неизвестных y_i, соответствующих ограничениям – неравенствам исходной задачи) образует систему ограничений двойственной задачи. Все ограничения двойственной задачи имеют вид неравенств, свободные члены которых находятся в правых частях, а члены с неизвестными $y_1, y_2, \ldots, y_m$ - в левых. Все знаки неравенств имеют вид «≥», если $F(Y)\to\min$, и «≤», если $F(Y)\to\max$.

Коэффициенты, с которыми неизвестные $y_1, y_2, ..., y_m$ входят в ограничение, соответствующее неизвестному x_j, совпадают с коэффициентами при этом x_j в ограничениях исходной задачи, а именно: коэффициент при y_i совпадает с тем же коэффициентом при x_j, с которым x_j входит в ограничение исходной задачи, соответствующее неизвестному y_i.

Пример. Составить задачу, двойственную к данной

$$Z(X) = x_1 + 4x_2 + 3x_3 \to \min,$$

$$\begin{cases} x_1 + x_2 + x_3 \le 10, \\ 2x_1 - x_2 \ge 6, \\ x_1 + 2x_2 + 3x_3 \ge 12, \end{cases}$$

$$x_j \ge 0, \quad j = 1,2,3,$$

Решение. Умножим первое ограничение – неравенство на -1. Задача примет вид исходной задачи симметричной пары двойственных задач (23):

$$Z(X) = x_1 + 4x_2 + 3x_3 \to \min$$

$$\begin{cases} -x_1 - x_2 - x_3 \ge -10, \\ 2x_1 - x_2 \ge 6, \\ x_1 + 2x_2 + 3x_3 \ge 12, \end{cases} \left| \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} \right.$$

$$x_j \ge 0, \ j = 1,2,3.$$

Умножим правые части ограничений на соответствующие переменные двойственной задачи и сложим их, получим целевую функцию

$$F(Y) = -10y_1 + 6y_2 + 12y_3 \to \max.$$

Функция $F(Y)$ максимизируется, так как целевая функция исходной задачи минимизируется.

Умножаем коэффициенты при x_1 на соответствующие переменные двойственной задачи и складываем их: $-1y_1 + 2y_2 + 1y_3$. Данная сумма меньше или равна коэффициенту при x_1 в целевой функции $-1y_1 + 2y_2 + 1y_3 \le 1$.

Неравенство имеет вид «≤», потому что целевая функция двойственной задачи максимизируется. Аналогично составляются еще два ограничения двойственной задачи (соответствуют переменным x_2, x_3):

$$-1y_1 - 1y_2 + 2y_3 \le 4,$$

$$-1y_1 + 3y_3 \le 3.$$

Все переменные двойственной задачи удовлетворяют условию неотрицательности, потому что все ограничения исходной задачи – неравенства.

Окончательно двойственная задача имеет вид

$$F(Y) = -10y_1 + 6y_2 + 12y_3 \to \max,$$

$$\begin{cases} -y_1+2y_2+y_3\le 1, \\ -y_1-y_2+2y_3\le 4, \\ -y_1+3y_3\le 3, \end{cases}$$

$y_i \ge 0$, $i=1,2,3.$

Пример. Составить задачу, двойственную данной

$$Z(X)=x_1-x_2-2x_3+3x_4\to \min,$$

$$\begin{cases} 3x_1+2x_2+x_3+x_4=7, \\ 4x_1+3x_2+2x_3+x_4=10, \end{cases} \left| \begin{array}{l} y_1 \\ y_2 \end{array} \right.$$

$x_j \ge 0$, $j=1,2,3,4$

Решение. Данная задача имеет вид исходной задачи второй несимметричной пары двойственных задач (26). Запишем двойственную задачу

$$F(Y)=7y_1+10y_2\to \max,$$

$$\begin{cases} 3y_1+4y_2\le 1, \\ 2y_1+3y_2\le -1, \\ y_1+2y_2\le -2, \\ y_1+y_2\le 3. \end{cases}$$

Переменные y_1, y_2 могут не удовлетворять условию неотрицательности, так как они соответствуют ограничениям – равенствам исходной задачи.

Пример. Составить задачу, двойственную данной

$$Z(X)=3-2x_1+x_3\to \min,$$

$$\begin{cases} 2x_1+3x_2-x_3\le 3, \\ x_1+2x_2-x_3\le -5, \\ -4x_1-2x_2-3x_3\le 8, \\ x_1-5x_2+4x_3=6, \end{cases}$$

$x_j \ge 0$, $j=1,2,3.$

Решение. Используем общие правила составления двойственных задач. Умножим ограничения – неравенства на -1, так как в задаче на минимум они должны иметь вид «≥» (правило 3). Исходная задача запишется в виде

$$Z(X)=3-2x_1+x_3\to \min,$$

$$\begin{cases} -2x_1-3x_2+x_3\ge -3, \\ -x_1-2x_2+x_3\ge 5, \\ 4x_1+2x_2+3x_3\ge -8, \\ x_1-5x_2+4x_3=6, \end{cases} \left| \begin{array}{l} y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \right.$$

$x_j \geq 0$, $j = 1,2,3.$

Составим двойственную задачу:

$$F(Y) = 3 - 3y_1 + 5y_2 - 8y_3 + 6y_4 \to \max,$$

$$\begin{cases} -2y_1 - y_2 - 4y_3 + y_4 \leq -2, \\ -3y_1 - 2y_2 + 2y_3 - 5y_4 \leq 0, \\ y_1 + y_2 + 3y_3 + 4y_4 \leq 1, \end{cases}$$

$$y_1 \geq 0,\ y_2 \geq 0,\ y_3 \geq 0.$$

Неизвестная y_4, соответствующая ограничению – равенству, может быть любого знака (правило 4).

Теоремы двойственности

Теоремы двойственности позволяют установить взаимосвязь между оптимальными решениями пары двойственных задач. Решив одну из пары двойственных задач, можно или найти оптимальное решение другой задачи, не решая ее, или установить его отсутствие. Возможны следующие случаи: 1) обе задачи из пары двойственных имеют оптимальные решения; 2)одна из задач не имеет решения ввиду неограниченности целевой функции, а другая – ввиду несовместности системы ограничений; 3) обе задачи из пары двойственных не имеют решений из – за неограниченности целевых функций.

Теорема (первая теорема двойственности). Если одна из пары двойственных задач имеет оптимальное решение, то и двойственная к ней имеет оптимальное решение; причем значения целевых функций задач на своих оптимальных решениях совпадают. Если же одна из пары двойственных задач не имеет решения ввиду неограниченности целевой функции, то другая не имеет решения ввиду несовместности системы ограничений.

Пусть имеется симметричная пара двойственных задач

$$Z(X) = \sum_{j=1}^{n} c_j x_j \to \max \qquad\qquad F(Y) = \sum_{i=1}^{m} b_i y_i \to \min,$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i,\ i = 1,2,\ldots,m, \qquad\qquad \sum_{i=1}^{m} a_{ij} y_i \geq c_j,\ j = 1,2,\ldots,n \tag{1.27}$$

$$x_j \geq 0, \quad j = 1,2,\ldots,n; \qquad\qquad y_i \geq 0, \quad i = 1,2,\ldots,m.$$

Теорема (вторая теорема двойственности). Для того чтобы допустимые решения $X = (x_1, x_2, \ldots, x_n)$, $Y = (y_1, y_2, \ldots, y_m)$ являлись оптимальными решениями пары двойственных задач, необходимо и достаточно, чтобы выполнялись следующие равенства:

$$x_j\left(\sum_{i=1}^{m} a_{ij}y_i - c_j\right) = 0, \quad j = 1, 2, \ldots, n; \tag{1.28}$$

$$y_i\left(\sum_{j=1}^{n} a_{ij}x_j - b_i\right) = 0, \; i = 1, 2, \ldots, m \tag{1.29}$$

Иначе, если при подстановке оптимального решения в систему ограничений i - е ограничение исходной задачи выполняется как строгое неравенство, то i - я координата оптимального решения двойственной задачи равна нулю, и, наоборот, если i - я координата оптимального решения двойственной задачи положительная, то i - е ограничение исходной задачи удовлетворяется оптимальным решением как равенством.

Пример. Для данной задачи составить двойственную, решить ее графическим методом и, используя вторую теорему двойственности, найти решение исходной задачи:

$$Z(X) = -2x_1 + 2x_2 + 10x_3 + 4x_4 + 2x_5 \to \min,$$

$$\begin{cases} -x_1 + x_2 + 2x_3 - 2x_5 = 2, & \Big| y_1 \\ -x_1 - x_2 + x_3 + x_4 + x_5 = 3, & \Big| y_2 \end{cases}$$

$$x_j \ge 0 \quad \forall j.$$

Решение. Составим двойственную задачу:

$$F(Y) = 2y_1 + 3y_2 \to \max,$$

$$\begin{cases} -y_1 - y_2 \le -2, \\ y_1 - y_2 \le 2, \\ 2y_1 + y_2 \le 10, \\ y_2 \le 4, \\ -2y_1 + y_2 \le 2. \end{cases}$$

Решим эту задачу графическим методом. На рисунке 1.9 изображены область допустимых решений задачи, нормаль $\bar{n} = (2,3)$ линий уровня $2y_1 + 3y_2 = c$ с оптимальным решением задачи $Y^* = (3,4)$. $Y^* = L_3 \cap L_4$,

$$+\begin{cases} 2y_1 + y_2 = 10 & \Big| \times 1 \\ y_2 = 4 & \Big| \times (-1) \end{cases}; \qquad \Rightarrow 2y_1 = 6 \Rightarrow y_1^* = 3, \qquad y_2^* = 4, \qquad Y^* = (3,4),$$

$$F(Y^*) = 2 \cdot 3 + 3 \cdot 4 = 18.$$

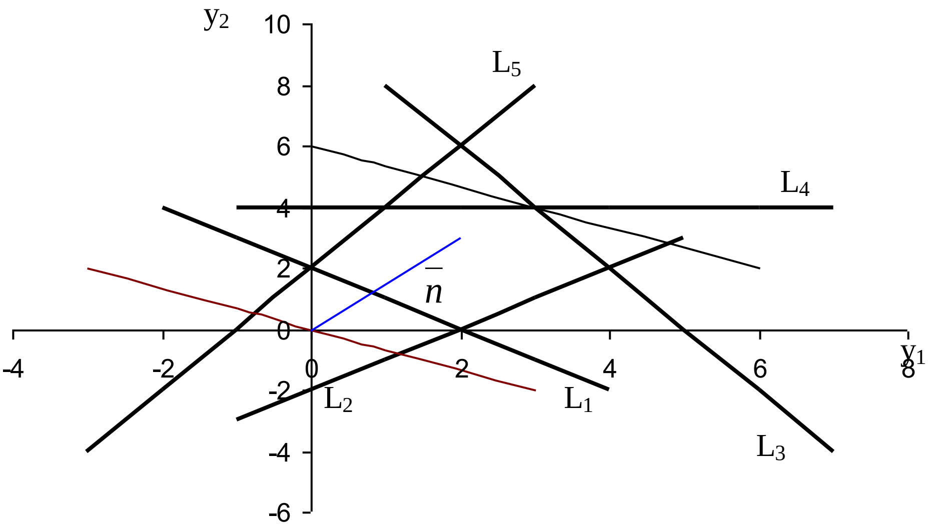


Рис. 1.9 Область допустимых решений

Подставим оптимальное решение $Y^* = (3,4)$ в систему ограничений. Получим, что первое, второе и пятое ограничения выполняются как строгие неравенства:

$$\begin{cases} -3-4<-2 \Rightarrow x_1^*=0, \\ 3-4<2 \Rightarrow x_2^*=0, \\ 2\cdot 3+4=10, \\ 4=4, \\ -2\cdot 3+4<2 \Rightarrow x_5^*=0. \end{cases}$$

По второй теореме двойственности следует, что соответствующие координаты оптимального решения двойственной задачи, т.е. исходной задачи, равны нулю: $x_1^* = x_2^* = x_5^* = 0$. Учитывая это, из системы ограничений исходной задачи получим

$$\begin{cases} 2x_3 = 2, \\ x_3 + x_4 = 3. \end{cases}$$

Отсюда находим $x_3^* = 1,\ x_4^* = 2$. Окончательно записываем $X^* = (0,0,1,2,0)$.

Ответ. $\min Z(X) = 18$ при $X^* = (0,0,1,2,0)$.

Транспортная задача линейного программирования

Под названием «транспортная задача» объединяется широкий круг задач с единой математической моделью. Данные задачи относятся к задачам линейного программирования и могут быть решены симплексным методом. Однако матрица системы ограничений транспортной задачи насколько своеобразна, что для ее решения разработаны специальные методы. Эти методы, как и симплексный метод, позволяют найти начальное опорное решение, а затем, улучшая его, получить оптимальное решение.

Формулировка транспортной задачи

Однородный груз сосредоточен у m поставщиков в объемах $a_1, a_2, ..., a_m$. Данный груз необходимо доставить n потребителям в объемах $b_1, b_2, ..., b_n$. Известны c_{ij}, $i = 1..m$, $j = 1..n$ - стоимости перевозки единицы груза от каждого i - го поставщика каждому j - у потребителю. Требуется составить такой план перевозок, при котором запасы всех поставщиков будут вывезены полностью, запросы всех потребителей полностью удовлетворены и суммарные затраты на перевозку всех грузов минимальны. Исходные данные транспортной задачи обычно записываются в таблице 8.

Таблица 8.

a_i \ b_i	b_1	b_2	…	b_n
a_1	c_{11}	c_{12}	…	c_{1n}
a_2	c_{21}	c_{22}	…	c_{2n}
…	…	…	…	…
a_m	c_{m1}	c_{m2}	…	c_{mn}

Исходные данные задачи могут быть представлены также в виде вектора запасов поставщиков $A = (a_1, a_2, ..., a_m)$, вектора запросов потребителей $B = (b_1, b_2, ..., b_n)$ и матрицы стоимостей

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

В транспортных задачах под поставщиками и потребителями понимаются различные промышленные и сельскохозяйственные предприятия, заводы, фабрики, склады, магазины и т.д. Однородными считаются грузы, которые могут быть перевезены одним видом транспорта.

Под стоимостью перевозок понимаются тарифы, расстояния, время, расход топлива и т.п.

Математическая модель транспортной задачи

Переменными (неизвестными) транспортной задачи являются x_{ij}, $i=1..m,\ j=1..n$ - объемы перевозок каждого i - го поставщика каждому j - у потребителю. Эти переменные можно записать в виде матрицы перевозок

$$X=\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}.$$

Так как произведение $c_{ij}x_{ij}$ определяет затраты на перевозку груза от i - го поставщика каждому j - у потребителю, то суммарные затраты на перевозку всех грузов равны $\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$. По условию задачи требуется обеспечить минимум суммарных затрат. Следовательно, целевая функция имеет вид $Z(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\to\min$. Система ограничений состоит их двух групп уравнений. Первая группа из m уравнений описывает тот факт, что запасы всех m поставщиков вывозятся полностью: $\sum_{j=1}^{n}x_{ij}=a_i,\quad i=1..m$. Вторая группа из n уравнений выражает требование полностью удовлетворить запасы всех n потребителей: $\sum_{i=1}^{m}x_{ij}=b_j,\quad j=1..n$. Учитывая условие неотрицательности объемов перевозок, математическую модель задачи можно записать так:

$$Z(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\to\min \tag{1.30}$$

$$\sum_{j=1}^{n}x_{ij}=a_i,\ i=1..m \tag{1.31}$$

$$\sum_{i=1}^{m}x_{ij}=b_j,\ j=1..n \tag{1.32}$$

$$x_{ij}\ge 0,\ i=1..m,\ j=1..n \tag{1.33}$$

В рассмотренной модели транспортной задачи предполагается, что суммарные запасы поставщиков равны суммарным запасам потребителей, т.е.

$$\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j \tag{1.34}$$

Такая задача называется *задачей с правильным балансом*, а ее модель – *закрытой*. Если же это равенство не выполняется, то задача называется *задачей с неправильным балансом*, а ее модель – *открытой.*

Математическая формулировка транспортной задачи такова: найти переменные $X=\left(x_{ij}\right)$, $i=1..m$, $j=1..n$, удовлетворяющие системе ограничений (1.31), (1.32), условиям неотрицательности (1.33) и обеспечивающие минимум целевой функции (1.30). Математическая модель транспортной задачи может быть записана в *векторном* виде. Для этого рассмотрим матрицу A системы уравнений - ограничений задачи (1.31), (1.32):

$$\begin{array}{c} \begin{array}{cccccc} x_{11} & x_{12}\dots x_{1n} & x_{21} & x_{22}\dots x_{2n}\dots & x_{m1} & x_{m2}\dots x_{mn} \end{array} \\ A=\begin{pmatrix} 1 & 1\dots1 & 1 & 0\dots\dots0\dots & 0 & 0\dots0 \\ 0 & 0\dots0 & 1 & 1\dots\dots1\dots & 0 & 0\dots0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0\dots0 & 0 & 0\dots0\dots & 1 & 1\dots1 \\ 1 & 0\dots0 & 1 & 0\dots0\dots & 1 & 0\dots0 \\ 0 & 1\dots0 & 0 & 1\dots0\dots & 0 & 1\dots0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0\dots1 & 0 & 0\dots1\dots & 0 & 0\dots1 \end{pmatrix} \end{array} \tag{1.35}$$

Сверху над каждым столбцом матрицы указана переменная задачи, коэффициентами при которой являются элементы соответствующего столбца в уравнениях системы ограничений.

Каждый столбец матрицы A, соответствующий переменной x_{ij}, является вектором – условием задачи и обозначается через A_{ij}. Каждый вектор имеет всего $m+n$ координат, и только две из них, отличные от нуля, равны единице. Первая единица вектора A_{ij} стоит на i - м месте, а вторая – на $\left(m+j\right)$ - м месте.

Обозначим через $\grave{A}_0$ вектор ограничений (правых частей уравнений (1.30), (1.31)) и представим систему ограничений задачи в векторном виде. Тогда математическая модель транспортной задачи запишется следующим образом:

$$Z\left(X\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\rightarrow min \tag{1.36}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}x_{ij}=A_0 \tag{1.37}$$

$$x_{ij}\geq 0,\ i=1..m,\ j=1..n \tag{1.38}$$

Пример. Составить математическую модель транспортной задачи, исходные данные которой приведены в таблице 9.

Таблица 9.

a_i \ b_j	20	30	40
40	3	5	7
50	4	6	10

Решение. Введем переменные задачи (матрицу перевозок)
$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{12} \end{pmatrix}.$

Запишем матрицу стоимостей $C = \begin{pmatrix} 3 & 5 & 7 \\ 4 & 6 & 10 \end{pmatrix}.$

Целевая функция задачи равна сумме произведений всех соответствующих элементов матриц $\tilde{N}$ и X:
$Z(X) = 3x_{11} + 5x_{12} + 7x_{13} + 4x_{21} + 6x_{22} + 10x_{23}.$

Данная функция, определяющая суммарные затраты на все перевозки, должна достигать минимального значения.

Составим систему ограничений задачи. Сумма всех перевозок, стоящих в первой строке матрицы X. Должна равняться запасам 1-го поставщика, а сумма перевозок во второй строке матрицы X - запасам 2-го поставщика:

$x_{11} + x_{12} + x_{13} = 40,$

$x_{21} + x_{22} + x_{23} = 50.$

Это означает, что запасы поставщиков вывозятся полностью. Суммы перевозок, стоящих в каждом столбце матрицы X, должны быть равны запросам потребителей:

$x_{11} + x_{12} = 20,$

$x_{12} + x_{22} = 30,$

$x_{13} + x_{23} = 40.$

Это означает, что запросы потребителей удовлетворятся полностью. Необходимо также учитывать, что перевозки не могут быть отрицательными: $x_{ij} \geq 0, \quad i = 1..m, \quad j = 1..n$. Следовательно, математическая модель рассматриваемой задачи такова: найти переменные задачи, обеспечивающие минимум функции

$$Z(X) = 3x_{11} + 5x_{12} + 7x_{13} + 4x_{21} + 6x_{22} + 10x_{23}$$

и удовлетворяющие системе ограничений

$$\begin{cases} x_{11} + x_{12} + x_{13} = 40, \\ x_{21} + x_{22} + x_{23} = 50, \\ x_{11} + x_{21} = 20, \\ x_{12} + x_{22} = 30, \\ x_{13} + x_{23} = 40. \end{cases}$$

и условиям неотрицательности $x_{ij} \geq 0,\ i = 1..m,\ j = 1..n$.

Теорема (необходимое и достаточное условия разрешимости транспортной задачи). Для того чтобы транспортная задача линейного программирования имела решение, необходимо и достаточно, чтобы суммарные запасы поставщиков равнялись суммарным запросам потребителей $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, т.е. задача должна быть с правильным балансом.

Теорема (свойство системы ограничений транспортной задачи). Ранг системы векторов – условий транспортной задачи равен $N = m + n - 1$.

Опорное решение транспортной задачи

Опорным решением транспортной задачи называется любое допустимое решение, для которого векторы – условия, соответствующие положительным координатам, линейно независимы.

Так как ранг системы векторов условий транспортной задачи равен $N = m + n - 1$, опорное решение не может иметь отличных от нуля координат более $m + n - 1$. Число отличных от нуля координат невырожденного опорного решения равно $m + n - 1$, а для вырожденного опорного решения меньше $m + n - 1$. Любое допустимое решение транспортной задачи можно записать в ту же таблицу, что и исходные данные. Клетки таблицы транспортной задачи, в которых находятся отличные от нуля или базисные нулевые перевозки, называются *занятыми*, остальные – *незанятыми* или *свободными.* Клетки таблицы нумеруются так, что клетка, содержащая перевозку x_{ij}, т.е. стоящая в i - ой строке и j - м столбце, имеет номер (i, j). Каждой клетке с номером (i, j) соответствует переменная x_{ij}, которой соответствует вектор – условие A_{ij}.

Для того чтобы избежать трудоемких вычислений при проверке линейной независимости векторов – условий, соответствующих положительным координатам допустимого решения, вводят понятие цикла. Циклы используются для перехода от одного опорного решения к другому.

Циклом называется такая последовательность клеток таблицы транспортной задачи (i_1, j_1), (i_1, j_2), (i_2, j_2), …, (i_{1k}, j_1), в которой две и только две соседние клетки расположены в одной строке или столбце, причем первая и последняя клетки также находятся в одной строке или столбце.

Цикл изображают в таблице транспортной задачи в виде замкнутой ломаной линии. В любой клетке цикла происходит поворот звена ломаной линии на 90^0. Простейшие циклы изображены на рисунке 1.10, где звездочкой отмечены клетки таблицы, включенные в состав цикла

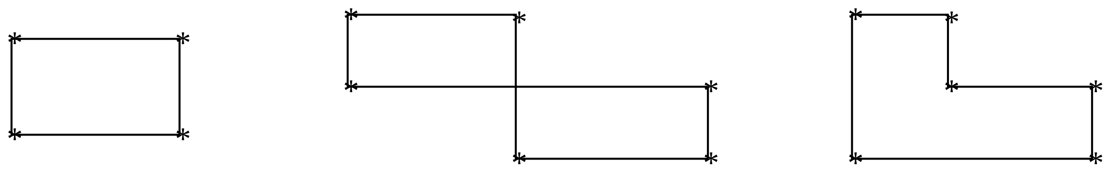


Рис. 1.10. Некоторые виды циклов

Теорема (о взаимосвязи линейной зависимости векторов – условий и возможности образования цикла). Для того чтобы система векторов - условий транспортной задачи была линейно - зависимой, необходимо и достаточно, чтобы из соответствующих клеток таблицы можно было выделить часть, которая образует цикл.

Следствие. Допустимое решение транспортной задачи $X=(x_{ij})$, $i=1..m,\ j=1..n$ является опорным тогда и только тогда, когда из занятых им клеток нельзя образовать ни одного цикла.

Метод вычеркивания

Метод вычеркивания позволяет проверить, является ли данное решение транспортной задачи опорным.

Пусть допустимое решение транспортной задачи, которое имеет $m+n-1$ отличную от нуля координату, записано в таблицу. Чтобы данное решение было опорным, векторы – условия, соответствующие положительным координатам, должны быть линейно – независимыми. Для этого занятые решением клетки таблицы должны быть расположены так, чтобы из них нельзя было образовать цикл.

Строка или столбец таблицы с одной занятой клеткой не может входить в какой-либо цикл, так как цикл имеет две и только две клетки в каждой строке или в столбце. Следовательно, можно вычеркнуть сначала либо все строки таблицы, содержащие по одной занятой клетке, либо все столбцы, содержащие по одной занятой клетке, далее вернуться к столбцам (строкам) и продолжить их вычеркивание. Если в результате вычеркиваний все строки и столбцы будут вычеркнуты, то из занятых клеток таблицы нельзя выделить часть, образующую цикл, и система соответствующих векторов – условий линейно – независима, а решение является опорным. Если же после вычеркивания останется часть клеток и эти клетки образуют цикл, то система соответствующих векторов – условий линейно – зависима, а решение не является опорным.

Ниже приведены примеры «вычеркиваемого» (опорного) и «невычеркиваемого» (неопорного) решений:

$$X = \begin{pmatrix} 5 & 0 & 0 & 15 \\ 0 & 9 & 2 & 0 \\ 0 & 0 & 7 & 1 \end{pmatrix} \qquad\qquad X = \begin{pmatrix} 5 & 6 & 0 & 9 \\ 0 & 3 & 8 & 0 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

«вычеркиваемое» «невычеркиваемое»

Методы построения начального опорного решения

Метод северо-западного угла

Существует ряд методов построения начального опорного решения, наиболее простым из которых является метод северо-западного угла. В данном методе запасы очередного поставщика используются для обеспечения запросов очередных потребителей до тех пор, пока не будут исчерпаны полностью, после чего используются запасы следующего по номеру поставщика.

Заполнение таблицы транспортной задачи начинается с левого верхнего угла и состоит из ряда однотипных шагов. На каждом шаге, исходя из запасов очередного поставщика и запросов очередного потребителя, заполняется только одна клетка и соответственно исключается из рассмотрения один поставщик или потребитель. Осуществляется это следующим образом:

1) если $a_i < b_j$, то $x_{ij} = a_i$ и исключается поставщик с номером i, $x_{ik} = 0$, $k = 1..n$, $k \neq j$, $b_j' = b_j - a_i$;

2) если $a_i > b_j$, то $x_{ij} = b_j$ и исключается потребитель с номером j, $x_{kj} = 0$, $k = 1..m$, $k \neq i$, $a_i' = a_i - b_j$;

3) если $a_i = b_j$, то $x_{ij} = a_i = b_j$ и исключается либо i поставщик $x_{ik} = 0$, $k = 1..n$, $k \neq j$, $b_j' = 0$, либо j потребитель, $x_{kj} = 0$, $k = 1..m$, $k \neq i$, $a_i' = 0$.

Нулевые перевозки принято заносить в таблицу только тогда, когда они попадают в клетку (i, j), подлежащую заполнению. Если в очередную клетку таблицы (i, j) требуется поставить перевозку, а i-й поставщик или j-й потребитель имеет нулевые запасы или запросы, то в клетку ставится перевозка, равная нулю (базисный нуль), и после этого из рассмотрения исключается соответствующий поставщик или потребитель. В таблицу заносим только базисные нули, остальные клетки с нулевыми перевозками остаются пустыми.

Чтобы избежать ошибок после построения начального опорного решения необходимо проверить, что число занятых клеток равно $m + n - 1$ и векторы – условия, соответствующие этим клеткам, линейно – независимы.

Теорема. Решение транспортной задачи, построенное методом северо-западного угла, является опорным.

Замечание. Метод северо-западного угла не учитывает стоимость перевозок, поэтому опорное решение, построенное данным методом, может быть далеко от оптимального.

Пример. Составить начальное опорное решение, используя метод северо-западного угла, для транспортной задачи, исходные данные которой приведены в таблице 10.

Таблица 10.

a_i \ b_i	150	200	100	100
100	1	3	4	2
250	4	5	8	3
200	2	3	6	7

Решение. Распределяем запасы первого поставщика. Так как его запасы $a_1 = 100$ меньше запросов первого потребителя $b_1 = 150$, то в клетку $(1{,}1)$ записываем перевозку $x_{11} = 100$ и исключаем из рассмотрения поставщика. Определяем оставшиеся неудовлетворенными запросы 1-го потребителя $b_1' = b_1 - a_1 = 150 - 100 = 50$.

Распределяем запасы второго поставщика. Так как его запасы $a_2 = 250$ больше оставшихся неудовлетворенными запросов 1-го потребителя $b_1' = 50$, то в клетку $(2{,}1)$ записываем перевозку $x_{21} = 50$ и исключаем из рассмотрения 1-го потребителя. Определяем оставшиеся запасы 2-го поставщика $a_2' = a_2 - b_1' = 250 - 50 = 200$. Так как $a_2' = b_2 = 200$, то в клетку $(2{,}2)$ записываем $x_{22} = 200$ и исключаем по своему усмотрению либо 2-го поставщика, либо 2-го потребителя. Пусть исключили 2-го поставщика. Вычисляем оставшиеся неудовлетворенными запросы 2-го потребителя $b_2' = b_2 - a_2' = 200 - 200 = 0$.

Распределяем запасы 3-го поставщика. Так как $a_3 > b_2'$ $(200 > 0)$, то в клетку $(3{,}2)$ записываем $x_{32} = 0$ и исключаем 2-го потребителя. Запасы 3-го поставщика не изменились $a_3' = a_3 - b_2' = 200 - 0 = 200$. Сравниваем a_3' и b_3 $(200 > 100)$, в клетку $(3{,}3)$ записываем $x_{33} = 100$, исключаем 3-го потребителя и вычисляем $a_3'' = a_3' - b_3 = 200 - 100 = 100$. Так как $a_3'' = b_4$, то в клетку $(3{,}4)$ записываем $x_{34} = 100$. Так как задача с правильным балансом запасы всех поставщиков исчерпаны и запросы всех потребителей удовлетворены полностью и одновременно. Результаты построения опорного решения приведены в таблице 11. Проверяем правильность построения опорного решения. Число занятых клеток должно быть равно $N = m + n - 1 = 3 + 4 - 1 = 6$. В таблице 11 занято шесть клеток. Применяя метод вычеркивания, убеждаемся, что данное решение является «вычеркиваемым»:

$$X = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 50 & 200 & 0 & 0 \\ 0 & 0^* & 100 & 0 \end{pmatrix}$$ (звездочкой отмечен базисный нуль).

Следовательно, векторы – условия, соответствующие занятым клеткам, линейно - независимы и построенное решение является опорным.

Таблица 11.

a_i \ b_i	150	200	100	100
100	1 100	3	4	2
250	4 50	5 200	8	3
200	2	3 0	6 100	7 100

Метод минимальной стоимости

Метод минимальной стоимости прост, он позволяет построить опорное решение, достаточно близкое к оптимальному, так как использует матрицу стоимостей транспортной задачи $C = (c_{ij})$, $i = 1..m$, $j = 1..n$. Как и метод северо-западного угла, он состоит из ряда однотипных шагов, на каждом из которых заполняется только одна клетка таблицы, соответствующая минимальной стоимости $\min_{i,j}\{c_{ij}\}$, и исключается из рассмотрения только одна строка (поставщик) или один столбец (потребитель). Очередную клетку, соответствующую $\min_{i,j}\{c_{ij}\}$, заполняют по тем же правилам, что и в методе северо-западного угла. Поставщик исключается из рассмотрения, если его запасы удовлетворены полностью. На каждом шаге исключается либо один поставщик, либо один потребитель. При этом если поставщик еще не исключен, но его запасы равны нулю, то на том шаге, когда от данного поставщика требуется поставить груз, в соответствующую клетку таблицы заносится базисный нуль и затем поставщик исключается из рассмотрения. Аналогично с потребителем.

Теорема. Решение транспортной задачи, построенное методом минимальной стоимости, является опорным.

Пример. Используя метод минимальной стоимости, построить начальное опорное решение транспортной задачи, исходные данные которой приведены в таблице 12.

Таблица 12.

b_i / a_i	40	60	80	60
60	1	3	4	2
80	4	5	8	3
100	2	3	6	7

Решение. Запишем отдельно матрицу стоимостей для того, чтобы удобнее было выбирать минимальные стоимости, вычеркивать строки и столбцы:

$$C=\begin{array}{c}\begin{pmatrix} \underline{1} & 3 & 4 & \underline{2} \\ 4 & 5 & 8 & \underline{3} \\ 2 & \underline{3} & \underline{6} & 7 \end{pmatrix}\begin{array}{c} 2 \\ 6 \\ 5 \end{array} \\ \begin{array}{cccc} 1 & 4 & 6 & 3 \end{array}\end{array}$$

Среди элементов матрицы стоимостей выбираем наименьшую стоимость $c_{11}=1$ и подчеркиваем ее. Это стоимость перевозки груза от 1-го поставщика 1-му потребителю. В соответствующую клетку $(1,1)$ записываем максимально возможный объем перевозки $x_{11}=min\{a_1,b_1\}=min\{60,40\}=40$ (таблица 13).

Таблица 13.

b_i / a_i	40	60	80	60
60	1 40	3	4	2 20
80	4	5	8 40	3 40
100	2	3 60	6 40	7

Запасы 1-го поставщика уменьшаем на 40, т.е. $a_1^{'}=a_1-b_1=60-40=20$. Исключаем из рассмотрения 1-го потребителя, так как его запросы удовлетворены. В матрице C вычеркиваем 1 столбец.

В оставшейся части матрицы C минимальной является стоимость $c_{14}=2$. Максимально возможная перевозка, которую можно осуществить от 1-го поставщика к 4-му потребителю, равна $x_{14}=min\{a_1^{'},b_4\}=min\{20,60\}=20$. В соответствующую клетку таблицы записываем перевозку $x_{14}=20$. Запасы 1-го поставщика исчерпаны, исключаем его из рассмотрения. В матрице C вычеркиваем 1 строку. Запросы 4-го потребителя уменьшаем на 20, т.е. $b_4^{'}=b_4-a_1^{'}=60-20=40$.

В оставшейся части матрицы C минимальная стоимость $c_{24} = c_{32} = 3$. Заполняем одну из двух клеток таблицы $(2{,}4)$ или $(3{,}2)$. Пусть в клетку $(2{,}4)$ записываем $x_{24} = min\{a_2, b_4\} = min\{80{,}40\} = 40$. Запросы 4-го потребителя удовлетворены, исключаем его из рассмотрения, вычеркиваем четвертый столбец в матрице C. Уменьшаем запасы 2-го поставщика $a_2^{'} = a_2 - b_4 = 80 - 40 = 40$.

В оставшейся части матрицы C минимальная стоимость $c_{32} = 3$. Записываем в клетку таблицы $(3{,}2)$ перевозку $x_{32} = min\{a_3, b_2\} = min\{100{,}60\} = 60$. Исключаем из рассмотрения 2-го потребителя, а из матрицы C второй столбец. Вычисляем $a_3^{'} = a_3 - b_2 = 100 - 60 = 40$.

В оставшейся части матрицы C минимальная стоимость $c_{33} = 6$. Записываем в клетку таблицы $(3{,}3)$ перевозку $x_{33} = min\{a_3, b_3\} = min\{40{,}80\} = 40$. Исключаем из рассмотрения 3-го потребителя, а из матрицы C третью строку. Определяем $b_3^{'} = b_3 - a_3^{'} = 80 - 40 = 40$.

В матрице C остается единственный элемент $c_{23} = 8$. Записываем в клетку таблицы $(2{,}3)$ перевозку $x_{23} = 40$.

Проверяем правильность построения опорного решения. Число занятых клеток таблицы равно $N = m + n - 1 = 3 + 4 - 1 = 6$. Методом вычеркивания проверяем линейную независимость векторов – условий, соответствующих положительным координатам решения. Порядок вычеркивания показан на матрице X:

$$X = \begin{array}{cl} \begin{pmatrix} 40 & 0 & 0 & 20 \\ 0 & 0 & 40 & 40 \\ 0 & 60 & 40 & 0 \end{pmatrix} & \begin{array}{l} 3 \\ \\ 4 \end{array} \\ \begin{array}{cccc} 1 & 2 & 5 & 6 \end{array} & \end{array}$$

Решение является «вычеркиваемым» и, следовательно, опорным.

Переход от одного опорного решения к другому

В транспортной задаче переход от одного опорного решения к другому осуществляется с помощью цикла. Для некоторой свободной клетки таблицы строится цикл, содержащий часть клеток, занятых опорным решением. По этому циклу перераспределяются объемы перевозок. Перевозка загружается в выбранную клетку и освобождается одна из занятых клеток, получается новое опорное решение.

Теорема (о существовании и единственности цикла). Если таблица транспортной задачи содержит опорное решение, то для любой свободной клетки таблицы существует единственный цикл, содержащий эту клетку и часть клеток, занятых опорным решением.

Цикл называется *означенным*, если его угловые клетки пронумерованы по порядку и нечетным клеткам приписан знак «+», а четным – знак «-» (Рис. 1.11).

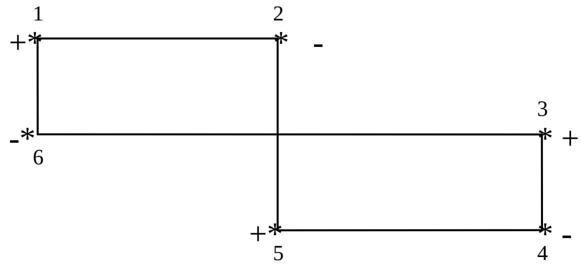


Рис. 1.11 Означенный цикл

Сдвигом по циклу на величину θ называется увеличение объемов перевозок во всех нечетных клетках цикла, отмеченных знаком «+», на θ и уменьшение объемов перевозок во всех четных клетках, отмеченных знаком «-», на θ.

Теорема. Если таблица транспортной задачи содержит опорное решение, то при сдвиге по любому циклу, содержащему одну свободную клетку, на величину $\theta = \min\limits_{"-"}\{x_{ij}\}$ получится опорное решение.

Распределительный метод

Один из наиболее простых методов решения транспортных задач - распределительный метод.

Пусть для транспортной задачи найдено начальное опорное решение X_1 и вычислено значение целевой функции на этом решении $Z(X_1)$. По теореме о существовании и единственности цикла для каждой свободной клетки таблицы задачи можно построить единственный цикл, который содержит эту клетку и часть клеток, занятых опорным решением. Означив этот цикл и осуществив сдвиг (перераспределение груза) по циклу на величину $\theta = \min\limits_{"-"}\{x_{ij}\}$, можно получить новое опорное решение X_2.

Определим, как изменится целевая функция при переходе к новому опорному решению. При сдвиге на единицу груза по циклу, соответствующему клетке (l,k) приращение целевой функции Δ_{lk} равно разности двух сумм:

$$\Delta_{lk} = \sum_{"+"} c_{ij} - \sum_{"-"} c_{ij} \ ,$$

где $\sum\limits_{"+"} \tilde{n}_{ij}$ - сумма стоимостей перевозок единиц груза в нечетных клетках цикла, отмеченных знаком «+»;

$\sum\limits_{"-"}\tilde{n}_{ij}$ - сумма стоимостей перевозок единиц груза в четных клетках цикла, отмеченных знаком «-».

В клетках, отмеченных знаком «+», величины груза прибавляются, что приводит к увеличению значения целевой функции $Z(X)$, а в клетках, отмеченных знаком «-», величины груза уменьшаются, что приводит к уменьшению значения целевой функции.

Если разность сумм для свободной клетки (l,k) меньше нуля, т.е. $\Delta_{lk} < 0$, то перераспределение величины θ по соответствующему циклу приведет к уменьшению значения $Z(X)$ на величину $\theta \cdot \Delta_{lk}$, т.е. опорное решение можно улучшить. Если же величина Δ_{lk}, называемые *оценками*, для всех свободных клеток таблицы транспортной задачи неотрицательны, то значение целевой функции нельзя уменьшить и опорное решение оптимально. Следовательно, *признаком оптимальности* распределительного метода является условие

$$\Delta_{lk} \geq 0 \qquad \forall x_{lk} = 0 \tag{1.39}$$

Для решения транспортной задачи распределительным методом необходимо найти начальное опорное решение. Затем для очередной опорной клетки (l,k) построить цикл и вычислить оценку Δ_{lk}. Если оценка неотрицательная, переходят к следующей свободной клетке. Если же оценка отрицательная, следует осуществить сдвиг по циклу на величину $\theta = \min\limits_{"-"}\{x_{ij}\}$. В результате получится новое опорное решение.

Для каждого нового опорного решения вычисление оценок начинается с первой свободной клетки таблицы. Очередность проверяемых свободных клеток целесообразно устанавливать в порядке возрастания стоимости перевозок c_{ij}, так как решается задача на нахождение минимума.

Пример. Решить распределительным методом транспортную задачу, исходные данные которой приведены в таблице 14.

Таблица 14.

a_i \ b_i	20	40	40
20	1	3	2
30	4	5	7
50	6	8	15

Решение. Строим начальное опорное решение методом минимальной стоимости (таблица 15).

Таблица 15.

a_i \ b_i	20	40	40

20	1 20	3 +	2 0 -
30	4	5 30	7
50	6	8 10 -	15 40 +

Затем вычисляем значение целевой функции на нем:

$Z(X_1) = 20 \cdot 1 + 30 \cdot 5 + 10 \cdot 8 + 40 \cdot 15 = 850$

Находим цикл для свободной клетки $(1,2)$ таблицы, он включает клетки $(1,2)$, $(1,3)$, $(3,3)$, $(3,2)$. Вычисляем оценку $\Delta_{12} = (3+15) - (2+8) = 8$. Так как $\Delta_{12} = 8 > 0$ переходим к следующей свободной клетке $(2,1)$. Для нее цикл таков: $(2,1)$, $(1,1)$, $(1,3)$, $(3,3)$, $(3,2)$, $(2,2)$ (таблица 16). Оценка $\Delta_{21} = (4+2+8) - (1+15+5) = 14 - 21 = -7$. Так как $\Delta_{21} = -7 < 0$, определяем величину груза, перераспределяемого по циклу, $\theta = \min\limits_{"-"}\{20,40,30\} = 20$. Приращение целевой функции $\Delta Z = -7 \cdot 20 = -140$. Получаем новое опорное решение X_2 (таблица 17). Значение целевой функции на нем $Z(X_2) = 20 \cdot 2 + 20 \cdot 4 + 10 \cdot 5 + 30 \cdot 8 + 20 \cdot 15 = 710$

Таблица 16.

$a_i \backslash b_i$	20	40	40
20	1 20 -	3	2 0 +
30	4 +	5 30 -	7
50	6	8 10 +	15 40 -

Таблица 17.

$a_i \backslash b_i$	20	40	40
20	1	3	2 20
30	4 20	5 10 -	7 +
50	6	8 30 +	15 20 -

Вычисляем

$\Delta_{11} = (1+15+5) - (2+8+4) = 7 > 0$, $\Delta_{12} = (3+15) - (2+8) = 8 > 0$,

$\Delta_{23} = (7+8) - (5+15) = -5 < 0$, $\Delta_{31} = (6+5) - (8+4) = -1 < 0$.

Оценки можно вычислять до первой отрицательной. Так как $\Delta_{23} = -5 < 0$, осуществляем сдвиг по циклу $(2,3)$, $(3,3)$, $(3,2)$, $(2,2)$ на величину $\theta=\min\limits_{"-"}\{10,20\}=10$. Приращение целевой функции $\Delta Z=-5\cdot 10=-50$. Получаем третье опорное решение X_3 (таблица 18). Значение целевой функции на нем $Z(X_3)=20\cdot 2+20\cdot 4+10\cdot 7+40\cdot 8+10\cdot 15=660$.

Таблица 18.

a_i \ b_i	20	40	40
20	1	3	2 20
30	4 20 -	5	7 10 +
50	6 +	8 40 +	15 10 -

Вычисляем оценки для свободных клеток:

$\Delta_{11}=(1+7)-(2+4)=2>0$, $\Delta_{12}=(3+15)-(2+8)=8>0$,

$\Delta_{22}=(5+15)-(7+8)=5>0$, $\Delta_{31}=(6+7)-(4+15)=-6<0$.

Так как $\Delta_{31}=-6<0$, осуществляем сдвиг по циклу $(3,1)$, $(2,1)$, $(2,3)$, $(3,3)$ на величину $\theta=\min\limits_{"-"}\{20,10\}=10$. Приращение целевой функции $\Delta Z=-6\cdot 10=-60$. Получаем четвертое опорное решение X_4 (таблица 19). Значение целевой функции на нем

$Z(X_4)=20\cdot 2+10\cdot 4+20\cdot 7+10\cdot 6+40\cdot 8=600$.

Таблица 19.

a_i \ b_i	20	40	40
20	1	3	2 20
30	4 10 -	5 +	7 20
50	6 10 +	8 40 -	15

Вычисляем оценки для свободных клеток:

$\Delta_{11}=(1+7)-(2+4)=2>0$, $\Delta_{12}=(3+7+6)-(2+4+8)=2>0$,

$\Delta_{22}=(5+6)-(4+8)=-1<0$. Так как $\Delta_{22}=-1<0$, осуществляем сдвиг по циклу $(2,2)$, $(3,2)$, $(3,1)$, $(2,1)$ на величину $\theta=\min\limits_{"-"}\{10,40\}=10$. Приращение целевой функции $\Delta Z=-1\cdot 10=-10$. Получаем четвертое опорное решение X_5 (таблица 20). Значение целевой функции на нем

$Z(X_5)=20\cdot 2+10\cdot 5+20\cdot 7+20\cdot 6+30\cdot 8=590$

Таблица 20.

b_i \ a_i	20	40	40
20	1	3	2 20
30	4 0	5 10	7 20
50	6 20	8 30	15

Вычисляем оценки для свободных клеток:

$\Delta_{11} = (1+7)-(2+4) = 2 > 0$, $\Delta_{12} = (3+7)-(2+5) = 3 > 0$,

$\Delta_{33} = (5+15)-(7+8) = 5 > 0$. Все оценки для свободных клеток положительные, следовательно, решение оптимально.

Ответ. $min\,Z(X) = 590$ при $X = \begin{pmatrix} 0 & 0 & 20 \\ 0 & 10 & 20 \\ 20 & 30 & 0 \end{pmatrix}$.

Метод потенциалов

Широко распространенным методом решения транспортных задач является метод потенциалов. Этот метод позволяет упростить наиболее сложную часть вычислений – нахождение оценок свободных клеток.

Теорема (признак оптимальности опорного решения). Если допустимое решение $X = (x_{ij})$, $i = 1..m$, $j = 1..n$ транспортной задачи является оптимальным, то существуют потенциалы (числа) поставщиков u_i, $i = 1..m$ и потребителей v_j, $j = 1..n$, удовлетворяющие следующим условиям:

$$u_i + v_j = c_{ij} \text{ при } x_{ij} > 0, \tag{1.40}$$

$$u_i + v_j \le c_{ij} \text{ при } x_{ij} = 0. \tag{1.41}$$

Группа равенств (40) $u_i + v_j = c_{ij}$ при $x_{ij} > 0$ используется как система уравнений для нахождения потенциалов. Эта система может иметь и другой вид, например $-u_i + v_j = c_{ij}$ или $u_i - v_j = c_{ij}$, если перед тем, как записать двойственную задачу, все уравнения одной из групп уравнений исходной задачи умножить на (-1).

Данная система уравнений имеет $m+n$ неизвестных u_i, $i = 1..m$ и v_j, $j = 1..n$. Число уравнений системы, как и число отличных от нуля координат невырожденного опорного решения, равно $m+n-1$. Так как число неизвестных системы на единицу больше числа уравнений, то одной из них можно задать значение произвольно, а остальные найти из системы.

Группа неравенств (1.41) $u_i + v_j \le c_{ij}$ при $x_{ij} = 0$ используется для проверки оптимальности опорного решения. Эти неравенства удобно записать в виде

$$\Delta_{ij} = u_i + v_j - c_{ij} \text{ при } x_{ij} = 0 \qquad (1.42)$$

Числа Δ_{ij} называются *оценками* свободных клеток таблицы или векторов условий транспортной задачи, не входящих в базис опорного решения. В этом случае *признак оптимальности* можно сформулировать так же, как в симплексном методе (для задачи на минимум): опорное решение является оптимальным, если для всех векторов – условий (клеток таблицы) оценки положительные.

Оценки для свободных клеток транспортной таблицы используются для улучшения опорного решения. С этой целью находят клетку (l,k) таблицы, соответствующую $max\{\Delta_{ij}\} = \Delta_{lk}$. Если $\Delta_{lk} \le 0$, то решение оптимальное. Если же $\Delta_{lk} > 0$, то для клетки (l,k) строят цикл и улучшают решение, перераспределяя груз $\theta = \underset{''-''}{min}\{x_{ij}\}$ по этому циклу.

Особенности решения транспортных задач с неправильным балансом

Мы рассматривали задачи с правильным балансом, но на практике чаще встречаются задачи с неправильным балансом. Каковы особенности их решения?

1. Пусть суммарные запасы поставщиков превосходят суммарные запросы потребителей, т.е. $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$. В этом случае при составлении оптимального плана перевозок часть запасов поставщиков, равная $b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$ останется не вывезенной. Поэтому в системе ограничений транспортной задачи первую группу уравнений (1.40) следует заменить неравенствами

$$\sum_{j=1}^{n} x_{ij} \le a_i,\ i = 1..m. \qquad (1.43)$$

Вторая группа уравнений остается без изменения, так как запросы всех потребителей удовлетворяются полностью. Для приведения к канонической форме в неравенстве (1.43) вводят дополнительные переменные $x_{1(n+1)}, x_{2(n+1)}, \ldots, x_{m(n+1)}$. В результате первые m ограничений задачи принимают вид

$$\sum_{j=1}^{n} x_{ij} + x_{i(n+1)} = a_i,\ i = 1..m$$

В целевую функцию дополнительные переменные не входят (входят с нулевыми коэффициентами). Математическая модель задачи принимает вид

$$Z(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}+\sum_{i=1}^{m}0\cdot x_{i(n+1)}\to min\ , \tag{1.44}$$

$$\sum_{j=1}^{n}x_{ij}+x_{i(n+1)}=a_i\ ,\ i=1..m \tag{1.45}$$

$$\sum_{i=1}^{m}x_{ij}=b_j\ ,\ j=1..n \tag{1.46}$$

$$x_{ij}\geq 0,\ i=1..m,\ j=1..n+1 \tag{1.47}$$

Запишем необходимое и достаточное условие разрешимости задачи:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\sum_{i=1}^{m}\left(a_i-x_{i(n+1)}\right)=\sum_{j=1}^{n}b_j\ .$$

Отсюда

$$\sum_{i=1}^{m}a_i-\sum_{i=1}^{m}x_{i(n+1)}=\sum_{j=1}^{n}b_j\Rightarrow\sum_{i=1}^{m}x_{i(n+1)}=\sum_{i=1}^{m}a_i-\sum_{j=1}^{n}b_j=b_{n+1}\ .$$

Чтобы задача в рассматриваемом случае имела решение, необходимо ввести фиктивного потребителя с запасами b_{n+1}, равными разности суммарных запасов поставщиков и запросов потребителей, и нулевыми стоимостями перевозок единиц груза $c_{i(n+1)}=0\quad\forall i$.

2. В случае, когда суммарные запросы потребителей превосходят суммарные запасы поставщиков, поступаем аналогично, т.е. $\sum_{i=1}^{m}a_i<\sum_{j=1}^{n}b_j$, часть запросов потребителей равная $a_{m+1}=\sum_{j=1}^{n}b_j-\sum_{i=1}^{m}a_i$, останется не удовлетворенной. Поэтому вторая группа уравнений системы ограничений (1.32) транспортной задачи заменяется неравенствами $\sum_{i=1}^{m}x_{ij}\leq b_j\ ,\quad j=1..n$. После введения дополнительных переменных $x_{(m+1)1},x_{(m+1)2},\ldots,x_{(m+1)n}$ в эти неравенства математическая модель задачи примет вид

$$Z(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}+\sum_{j=1}^{n}0\cdot x_{(m+1)j}\to min \tag{1.48}$$

$$\sum_{j=1}^{n}x_{ij}=a_i\ ,\ i=1..m \tag{1.49}$$

$$\sum_{i=1}^{m}x_{ij}+x_{(m+1)j}=b_j\ ,\ j=1..n \tag{1.50}$$

$$x_{ij}\geq 0,\ i=1..m+1,\ j=1..n \tag{1.51}$$

Для того чтобы задача имела решение, необходимо и достаточно, чтобы

$$\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\sum_{j=1}^{n}\left(b_j-x_{(m+1)j}\right)=\sum_{i=1}^{m}a_i\ .$$

Отсюда

$$\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j-\sum_{j=1}^{n}x_{(m+1)j}\Rightarrow\sum_{j=1}^{n}x_{(m+1)j}=\sum_{j=1}^{n}b_j-\sum_{i=1}^{m}a_i=a_{m+1}.$$

Чтобы в этом случае задача имела решение, необходимо ввести фиктивного поставщика с запасами a_{m+1}, равными разности суммарных запросов потребителей и запасов поставщиков, и нулевыми стоимостями перевозок единиц груза $c_{(m+1)j} = 0 \quad \forall j$.

Замечание. При составлении начального опорного решения в последнюю очередь следует распределить запасы фиктивного поставщика и удовлетворять запросы фиктивного потребителя, несмотря на то, что им соответствует наименьшая стоимость перевозок, равная нулю.

Алгоритм решения транспортной задачи методом потенциалов

Порядок решения транспортных задач методом потенциалов следующий.

1. Проверяют выполнение необходимого и достаточного условия разрешимости задачи. Если задача имеет неправильный баланс, то вводят фиктивного поставщика или потребителя с недостающими запасами или запросами и нулевыми стоимостями перевозок.

2. Строят начальное опорное решение (методом минимальной стоимости или каким – либо методом) и проверяют правильность его построения, для чего подсчитывают количество занятых клеток (их должно быть $m+n-1$) и убеждаются в линейной независимости векторов – условий (методом вычеркивания).

3. Строят систему потенциалов, соответствующих опорному решению. Для этого решают систему уравнений $u_i + v_j = c_{ij}$ при $x_{ij} > 0$. Для того, чтобы найти частное решение системы, одному из потенциалов (обычно тому, которому соответствует большее число занятых клеток) задают произвольно некоторое значение (чаще нуль). Остальные потенциалы однозначно определяются по формулам:

$$u_i = c_{ij} - v_j \text{ при } x_{ij} > 0, \qquad (1.52)$$

если известен потенциал v_j, и

$$v_j = c_{ij} - u_i \text{ при } x_{ij} > 0 \qquad (1.53)$$

Если известен потенциал u_i.

4. Проверяют, выполняется ли условие оптимальности для свободных клеток таблицы. Для этого вычисляют оценки для всех свободных клеток по формулам $\Delta_{ij} = u_i + v_j - c_{ij}$ и те оценки, которые больше нуля, записывают в левые нижние углы клеток. Если для всех свободных клеток $\Delta_{ij} \le 0$, то вычисляют значение целевой функции, и решение задачи заканчивается, так как полученное решение является оптимальным. Если же имеется хотя бы одна клетка с положительной оценкой, то опорное решение не является оптимальным.

5. Переходят к новому опорному решению, на котором значение целевой функции будет меньше. Для этого находят клетку таблицы задачи, которой соответствует наибольшая положительная оценка $\max_{\Delta_{ij}>0}\{\Delta_{ij}\} = \Delta_{lk}$.

Строят цикл, включающий в свой состав данную клетку и часть клеток, занятых опорным решением. В клетках цикла расставляют поочередно знаки «+» и «-», начиная с «+» в клетке с наибольшей положительной оценкой. Осуществляют сдвиг (перераспределение груза) по циклу на величину $\theta = \min_{"-"}\{x_{ij}\}$. Клетка со знаком «-», в которой достигается $\min_{"-"}\{x_{ij}\}$, остается пустой. Если минимум достигается в нескольких клетках, то одна из них остается пустой, а в остальных проставляют базисные нули, чтобы число занятых клеток оставалось равным $m+n-1$.

Далее возвращаемся к пункту 3 алгоритма.

Пример. Решить транспортную задачу, исходные данные которой приведены в таблице 21.

Решение. 1. Проверяем выполнение необходимого и достаточного условия разрешимости задачи. Находим суммарные запасы поставщиков и запросы потребителей:

$$\sum_{i=1}^{3} a_i = 100 + 200 + 300 = 600,\ \sum_{j=1}^{4} b_j = 100 + 100 + 300 + 300 = 800$$

Таблица 21

b_i / a_i	100	100	300	300
100	1	2	3	1
200	2	3	4	6
300	3	4	7	12

Задача с неправильным балансом. Вводим четвертого, фиктивного поставщика с запасами $a_4 = 800 - 600 = 200$ и нулевыми стоимостями перевозки единиц груза (таблица 22).

Таблица 22

b_i / a_i	100	100	300	300
100	1 100	2	3	1
200	2 0	3 100	4 100	6
300	3	4	7 200	12 100

200	0	0	0	0 200

2. Составляем начальное опорное решение методом минимальной стоимости (таблица 22). Записываем матрицу стоимостей $\tilde{N}$:

$$C=\begin{array}{cc} \begin{pmatrix} \underline{1} & 2 & 3 & 1 \\ \underline{2} & \underline{3} & \underline{4} & 6 \\ 3 & 4 & 7 & 12 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 1 \\ 4 \\ 6 \\ 7 \end{array} \\ \begin{array}{cccc} 2 & 3 & 5 & 7 \end{array} & \end{array}$$

Находим в этой матрице наименьшие на каждом шаге стоимости и направляем в клетку, которая соответствует этим стоимостям, максимально допустимые объемы перевозок грузов. При этом исключаем на каждом шаге одного поставщика или одного потребителя. Подчеркнутые числа в матрице $\tilde{N}$ - минимальные элементы, а цифры рядом со строками и столбцами – порядок исключения из рассмотрения поставщиков и потребителей. Запасы фиктивного поставщика вывозятся в последнюю очередь.

Полученное решение X_1 имеет $m+n-1=4+4-1=7$ базисных переменных. Опорное решение является вырожденным, так как одна из его координат равна нулю. Вычислим значение целевой функции на этом опорном решении $Z(X_1)=100\cdot 1+0\cdot 2+100\cdot 3+100\cdot 4+200\cdot 7+100\cdot 12+200\cdot 0=3400$.

3. Для проверки оптимальности опорного решения необходимо найти потенциалы. По признаку оптимальности в каждой занятой опорным решением клетке таблицы транспортной задачи сумма потенциалов равна стоимости ($u_i+v_j=c_{ij}$ при $x_{ij}>0$). Записываем систему уравнений для нахождения потенциалов:

$$\begin{cases} u_1+v_1=1, \\ u_2+v_1=2, \\ u_2+v_2=3, \\ u_2+v_3=4, \\ u_3+v_3=7, \\ u_3+v_4=12, \\ u_4+v_4=0. \end{cases}$$

Система состоит из семи уравнений и имеет восемь переменных. Система неопределенная. Одному из потенциалов задаем значение произвольно: пусть $u_2=0$. Остальные потенциалы находятся однозначно:

$u_2 = 0,$
$v_1 = 2 - u_2 = 2 - 0 = 2,$
$v_2 = 3 - u_2 = 3 - 0 = 3,$
$v_3 = 4 - u_2 = 4 - 0 = 4,$
$u_1 = 1 - v_1 = 1 - 2 = -1,$
$u_3 = 7 - v_3 = 7 - 4 = 3,$
$v_4 = 12 - u_3 = 12 - 3 = 9,$
$u_4 = 0 - v_4 = 0 - 9 = -9.$

Значения потенциалов записываем рядом с запасами или запросами соответствующих поставщиков и потребителей (таблица 23).

Система уравнений для нахождения потенциалов достаточно проста, обычно ее решают устно, глядя на таблицу задачи, ее занятые клетки и известные потенциалы. Любой неизвестный потенциал, соответствующий занятой клетке, равен находящейся в этой клетке стоимости минус известный потенциал, соответствующий этой же клетке (см. формулы (1.52), (1.53)).

Таблица 23. Опорное решение X_1.

		$v_1 = 2$	$v_2 = 3$	$v_3 = 4$	$v_4 = 9$
	a_i \ b_i	100	100	300	300
$u_1 = -1$	100	1 100 -	2 *0*	3 *0*	1 *7* +
$u_2 = 0$	200	2 0 +	3 100	4 100 -	6 *3*
$u_3 = 3$	300	3 *2*	4 *2*	7 200 +	12 100 -
$u_4 = -9$	200	0 -	0 -	0 -	0 200

4. Проверяем опорное решение X_1 на оптимальность. Для этого вычисляем оценки Δ_{ij} для всех незаполненных клеток таблицы (для всех занятых клеток Δ_{ij} =0):

$\Delta_{12} = u_1 + v_2 - c_{12} = -1 + 3 - 2 = 0,$ $\Delta_{13} = u_1 + v_3 - c_{13} = -1 + 4 - 3 = 0,$
$\Delta_{14} = u_1 + v_4 - c_{14} = -1 + 9 - 1 = 7,$ $\Delta_{24} = u_2 + v_4 - c_{24} = 0 + 9 - 6 = 3,$
$\Delta_{31} = u_3 + v_1 - c_{31} = 3 + 2 - 3 = 2,$ $\Delta_{32} = u_3 + v_2 - c_{32} = 3 + 3 - 4 = 2,$
$\Delta_{41} = u_4 + v_1 - c_{41} = -9 + 2 - 0 = -7,$ $\Delta_{42} = u_4 + v_2 - c_{42} = -9 + 3 - 0 = -6,$

$\Delta_{43} = u_4 + v_3 - c_{43} = -9 + 4 - 0 = -5.$

Положительные оценки записываем (курсивом) в левые нижние углы соответствующих клеток таблицы, вместо отрицательных ставим знак «-». Начальное опорное решение не является оптимальным, так как имеются положительные оценки.

5. Переходим к новому опорному решению. Находим клетку таблицы, которой соответствует наибольшая положительная оценка: $\max\limits_{\Delta_{ij}>0}\{\Delta_{ij}\} = \max\{7,3,2,2\} = 7$ для клетки (1,4). Для этой клетки строим цикл. Ставим в нее знак «+», присоединяем ее к занятым клеткам и применяем метод вычеркивания. После проведения вычеркиваний в таблице остаются только образующие цикл клетки. Цикл изображен в таблице 23. На основании теоремы о существовании и единственности цикла такой цикл единственный. В угловых точках цикла расставляем поочередно знаки «+» и «-», начиная с «+» в клетке (1,4). В клетки, отмеченные знаком «+», прибавляется груз θ, а из клеток, отмеченных знаком «-», вычитается такой же по величине груз. Определяем величину груза θ, перераспределяемого по циклу. Она равна значению наименьшей из перевозок в клетках цикла, отмеченных знаком «-», $\theta = \min\limits_{"-"}\{100, 100, 100\} = 100$. Осуществляем сдвиг по циклу на величину $\theta = 100$. Получаем второе опорное решение X_2 (таблица 24). В данном случае минимум перевозок в клетках, отмеченных знаком «-», достигается в трех клетках, поэтому для того, чтобы число занятых клеток опорного решения было по-прежнему равно $m + n - 1 = 7$, в клетки с номерами (1,1) и (2,3) поставлены нулевые базисные перевозки. Следует освобождать клетку с большей стоимостью перевозки, т.е. клетку (3,4).

Таблица 24. Второе опорное решение X_2

		$v_1 = 2$	$v_2 = 3$	$v_3 = 4$	$v_4 = 2$
	a_i \ b_i	100	100	300	300
$u_1 = 1$	100	1 0	2 *0*	3 *0*	1 100
$u_2 = 0$	200	2 100	3 100 -	4 - +	6 -
$u_3 = 3$	300	3 *2*	4 *2* + 0	7 300 -	12 -
$u_4 = -2$	200	0 *0*	0 *1*	0 *2*	0 200

Вычисляем значение целевой функции на втором опорном решении:
$Z(X_2) = 0 \cdot 1 + 100 \cdot 1 + 100 \cdot 2 + 100 \cdot 3 + 0 \cdot 4 + 300 \cdot 7 + 200 \cdot 0 = 2700.$

6. Проверяем второе оптимальное решение X_2 на оптимальность. Находим потенциалы и оценки (таблица 24). Решение не является оптимальным, так как имеются положительные оценки

$$\Delta_{31} = u_3 + v_1 - c_{31} = 2, \qquad \Delta_{32} = u_3 + v_2 - c_{32} = 20,$$
$$\Delta_{42} = u_4 + v_2 - c_{42} = 1, \qquad \Delta_{43} = u_4 + v_3 - c_{43} = 2.$$

Наибольшая из них равна 2 одновременно для трех клеток (3,1), (3,2) и (4,3). В одну из них, например, в клетку (3,2), ставим знак «+». Для этой клетки строим цикл (таблица 24) и находим величину груза для перераспределения по циклу: $\theta = \min\limits_{"-"}\{300, 100\} = 100$. Осуществим сдвиг по циклу на величину $\theta = 100$. Получаем третье опорное решение X_3 (таблица 25).

Таблица 25. Третье опорное решение X_3.

	a_i \ b_i	$v_1 = 1$ 100	$v_2 = 0$ 100	$v_3 = 3$ 300	$v_4 = 1$ 300
$u_1 = 0$	100	1 0	2 -	3 *0*	1 100
$u_2 = 1$	200	2 100 -	3 -	4 100 +	6 -
$u_3 = 4$	300	3 *2* +	4 100	7 200 -	12 -
$u_4 = -1$	200	0 *0*	0 -	0 *2*	0 200

Вычисляем значение целевой функции на третьем опорном решении:

$$Z(X_3) = 0 \cdot 1 + 100 \cdot 1 + 100 \cdot 2 + 100 \cdot 4 + 100 \cdot 4 + 200 \cdot 7 + 200 \cdot 0 = 2500$$

7. Проверяем третье опорное решение на оптимальность. Находим потенциалы и оценки. Они приведены в таблице 25. Решение не является оптимальным, так как имеются положительные оценки

$$\Delta_{31} = u_3 + v_1 - c_{31} = 2, \qquad \Delta_{43} = u_4 + v_3 - c_{43} = 2.$$

В одну из клеток с положительной оценкой, пусть в клетку (3,1), ставим знак «+». Для этой клетки строим цикл (таблица 25) и находим величину груза для перераспределения по циклу: $\theta = \min\limits_{"-"}\{100, 200\} = 100$. Осуществляем сдвиг по циклу на величину $\theta = 100$. Получаем четвертое опорное решение X_4 (таблица 26).

Таблица 26. Четвертое опорное решение $\tilde{O}_4$

		$v_1 = 3$	$v_2 = 4$	$v_3 = 7$	$v_4 = 3$
	a_i \ b_i	100	100	300	300
$u_1 = -2$	100	1 0 -	2 *0*	3 *2*	1 100 +
$u_2 = -3$	200	2 -	3 -	4 200	6 -
$u_3 = 0$	300	3 100 +	4 100	7 100 -	12 -
$u_4 = -3$	200	0 *0*	0 *1*	0 *4* +	0 200 -

Вычисляем значение целевой функции на четвертом опорном решении:
$Z(X_4) = 0 \cdot 1 + 100 \cdot 1 + 200 \cdot 4 + 100 \cdot 3 + 100 \cdot 4 + 100 \cdot 7 + 200 \cdot 0 = 2300$.

8. Проверяем решение $\tilde{O}_4$ на оптимальность. Находим потенциалы и оценки. Они приведены в таблице 26. Положительными являются оценки $\Delta_{13} = u_1 + v_3 - c_{13} = 2$, $\Delta_{42} = u_4 + v_2 - c_{42} = 1$, $\Delta_{43} = u_4 + v_3 - c_{43} = 4$. Для клетки (4,3), которой соответствует наибольшая оценка, строим цикл (таблица 26) и находим величину груза для перераспределения по циклу: $\theta = \min\limits_{"-"}\{200{,}0{,}200\} = 0$. Осуществляем сдвиг по циклу на величину $\theta = 0$. Получаем пятое опорное решение (таблица 27).

Решение $\tilde{O}_5$ является оптимальным, так как все оценки отрицательные. Значение целевой функции $Z(X_5) = Z(X_4) = 2300$.

Таблица 27. Пятое опорное решение $\tilde{O}_5$

		$v_1 = 3$	$v_2 = 4$	$v_3 = 7$	$v_4 = 7$
	a_i \ b_i	100	100	300	300
$u_1 = -6$	100	1 -	2 -	3 -	1 100
$u_2 = -3$	200	2 -	3 -	4 200	6 -
$u_3 = 0$	300	3 100	4 100	7 100	12 -
$u_4 = -7$	200	0 -	0 -	0 -	0 200

Ответ. $\min Z(X) = 2300$.

Транспортная задача с ограничениями на пропускную способность

Пусть требуется при решении транспортной задачи ограничить перевозки от поставщика с номером l к потребителю с номером k. Возможны ограничения двух типов: 1) $x_{lk} \geq a$; 2) $x_{lk} \leq b$, где a и b - постоянные величины.

1. Если $x_{lk} \geq a$, то необходимо прежде, чем решать задачу, сократить (уменьшить) запасы l - го поставщика и запросы k - го потребителя на величину a (зарезервировать перевозку $x_{lk} = a$). После решения задачи в оптимальном решении следует увеличить объем перевозки x_{lk} на величину a.

2. Если $x_{lk} \leq b$, то необходимо вместо k - го потребителя с запросами b_k ввести двух других потребителей. Один из них с номером k должен иметь запросы $b_k' = b$, а другой с номером $n+1$ - запросы $b_{n+1} = b_k - b$. Стоимости перевозок для этих потребителей остаются прежними, за исключением стоимости $c_{l(n+1)}$, которая принимается равной сколь угодно большому числу $M\,(M >> 1)$. После получения оптимального решения величины грузов, перевозимых к $(n+1)$ – му потребителю, прибавляются к величинам перевозок k - го потребителя. Так как $c_{l(n+1)} = Ì$ - самая большая стоимость перевозки, то в оптимальном решении клетка с номером $(l, n+1)$ останется пустой, $x_{l(n+1)} = 0$ и объем перевозки x_{lk} не превзойдет b.

Пример. Решить транспортную задачу, исходные данные которой приведены в таблице 28, при дополнительных условиях: объем перевозки груза от 1 – го поставщика 2 – му потребителю должен быть не менее 100 единиц $(x_{12} \geq 100)$, а от 3 – го 1 – му не более 200 $(x_{31} \leq 200)$.

Таблица 28.

a_i \ b_i	500	400	300
200	1	5	6
300	2	6	7
500	3	7	8

Решение. Для того чтобы в оптимальном решении объем перевозки x_{12} был не менее 100 единиц, при решении задачи будем предполагать, что запасы первого поставщика a_1 и запросы второго потребителя b_2 меньше фактических на 100 единиц. После получения оптимального решения объем перевозки x_{12} увеличим на 100 единиц.

Для того чтобы удовлетворить требованию $(x_{31} \le 200)$, вместо первого потребителя введем двух других. Один из них под прежним первым номером имеет запросы $b_1 = 200$ единиц и прежние стоимости перевозок единиц груза. Другому присвоим четвертый номер. Его запросы равны $b_4 = 500 - 200 = 300$ единиц и стоимости перевозок единиц груза те же, что и у 1-го потребителя, за исключением c_{34}, которую примем равной сколь угодно большому числу M, т.е. $c_{34} = M$. После нахождения оптимального решения задачи объемы перевозок для 4 – го потребителя необходимо прибавить к соответствующим объемам перевозок для 1 –го потребителя.

В результате указанных преобразований таблица исходных данных задачи будет иметь следующий вид (таблица 29)

Таблица 29.

a_i \ b_i	500	400	300	300
200	1	5	6	1
300	2	6	7	2
500	3	7	8	M

Далее задачу решаем обычным методом потенциалов. Проверяем выполнение необходимого и достаточного условия существования решения задачи. Находим суммарные запасы поставщиков и запросы потребителей:

$a_1 + a_2 + a_3 = 100 + 300 + 500 = 900$;

$b_1 + b_2 + b_3 + b_4 = 200 + 300 + 300 + 300 = 1100/$

Задача с неправильным балансом. Вводим фиктивного поставщика с запасами $a_4 = 1100 - 900 = 200$ (таблица 30).

Составляем начальное опорное решение X_1 методом минимальной стоимости. Записываем матрицу стоимостей $\tilde{N}$:

$$C = \begin{array}{cc} \begin{pmatrix} \underline{1} & 5 & 6 & 1 \\ \underline{2} & 6 & 7 & \underline{2} \\ 3 & \underline{7} & \underline{8} & \grave{I} \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 1 \\ 3 \\ 5 \\ 7 \end{array} \\ \begin{array}{cccc} 2 & 4 & 6 & \end{array} & \end{array}$$

В матрице $\tilde{N}$ подчеркнутые числа – минимальные элементы, а цифры рядом со строками и столбцами – порядок исключения из рассмотрения поставщиков и потребителей.

Таблица 30. Первое опорное решение X_1

		$v_1 = 1$	$v_2 = 0$	$v_3 = 1$	$v_4 = 1$
	a_i \ b_i	200	300	300	300
$u_1 = 0$	100	1 100	5 -	6 -	1 *0*
$u_2 = 1$	300	2 100 -	6 -	7 -	2 200 +
$u_3 = 7$	500	3 *5* +	7 300	8 200 -	M -
$u_4 = -1$	200	0 *0*	0 -	0 100 +	0 100 -

Полученное решение X_1 имеет $m+n-1=4+4-1=7$ базисных переменных. Вычислим значение целевой функции на этом опорном решении:

$$Z(X_1)=100\cdot 1+100\cdot 2+200\cdot 2+300\cdot 7+200\cdot 8+100\cdot 0+100\cdot 0=4400$$

Для проверки оптимальности опорного решения находим потенциалы. Записываем систему уравнений для нахождения потенциалов и решаем ее:

$$\begin{cases} u_1+v_1=1, \\ u_2+v_1=2, \\ u_2+v_4=2, \\ u_3+v_2=7, \\ u_3+v_3=8, \\ u_4+v_3=0, \\ u_4+v_4=0. \end{cases}$$

Система состоит из семи уравнений и имеет восемь переменных. Так как число неизвестных на единицу больше числа уравнений, то одному из потенциалов задаем значение произвольно: пусть $u_1=0$. Остальные потенциалы находятся однозначно:

$u_1=0,$

$v_1=1-u_1=1-0=1,$

$u_2=2-v_1=2-1=1,$

$v_4=2-u_2=2-1=1,$

$u_4=0-v_4=0-1=-1,$

$v_3=0-u_4=0-(-1)=1,$

$u_3=8-v_3=8-1=7,$

$v_2=7-u_3=7-7=0.$

Значения потенциалов приведены в таблице 30. Находим оценки для свободных клеток таблицы:

$\Delta_{12} = u_1 + v_2 - c_{12} = 0 + 0 - 5 = -5,$ $\Delta_{13} = u_1 + v_3 - c_{13} = 0 + 1 - 6 = -5,$

$\Delta_{14} = u_1 + v_4 - c_{14} = 0 + 1 - 1 = 0,$ $\Delta_{22} = u_2 + v_2 - c_{22} = 1 + 0 - 6 = -5,$

$\Delta_{23} = u_2 + v_3 - c_{23} = 1 + 1 - 7 = -5,$ $\Delta_{31} = u_3 + v_1 - c_{31} = 7 + 1 - 3 = 5,$

$\Delta_{34} = u_3 + v_4 - c_{34} = 7 + 1 - \grave{I} < 0,$ $\Delta_{41} = u_4 + v_1 - c_{41} = -1 + 1 - 0 = 0,$

$\Delta_{42} = u_4 + v_2 - c_{42} = -1 + 0 - 0 = -1.$

Опорное решение не является оптимальным, так как имеется положительная оценка для клетки $(3,1)$. Отмечаем эту клетку знаком «+». Находим цикл для улучшения опорного решения (таблица 30). Определяем величину груза для перераспределения по циклу $\theta = \min\limits_{''-''}\{100,200,100\} = 100$. Осуществляем сдвиг по циклу на величину $\theta = 100$. Получаем второе опорное решение X_2 (таблица 31).

Таблица 31. Второе опорное решение X_2.

		$v_1 = 1$	$v_2 = 5$	$v_3 = 6$	$v_4 = 1$
	a_i \ b_i	200	300	300	300
$u_1 = 0$	100	1 100	5 *0*	6 *0*	1 *0*
$u_2 = 1$	300	2 0	6 *0*	7 *0*	2 300
$u_3 = 2$	500	3 100	7 300	8 100	M -
$u_4 = -6$	200	0 -	0 -	0 200	0 -

В таблице 31 также записаны потенциалы и оценки для свободных клеток. Решение X_2 оптимальное, так как все оценки неположительные. Запишем оптимальное решение исходной задачи. Для этого увеличим объем перевозки x_{12} на 100 единиц и объединим объемы перевозок 4 – го потребителя с объемами перевозок 1 – го потребителя.

Получим

$$X^* = \begin{pmatrix} 100 & 100 & 0 \\ 300 & 0 & 0 \\ 100 & 300 & 100 \end{pmatrix}.$$

Вычислим значение целевой функции на этом решении:

$Z(X^*) = 100 \cdot 1 + 100 \cdot 5 + 300 \cdot 2 + 100 \cdot 3 + 300 \cdot 7 + 100 \cdot 8 + 100 = 4400$

Ответ: $\min Z(X) = 4400$.

В некоторых задачах требуется запретить перевозки от отдельных поставщиков отдельным потребителям. В таких случаях либо зачеркивают клетку таблицы транспортной задачи, либо назначают соответствующую этой клетке стоимость перевозки единицы груза сколь угодно большой, равной $M >> 1$. В остальном задача решается обычным способом. Для разрешимости данной задачи необходимо существование начального опорного решения.

Транспортная задача по критерию времени

Задача по критерию времени возникает при перевозке срочных грузов. Как и в обычной транспортной задаче, имеется m поставщиков с запасами однородного груза в количестве $a_1, a_2, \ldots, a_m$ и n потребителей, которым этот груз должен быть доставлен в объеме $b_1, b_2, \ldots, b_n$. Известно время t_{ij}, за которое груз доставляется от каждого i - го поставщика каждому j - му потребителю, $i = 1..m$, $j = 1..n$. Требуется составить такой план перевозок груза, при котором запасы всех поставщиков вывозятся полностью, запросы всех потребителей удовлетворяются полностью и наибольшее время доставки всех грузов является минимальным.

Составим математическую модель этой задачи. Обозначим x_{ij} - объем перевозимого груза от i - го поставщика j - му потребителю. Система ограничений задачи не отличается от системы ограничений обычной транспортной задачи. Пусть $X = (x_{ij})$, $i = 1..m$, $j = 1..n$ - некоторое опорное решение задачи. Запишем целевую функцию задачи. Обозначим через $T(X)$ наибольшее значение элементов матрицы $T = (t_{ij})$, $i = 1..m$, $j = 1..n$, соответствующих клеткам таблицы, занятым опорным решением: $T(X) = \max\limits_{x_{ij} > 0} \{t_{ij}\}$. Таким образом, за время $T(X)$ план перевозок будет выполнен полностью. Математическая модель имеет вид

$$T(X) = \max_{x_{ij} > 0} \{t_{ij}\} \to min, \tag{1.54}$$

$$\sum_{j=1}^{n} x_{ij} = a_i, i = 1..m, \tag{1.55}$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \ j = 1..n, \tag{1.56}$$

$$x_{ij} \geq 0, \ i = 1..m, \ j = 1..n. \tag{1.57}$$

Задача решается в следующем порядке. Находится начальное опорное решение X_1. Определяется значение целевой функции $T(X_1) = \max\limits_{x_{ij} > 0} \{t_{ij}\} = t_{l_1 k_1}$. Все свободные клетки, которым соответствуют значения $t_{ij} > T(X_1)$, исключаются из рассмотрения (перечеркиваются). Занимать эти клетки нецелесообразно, так как повысится значение целевой функции. Чтобы понизить ее значение, необходимо освободить клетку (l_1, k_1), в которой t_{ij} достигает максимума. Для этого строят «*разгрузочные циклы»,* которые могут включать в свой состав несколько свободных клеток. В каждом разгрузочном цикле, начиная с разгружаемой клетки (l_1, k_1), расставляются поочередно знаки «-» и «+» и осуществляется сдвиг на величину $\theta = \min_{"-"}\{x_{ij}\}$. Если удается эту клетку разгрузить, то она исключается из рассмотрения (зачеркивается). Получается новое опорное решение X_2, на котором значение целевой функции меньше, чем на X_1. Далее снова пытаемся разгрузить клетку, соответствующую $T(X_2) = \max_{x_{ij}>0}\{t_{ij}\} = t_{l_2 k_2}$. Процесс продолжается до тех пор, пока возможность разгрузить соответствующую клетку не исчезнет.

Пример. Найти минимальное время на осуществление всех перевозок для задачи, исходные данные которой приведены в таблице 32.

Таблица 32.

a_i \ b_i	20	30	40	60
20	10	6	3	2
30	5	8	7	4
50	2	4	5	12
50	15	5	9	4

Решение. Составим начальное опорное решение X_1 по методу северо–западного угла (таблица 33). Базисные нули не записываем. Максимум целевой функции $T(X_1) = \max_{x_{ij}>0}\{10,8,5,12,4\} = 12$ достигается в клетке $(3,4)$. Перечеркиваем клетку $(4,1)$, в которой время доставки груза $t_{41} = 15$ больше $T(X_1) = 12$.

Таблица 33.

a_i \ b_i	20	30	40	60
20	10 20	6	3	2
30	5	8 30 -	7	4 +
50	2	4 +	5	12 10 -
50	~~15~~	5	9	4 50

Для улучшения решения разгрузим клетку $(3{,}4)$ с помощью цикла $(3{,}4)$, $(2{,}4)$, $(2{,}2)$, $(3{,}2)$ (таблица 33). Обозначим цикл, найдем $\theta = \min\limits_{''-''}\{10{,}30\} = 10$. Осуществив сдвиг по циклу, получим второе опорное решение X_2 (таблица 34). Максимум целевой функции на этом опорном решении $T(X_2) = \max\limits_{x_{ij}>0}\{10{,}8{,}4{,}5{,}4\} = 10$ достигается в клетке $(1{,}1)$. Перечеркнем клетку $(3{,}4)$, так как время $t_{34} = 12$ больше, чем $T(X_2) = 10$. Разгрузим клетку $(1{,}1)$ с помощью цикла $(1{,}1)$, $(1{,}2)$, $(2{,}2)$, $(2{,}1)$. Обозначим цикл, найдем $\theta = \min\limits_{''-''}\{20{,}20\} = 20$. Осуществив сдвиг по циклу, получим третье опорное решение X_3 (таблица 35).

Таблица 34.

a_i \ b_i	20	30	40	60
20	10 20 -	6 +	3	2
30	5 +	8 20 -	7	4 10
50	2	4 10	5 40	~~12~~
50	~~15~~	5	9	4 50

Максимум целевой функции на третьем опорном решении $T(X_3) = \max\limits_{x_{ij}>0}\{6{,}5{,}4{,}4{,}5{,}4\} = 6$ и достигается в клетке $(1{,}2)$. Перечеркнем клетки $(1{,}1)$, $(2{,}2)$, $(2{,}3)$ и $(4{,}3)$: в них время $t_{11} = 10$, $t_{22} = 8$, $t_{23} = 7$ и $t_{43} = 9$ больше, чем $T(X_3) = 6$. Разгрузим клетку $(1{,}2)$ с помощью цикла $(1{,}2)$, $(1{,}3)$, $(3{,}3)$, $(3{,}2)$. Обозначим цикл, найдем $\theta = \min\limits_{''-''}\{20{,}20\} = 20$. Осуществив сдвиг по циклу, получим четвертое опорное решение X_4 (таблица 36).

Таблица 35.

a_i \ b_i	20	30	40	60
20	~~10~~	6 20 -	3 +	2
30	5 20	~~8~~	~~7~~	4 10
50	2	4 10 +	5 40 -	~~12~~
50	~~15~~	5	~~9~~	4 50

Таблица 36.

a_i \ b_i	20	30	40	60
20	10	6	3 20	2
30	$\underline{5}$ 20	8	7	4 10
50	2	4 30	$\underline{5}$ 20	12
50	15	5	9	4 50

Максимум целевой функции на этом опорном решении $T(X_4) = \max\limits_{x_{ij}>0}\{5,4,4,5,4\} = 5$ и достигается в клетках $(2,1)$ и $(3,3)$. Перечеркнем клетки $(1,2)$ и $(4,2)$, в которых время перевозок не менее $t_{21} = 5$. С помощью оставшихся невычеркнутых клеток разгрузить клетки $(2,1)$ и $(3,3)$ не удается, поэтому X_4 является оптимальным решением.

Ответ: $\min T(X) = 5$.

ТЕОРИЯ ИГР

Общая задача теории игр

Теория игр − математическая теория конфликтных ситуаций. Конкуренция в бизнесе, спортивные встречи, боевые операции − примеры конфликтных ситуаций.

Если в игре сталкиваются интересы двух противников, то говорят о парной игре или игре двух лиц. Если игроков больше двух, то говорят, что игра множественная.

Стратегией игрока называется система правил, однозначно определяющих выбор поведения игрока на каждом шаге в зависимости от ситуации, сложившейся в процессе игры. В зависимости от числа возможных стратегий игры делятся на *конечные* и *бесконечные*.

Обычно решение (выбор) при каждом личном ходе принимается игроком в ходе самой игры в зависимости от сложившейся конкретной ситуации. Однако теоретически ни чего не изменится, если мы представим себе, что все эти решения принимаются игроком заранее. Для этого игрок должен был бы заблаговременно составить перечень всех возможных в ходе игры ситуаций и предусмотреть свое решение для каждой из них. В принципе (если не практически) это возможно для любой игры. Если такая система решений будет принята, это будет означать, что игрок выбрал определенную стратегию.

Обозначим S_i− множество стратегий i-го игрока, элементы которого обозначаются s_i, $i=\overline{1,n}$, где n− общее число игроков. Процесс игры состоит в выборе каждым игроком i одной своей стратегии $s_i \in S_i$. В результате сложившийся ситуации $s=(s_1,s_2,\ldots,s_n)$ игрок i получает выигрыш $H_i(s)$.

Игры, в которых целью каждого участника является получение по возможности большего индивидуального выигрыша, называются *бескоалиционными* или *антагонистическими*. Если в игре действия игроков направлены на максимизацию выигрышей коллективов (коалиции), то такую игру называют *коалиционной*.

По определению бескоалиционной игрой называют упорядоченную тройку

$$\Gamma=\left(I,\{S_i\}_{i\in I},\{H_i\}_{i\in I}\right),$$

в которой $I=\{1,2,\ldots,n\}$− множество индексов, S_i − непустые множества, H_i− функции на множестве $S=\prod_{i=1}^{n} S_i$, принимающие вещественные значения, т.е.

$$H_i : S \to R.$$

Антагонистическая игра называется *игрой с постоянной суммой*, если существует такое постоянное C, что

$$\sum_{i=1}^{n} H_i(s) = C \quad \forall s \in S.$$

Ситуация s в игре называется *приемлемой* для игрока i, если этот игрок, изменяя в ситуации s свою стратегию s_i на какую-либо другую s'_i, не может увеличить свой выигрыш.

Ситуация s, приемлемая для всех игроков, называется *ситуацией равновесия*.

Процесс нахождения ситуации равновесия в бескоалиционной игре является процессом решения игры.

Матричные игры

Определение матричной игры

Парная игра называется *игрой с нулевой суммой*, если один игрок выигрывает столько, сколько проигрывает другой в той же игре т.е. сумма выигрышей обеих сторон равна нулю.

Каждая фиксированная стратегия, которую может выбрать игрок, называется его *чистой стратегией*.

Матричной называют конечную парную игру с нулевой суммой.

Пусть первый игрок имеет m чистых стратегий, а второй − n.

Матричная игра задается формально системой чисел − матрицей $A = \|a_{ij}\|$, $i = \overline{1,m}$, $j = \overline{1,n}$. i-й строке матрицы соответствует i-я стратегия первого игрока, а j-му столбцу матрицы соответствует j-я стратегия второго игрока. Если первый игрок выбирает i-ю стратегию, а второй игрок выбирает j-ю стратегию, то элемент a_{ij} матрицы A определяют выигрыш первого игрока (и соответственно проигрыш второго) в ситуации при выбранных стратегиях.

Матрица $A = \|a_{ij}\|$ называется *платежной матрицей* или *матрицей игры*. Элементы матрицы могут быть как положительными, так и отрицательными (отрицательный выигрыш означает проигрыш).

Задача первого игрока − максимизировать свой выигрыш. Задача второго игрока так же максимизировать свой выигрыш. Задача максимизации выигрыша второго игрока сводится к минимизации его проигрыша.

В дальнейшем для удобства первого игрока иногда будем именовать «мы», а второго игрока – «противник».

Чистые стратегии

Выбирая i-ю стратегию мы всегда должны рассчитывать на то, что противник ответит на нее такой j-й стратегией, для которой наш выигрыш a_{ij} минимален. Отсюда, гарантированный выигрыш первого игрока, применяющего чистую i-ю стратегию:

$$\alpha_i = \min_{1 \le j \le n} a_{ij}$$

есть минимальное значение из чисел a_{ij} в i-й строке.

Выбирая какую-нибудь -ю стратегию, мы должны рассчитывать на то, что в результате разумных действий противника мы не выиграем больше чем α_i. Естественно нам следует выбрать ту стратегию, при которой число α_i является максимальным. Обозначим это максимальное значение $\underline{v}$.

Число

$$\underline{v} = \max_{1 \le i \le m} \alpha_i = \max_{1 \le i \le m} \left\{ \min_{1 \le j \le n} a_{ij} \right\}$$

называется *нижним значением игры* или *нижней ценой игры*, а соответствующая чистая стратегия i_0, при которой достигается $\underline{v}$, называется максиминной стратегией первого игрока.

Очевидно, что если мы будем придерживаться максиминной стратегии, то нам при любом поведении противника гарантирован выигрыш, во всяком случае не меньший $\underline{v}$.

Аналогично,

$$\bar{v} = \min_{1 \le j \le n} \left\{ \max_{1 \le i \le m} a_{ij} \right\}$$

называется *верхним значением игры*, а чистая стратегия j_0 второго игрока, при которой достигается $\bar{v}$, называется *минимаксной стратегией* второго игрока.

Придерживаясь своей наиболее осторожной минимаксной стратегии, противник гарантирует себе следующее: что бы мы не предпринимали против него, он проиграет сумму, не большую чем $\bar{v}$.

Всегда справедливо неравенство $\underline{v} \le \bar{v}$.

Если $\underline{v} = \bar{v} = v$, то существует элемент матрицы, являющийся минимальным в своей строке и максимальным в своем столбце. Элемент матрицы, обладающий этим свойством, называется *седловой точкой.*

Число v называется *значением* или *ценой* игры.

Если $a_{i_0 j_0}$ – седловая точка, то соответствующие стратегии с номерами i_0 и j_0 называются *оптимальными*, а их совокупность – *решением игры*.

Решение игры обладает следующим замечательным свойством. Если один из игроков придерживается своей оптимальной стратегии, а другой игрок выберет другую (не оптимальную свою стратегию), то для игрока, допустившего отклонение от оптимальной стратеги, это никогда не будет выгодно. Такое отклонение может в лучшем случае оставить выигрыш неизменным, а в худшем – уменьшить значение функции выигрыша для игрока, выбравшего не оптимальную стратегию.

Пример. Найти оптимальные стратегии игроков и цену игры, заданной матрицей:

$$A = \begin{pmatrix} 2 & 10 & 3 & 14 & 5 \\ 8 & 9 & 5 & 6 & 7 \\ 10 & 8 & 4 & 8 & 12 \end{pmatrix}.$$

Минимизируя элементы первой строки, получим, что
$\alpha_1 = \min\{2,10,3,14,5\} = 2$; аналогично, $\alpha_2 = \min\{8,9,5,6,7\} = 5$;
$\alpha_3 = \{10,8,4,8,12\} = 4$. Следовательно
$\underline{v} = \max\limits_{i=\overline{1,m}} \alpha_i = \max\{2,5,4\} = 5$.

Максимизируя элементы по столбцам, получим: $\beta_1 = \max\{2,8,10\} = 10$; $\beta_2 = \max\{10,9,8\} = 10$; $\beta_3 = \max\{3,5,4\} = 5$; $\beta_4 = \max\{14,6,8\} = 14$; $\beta_5 = \max\{5,7,12\} = 12$. Отсюда
$\overline{v} = \min\limits_{j=\overline{1,n}} \beta_j = \min\{10,10,5,14,12\} = 5$.

Мы получили, что $\underline{v} = \overline{v} = 5$, следовательно цена игры $v = 5$.

Данная игра имеет седловую точку $a_{23} = 5$, при этом вторая стратегия первого игрока и третья стратегия второго игрока являются оптимальными стратегиями. Задача разрешима в чистых стратегиях. Оптимальный способ игры найден. Придерживаясь второй чистой стратегии, первый игрок обеспечит себе выигрыш, не меньший 5, второй игрок, применяя третью чистую стратегию, проигрывает не более 5.

Смешанные стратегии

Среди конечных игр сравнительно редко встречаются игры с седловой точкой. Как правило $\underline{v} < \overline{v}$.

Из предыдущего мы пришли к выводу, что если каждому игроку предоставлен выбор одной стратегии, то придерживаясь своей максиминной стратегии, первый игрок при любом поведении противника заведомо гарантирует себе выигрыш, не меньший нижней цены игры. Возникает вопрос: нельзя ли гарантировать первому игроку средний выигрыш, больший $\underline{v}$, если применять не одну-единственную чистую стратегию, а чередовать случайным образом несколько стратегий?

Такие комбинированные стратегии, состоящие в применении нескольких чистых стратегий, чередующихся по случайному закону с определенным распределением частот, называются *смешанными стратегиями*.

Очевидно, что каждая чистая стратегия является частным случаем смешанной, в которой все стратегии, кроме одной, применяются с нулевыми частотами, а данная – с частотой 1.

Оказывается, что, применяя не только чистые, но и смешанные стратегии, модно для каждой конечной игры получить решение, т.е. пару таких стратегий, что при применении их обоими игроками выигрыш будет равен цене игры, а при любом одностороннем отклонении от оптимальной стратегии выигрыш может измениться только в худшую сторону для отклоняющегося.

Высказанное утверждение составляет содержание так называемой основной теоремы теории игр. Эта теорема была доказана фон Нейманом в 1928 г.

Терема. *Каждая конечная игра имеет, по крайней мере, одно решение в области смешанных стратегий.*

Обозначим через s_x смешанную стратегию первого игрока в которой он выбирает первую, вторую, … , m-ю чистую стратегию с вероятностями (частотами) $x_1, x_2, \dots, x_m$ соответственно, так что

$$x_i \geq 0, \sum_{i=1}^{m} x_i = 1,\ i = \overline{1,m}.$$

Через s_y – обозначим смешанную стратегию второго игрока, применяющего свои первую, вторую,…, n-ю чистые стратегии с соответствующими вероятностями $y_1, y_2, \dots, y_n$, причем

$$y_j \geq 0, \sum_{j=1}^{n} y_j = 1, j = \overline{1,n}.$$

Очевидно, что смешанные стратегии s_x и s_y однозначно определяются наборами чисел $x = (x_1, x_2, \dots, x_m)$ и $y = (y_1, y_2, \dots, y_n)$ соответственно, при заданных чистых стратегиях.

Каждый игрок имеет бесчисленное множество смешанных стратегий. Множество смешанных стратегий первого игрока обозначим через S_1, а множество смешанных стратегий второго игрока – через S_2.

Задача первого игрока состоит в выборе такой стратегии $s_{x^*} \in S_1$, чтобы при отсутствии информации о выборе другого максимизировать свой выигрыш. Задача второго игрока состоит в выборе такой стратегии $s_{y^*} \in S_2$, чтобы при отсутствии информации о поведении первого игрока минимизировать выигрыш первого.

Если первый игрок применяет стратегию $s_x \in S_1$, второй – стратегию $s_y \in S_2$, то средний выигрыш $M(s_x, s_y)$ первого игрока равен

$$M(s_x, s_y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j.$$

Выигрыш $M(s_x, s_y)$ называют *функцией игры*.

Например, в игре с матрицей

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

Первый игрок имеет две чистые стратегии с вероятностями $x^{(1)} = (1{,}0)$ и $x^{(2)} = (0{,}1)$ и бесчисленное множество смешанных стратегий с частотами, такими, как $x^{(3)} = (1/2, 1/2)$, $x^{(4)} = (1/17, 16/17)$, $x^{(5)} = (12/25, 13/25)$ и т.п.; все они являются элементами множества

$S_1 = \{s_x : x = (x_1, x_2),\ x_1 \geq 0, x_2 \geq 0,\ x_1 + x_2 = 1\}$.

Второй игрок имеет четыре чистые стратеги с соответствующими вероятностями $y^{(1)} = (1{,}0{,}0{,}0)$, $y^{(2)} = (0{,}1{,}0{,}0)$, $y^{(3)} = (0{,}0{,}1{,}0)$, $y^{(4)} = (0{,}0{,}0{,}1)$ и бесконечное множество смешанных стратегий с вероятностями, такими, как $y^{(5)} = (1/4, 0, 1/4, 0\ 1/2)$, $y^{(6)} = (1/3, 1/6, 1/4, 1/4)$, являющихся элементами множества

$$S_2 = \left\{ s_y : y = (y_1, y_2, y_3, y_4),\ \ y_j \geq 0,\ \ \sum_{j=1}^{4} y_j = 1 \right\}.$$

Если первый игрок применяет стратегию $s_{x^{(0)}}$, где $x^{(0)} = (1/6, 5/6)$, а второй применяет стратегию $y^{(0)} = (1/3, 0, 1/3, 1/3)$, то средний выигрыш первого игрока, определяемый функцией игры, окажется равным

$$M\left(s_{x^{(0)}}, s_{y^{(0)}}\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i^{(0)} y_j^{(0)} =$$

$$= 2 \cdot \frac{1}{6} \cdot \frac{1}{3} + 3 \cdot \frac{1}{6} \cdot 0 - 1 \cdot \frac{1}{6} \cdot \frac{1}{3} + 4 \cdot \frac{1}{6} \cdot \frac{1}{3} +$$

$$+5 \cdot \frac{5}{6} \cdot \frac{1}{3} - 4 \cdot \frac{5}{6} \cdot 0 + 3 \cdot \frac{5}{6} \cdot \frac{1}{3} - 5 \cdot \frac{5}{6} \cdot \frac{1}{3} = \frac{10}{9}.$$

Если же первый игрок применяет стратегию

$x^{(0)} = (1/6, 5/6)$, а второй стратегию $y^{(7)} = (1/3, 0, 1/2, 1/6)$, то $M\left(s_{x^{(0)}}, s_{y^{(7)}}\right) = 2 \cdot \frac{1}{6} \cdot \frac{1}{3} + 3 \cdot \frac{1}{6} \cdot 0 - 1 \cdot \frac{1}{6} \cdot \frac{1}{2} + 4 \cdot \frac{1}{6} \cdot \frac{1}{6} +$

$$+5 \cdot \frac{5}{6} \cdot \frac{1}{3} - 4 \cdot \frac{5}{6} \cdot 0 + 3 \cdot \frac{5}{6} \cdot \frac{1}{2} - 5 \cdot \frac{5}{6} \cdot \frac{1}{6} = \frac{25}{12}.$$

Седловая точка в смешанных стратегиях

Гарантированный выигрыш первого игрока, применяющего смешанную стратегию s_x,

$$v(s_x) = \min_{s_y \in S_2} M\left(s_x, s_y\right),$$

$$S_2 = \left\{ s_y : y = (y_1, y_2, \dots, y_n),\ \ y_j \geq 0, j = \overline{1, n},\ \ \sum_{j=1}^{n} y_j = 1 \right\}.$$

Стратегия s_x^*, при которой гарантированный выигрыш первого игрока достигает максимального значения, называется *оптимальной стратегией* первого игрока. Следовательно

$$v(s_x^*) = \max_{s_x \in S_1} v(s_x) = \max_{s_x \in S_1} \min_{s_y \in S_2} M\left(s_x, s_y\right),$$

$$S_1 = \left\{ s_x : x = (x_1, x_2, \dots, x_m),\ \ x_i \geq 0, i = \overline{1, m},\ \ \sum_{i=1}^{m} x_i = 1 \right\}.$$

Гарантированный проигрыш второго игрока (т.е. он проиграет не больше своего гарантированного проигрыша):

$$u(y) = \max_{s_x \in S_1} M\left(s_x, s_y\right).$$

Аналогично, s_y^* – оптимальная стратегия второго игрока, если
$u(s_y^*) = \min\limits_{s_y \in S_2} u(s_y) = \min\limits_{s_y \in S_2} \max\limits_{s_x \in S_1} M(s_x, s_y).$

Совокупность оптимальных стратегий s_x^* и s_y^* называется *решением игры*.

Если s_x^* и s_y^* – оптимальные стратегии первого и второго игроков, то

$$M(s_x, s_y^*) \le M(s_x^*, s_y^*) \le M(s_x^*, s_y). \qquad (2.1)$$

Пара смешанных стратегий (s_x^*, s_y^*), для которых выполняется неравенство (2.1), называется *седловой точкой* функции $M(s_x, s_y)$.

Каждая матричная игра с нулевой суммой имеет решение в смешанных стратегиях, т.е. существуют такие смешанные стратегии s_x^* и s_y^* первого и второго игроков соответственно, что выполняется условие (2.1).

Гарантированный выигрыш первого игрока, применяющего свою оптимальную стратегию, равен гарантированному проигрышу второго игрока, применяющего свою оптимальную стратегию:

$v(s_x^*) = u(s_y^*) = v^*,$

где v^* – цена игры.

В заключении параграфа отметим важное свойство оптимальных стратегий, которое широко применяется при решении игр.

Предположим, что нами найдено решение игры, состоящее из двух оптимальных смешанных стратегий s_x^* и s_y^*. В общем случае не все чистые стратегии входят в его оптимальную смешанную стратегию, а только некоторые (но возможно и все). Будем называть чистые стратегии, входящие в оптимальную смешанную стратегию игрока, его *полезными стратегиями*.

Оказывается, что решение игры обладает следующим свойством: *если один из игроков придерживается своей оптимальной смешанной стратегии s_x^* (s_y^*), то выигрыш останется неизменным и равным цене игры v^*, независимо от того, что делает другой игрок, если он только не выходит за пределы своих полезных стратегий.* Он, например, может пользоваться одной из своих полезных стратегий в чистом виде (использовать чистую стратегию), а также может смешивать их в любой пропорции.

Так же в теории игр доказано, что у любой конечной игры размера $m \times n$ имеется решение, в котором число полезных стратегий той и другой стороны не превосходит наименьшего из двух чисел m и n.

Графический метод решения задач теории игр

Если матрица игры A имеет размер $2 \times n$ или $m \times 2$, то решение задачи может быть получено графически.

Иногда процесс решения задачи можно упростить, если вычеркнуть из матрицы доминируемые (неполезные) стратегии первого игрока и доминирующие стратегии второго игрока.

Если матрица содержит две одинаковые строки (столбца), то одну из строк (столбцов) можно вычеркнуть. Тем самым мы убираем одну повторяющуюся стратегию.

Далее, если каждый элемент строки i_0 меньше или равен соответствующего элемента строки i_1, то говорят, что стратегия i_0 является *доминируемой*. Очевидно, что доминируемой стратегией i_0 первого игрока мы никогда не должны пользоваться, поскольку она является заведомо невыгодной, поэтому строку матрицы с номером i_0 можно вычеркнуть.

Так же стратегия i_0 является доминируемой, если существуют такие числа

$$\mu_i \geq 0, i \neq i_0, \sum_{\substack{i=1 \\ i\neq i_0}}^{m} \mu_i \geq 1,$$

что

$$\sum_{\substack{i=1 \\ i\neq i_0}}^{m} \mu_i a_{ij} \geq a_{i_0 j}\ , j = \overline{1,n}.$$

Т.е. если элементы строки i_0 не превосходят соответствующих элементов линейной комбинации остальных строк.

Если существует столбец матрицы j_0 все элементы которого больше или равны соответствующих элементов столбца j_1, то говорят что j_0 является *доминирующей* стратегией второго игрока. Доминирующая стратегия второго игрока заведомо невыгодна, поэтому можно вычеркнуть столбец матрицы с номером j_0.

Пример. Найти оптимальные стратегии игроков и цену игры, заданной матрицей

$$A = \begin{pmatrix} 3 & 7 & 2 & 1 & 5 & 8 \\ 4 & 9 & 3 & 6 & 2 & 1 \\ 2 & 3 & 1 & 4 & 9 & 20 \end{pmatrix}.$$

В чистых стратегиях решения игры нет, так как

$$\underline{v} = \max_i \min_j a_{ij} = \max\{1,1,1\} = 1,$$

$$\overline{v} = \min_j \max_i a_{ij} = \min\{4,9,3,6,9,20\} = 3,$$

$$\underline{v} < \overline{v}.$$

Упростим матрицу игры. Все элементы второго столбца больше соответствующих элементов первого столбца, следовательно, вторая стратегия второго игрока является доминирующей. Вычеркивая второй столбец, получим

$$A^1 = \begin{pmatrix} 3 & 2 & 1 & 5 & 8 \\ 4 & 3 & 6 & 2 & 1 \\ 2 & 1 & 4 & 9 & 20 \end{pmatrix}.$$

В полученной матрице все элементы первого столбца больше соответствующих элементов второго столбца. Вычеркиваем первый столбец.

$$A^2 = \begin{pmatrix} 2 & 1 & 5 & 8 \\ 3 & 6 & 2 & 1 \\ 1 & 4 & 9 & 20 \end{pmatrix}.$$

В матрице A^2 выполняются неравенства:

$$a_{1j} \leq \frac{1}{2}a_{2j} + \frac{1}{2}a_{3j}, \qquad j = \overline{1,4},$$

т.е. элементы первой строки не превосходят элементов линейной комбинации остальных строк, следовательно, стратегия, соответствующая первой строке матрицы A^2 является доминируемой. Вычеркивая первую строку, получим:

$$A^3 = \begin{pmatrix} 3 & 6 & 2 & 1 \\ 1 & 4 & 9 & 20 \end{pmatrix}.$$

В матрице A^3 элементы второго столбца превосходят соответствующие элементы первого столбца. Вычеркиваем второй столбец и получаем матрицу

$$A^4 = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 9 & 20 \end{pmatrix}.$$

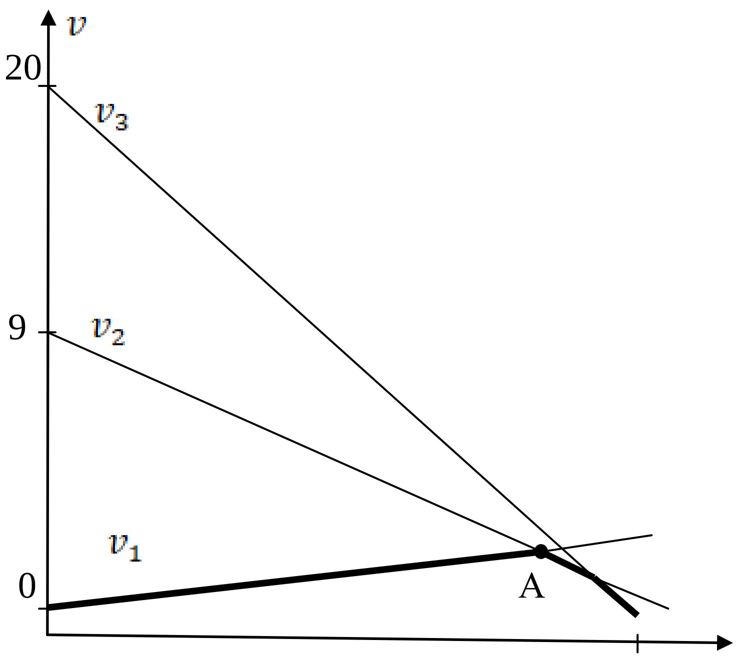
В результате получили задачу с матрицей A^4 размера 2×3. Решим эту игру геометрически.

Пусть первый игрок применяет свою первую стратегию с вероятностью $x_1 = x$. Поскольку $x_1 + x_2 = 1$, то вторая стратегия будет использоваться с вероятностью $x_2 = 1 - x$. Обозначим $v_j(x)$ – ожидаемый средний выигрыш первого игрока, применяющего первую стратегию с вероятностью x при условии, что второй игрок отвечает своей чистой стратегией j ($j = 1,2,3$).

В плоскости переменных (x, v) построим графики функций $v_j(x)$:

$v_1(x) = 3x + 1(1 - x) = 2x + 1,$
$v_2(x) = 2x + 9(1 - x) = -7x + 9,$
$v_3(x) = x + 20(1 - x) = -19x + 20,$

Поскольку вероятность x принимает значения $0 \leq x \leq 1$, то нас будут интересовать графики функций на отрезке $[0,1]$ (см. Рис. 2.1).



0 1 x

Рис. 2.1 Гарантированный выигрыш первого игрока

Гарантированный средний выигрыш первого игрока (изображенный жирной линией) является наименьшим из значений функций $v_j(x)$, $j = \overline{1,n}$. Очевидно, что своего наибольшего значения он достигает в точке А, являющейся пересечением прямых $v_1(x)$ и $v_2(x)$.

Найдем абсциссу точки А, приравняв правые части уравнений прямых:

$$2x + 1 = -7x + 9,$$

$$9x = 8,$$

$$x = \frac{8}{9}.$$

Следовательно, $x_1 = \frac{8}{9}$, $x_2 = 1 - x_1 = 1 - \frac{8}{9} = \frac{1}{9}$. Отсюда оптимальная стратегия первого игрока упрощенной задачи $s_x^* = \left(\frac{8}{9}, \frac{1}{9}\right)$.

Цена игры численно равна ординате точки А. Подставив $x = 8/9$ в правую часть уравнения прямой $v_1(x)$, получим, что цена игры $v = v_1\left(\frac{8}{9}\right) = 2 \cdot \frac{8}{9} + 1 = \frac{25}{9}$.

Найдем оптимальную стратегию второго игрока. В данном случае оптимальная стратегия второго игрока получается применением смеси двух полезных стратегий s_{y_1} и s_{y_2}, поскольку точка А получена пересечением двух прямых. Очевидно, что в этом случае $y_3 = 0$, т.к. третья стратегия не используется в оптимальном решении.

Для этого решим, например, систему

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + a_{13}y_3 = v; \\ y_1 + y_2 = 1. \end{cases}$$

Или

$$\begin{cases} 3y_1 + 2y_2 = \frac{25}{9}; \\ y_1 + y_2 = 1. \end{cases}$$

Умножая второе уравнение на 2, и вычитая его из первого уравнения, получим

$$y_1 = \frac{7}{9}.$$

Подставляя найденное значение y_1 во второе уравнение, находим $y_2 = \frac{2}{9}$. Следовательно, оптимальное решение упрощенной задачи для второго игрока $s_y^* = \left(\frac{7}{9}, \frac{2}{9}, 0\right)$.

Осталось вернуться к исходной задаче. Стратегии, соответствующие вычеркнутым строкам, не используются в оптимальном решении, следовательно, вероятности таких стратегий равны нулю. Отсюда получаем следующее оптимальное решение данной задачи:

$$s_x^* = \left(0, \frac{8}{9}, \frac{1}{9}\right), s_y^* = \left(0,0, \frac{7}{9}, 0, \frac{2}{9}, 0\right), v = \frac{25}{9}.$$

Рассмотренный пример соответствует классу задач размера $2 \times n$. Задача размера $m \times 2$ решается аналогично. Здесь возможны два варианта решения. Либо свести задачу к методу решения предыдущего класса задач, для этого достаточно изменить знак выигрыша на противоположный (и поменять знаки элементов платежной матрицы), тем самым превратить первого игрока из выигрывающего в проигрывающего. Во тором варианте решения рассматривается задача непосредственно для второго игрока, но строится не нижняя, а верхняя граница гарантированного выигрыша. На границе ищется точка с минимальной ординатой. Эта ордината численно равна цене игры.

Сведение задачи теории игр к задаче линейного программирования

Если задача теории игр не имеет решения в чистых стратегиях и не может быть решена графически, для получения решения игры используют методы линейного программирования.

Рассмотрим задачу о нахождении решения игры, заданной платежной матрицей $\|a_{ij}\|$ размера $m \times n$.

Наша оптимальна стратегия s_x^* должна обеспечивать нам выигрыш, не меньший v, при любом поведении противника, и выигрыш, равный v, при его оптимальной стратегии s_y^*. Аналогично стратегия s_y^* должна обеспечивать второму игроку проигрыш, не больший v, при любом поведении первого игрока и равный v при нашей оптимальной стратегии.

Величина цены игры v в данном случае нам неизвестна. Будем считать, что она равна некоторому положительному числу. Для того чтобы $v > 0$, очевидно, достаточно, чтобы все элементы матрицы $\|a_{ij}\|$ были неотрицательными. Этого всегда можно добиться, прибавляя ко всем элементам a_{ij} некоторое положительное число c. Оптимальные стратегии игроков при этом не изменятся, а цена игры увеличится на c.

Пусть мы выбрали свою оптимальную стратегию s_x^*. Тогда наш средний выигрыш при стратегии s_{y_j} противника будет равен:

$$a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{mj}x_m .$$

Наша оптимальная стратегия обладает тем свойством, что при любой стратегии противника обеспечивает выигрыш не меньший, чем v. Отсюда получаем ряд условий:

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + \cdots + a_{m1}x_m \geq v; \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{m2}x_m \geq v; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{mn}x_m \geq v. \end{cases} \tag{2.2}$$

Разделим неравенства (2.2) на положительную величину v и обозначим

$$\frac{x_1}{v} = \xi_1, \frac{x_2}{v} = \xi_2, \dots, \frac{x_m}{v} = \xi_m.$$

Тогда условия (2.2) запишутся в виде

$$\begin{cases} a_{11}\xi_1 + a_{21}\xi_2 + \dots + a_{m1}\xi_m \geq 1; \\ a_{12}\xi_1 + a_{22}\xi_2 + \dots + a_{m2}\xi_m \geq 1; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + a_{mn}\xi_m \geq 1, \end{cases} \quad (2.3)$$

где ξ_i– неотрицательные числа, $i = \overline{1,m}$. Так как $x_1 + x_2 + \dots + x_m = 1$, то величины ξ_i удовлетворяют условию

$$\xi_1 + \xi_2 + \dots + \xi_m = \frac{1}{v}. \quad (2.4)$$

Мы хотим сделать свой гарантированный выигрыш максимально возможным. Очевидно, что при этом правая часть равенства (2.4) должна достигать минимального значения.

Таким образом, задача нахождения решения игры сводится к следующей задаче линейного программирования:

$$f = \xi_1 + \xi_2 + \dots + \xi_m \to min,$$

$$\begin{cases} a_{11}\xi_1 + a_{21}\xi_2 + \dots + a_{m1}\xi_m \geq 1; \\ a_{12}\xi_1 + a_{22}\xi_2 + \dots + a_{m2}\xi_m \geq 1; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + a_{mn}\xi_m \geq 1, \end{cases}$$

$$\xi_i \geq 0, i = \overline{1,n}.$$

Рассуждая аналогично для второго игрока, получим задачу линейного программирования, двойственную к задаче линейного программирования для первого игрока.

Выпишем пару взаимно двойственных задач линейного программирования:

$$F = v \to max,$$

$$\sum_{i=1}^{m} a_{ij}x_i \geq v, \ \ j = \overline{1,n},$$

$$\sum_{i=1}^{m} x_i = 1,$$

$$x_i \geq 0, i = \overline{1,m}.$$

$$\Phi = u \to min,$$

$$\sum_{j=1}^{n} a_{ij}y_j \leq u, i = \overline{1,m},$$

$$\sum_{j=1}^{n} y_j = 1,$$

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

Процесс решения упрощается, если перейти к переменным

$$\xi_i = \frac{x_i}{v} \ (i = \overline{1,m}), \eta_i = \frac{y_j}{u} \ (j = \overline{1,n}).$$

Это возможно, если $a_{ij} \geq 0$. Иначе, добиваемся этого условия, прибавляя ко всем элементам a_{ij} некоторое положительное число c. Имеем:

$$f = \sum_{i=1}^{m} \xi_i \to min,$$

$$\varphi = \sum_{j=1}^{n} \eta_j \to max,$$

$$\sum_{i=1}^{m} a_{ij}\xi_i \geq 1, j = \overline{1,n},$$
$$\xi_i \geq 0, i = \overline{1,m}.$$

$$\sum_{j=1}^{n} a_{ij}\eta_j \leq 1, i = \overline{1,m},$$
$$\eta_j \geq 0, j = \overline{1,n}.$$

Целесообразно сначала решать задачу второго игрока симплекс-методом. В последней симплекс-таблице, содержащей оптимальное решение второго игрока, можно найти и оптимальное решение двойственной задачи– задачи первого игрока.

Пример. Получить решение задачи теории игр, если матрица выигрышей первого игрока:

$$A = \begin{pmatrix} 0 & -2 & 7 & 4 \\ -1 & 1 & 4 & -2 \\ 2 & 0 & -1 & 1 \end{pmatrix}.$$

Нетрудно проверить, что матрица A не имеет седловой точки.
Геометрический способ решения не применим.

Сначала перейдем к матрице с положительными элементами. Прибавим к элементам матрицы A число 2, получим:

$$\tilde{A} = \begin{pmatrix} 2 & 0 & 9 & 6 \\ 1 & 3 & 6 & 0 \\ 4 & 2 & 1 & 3 \end{pmatrix}.$$

Построим пару взаимно двойственных задач линейного программ-мирования, эквивалентную игре с матрицей $\tilde{A}$.

$$f = \xi_1 + \xi_2 + \xi_3 \to min,$$
$$\begin{cases} 2\xi_1 + \xi_2 + 4\xi_3 \geq 1, \\ \qquad 3\xi_2 + 2\xi_3 \geq 1, \\ 9\xi_1 + 6\xi_2 + \xi_3 \geq 1, \\ 6\xi_1 \qquad + 3\xi_3 \geq 1, \\ \xi_i \geq 0, \ \ i = 1,2,3. \end{cases}$$

$$\varphi = \eta_1 + \eta_2 + \eta_3 + \eta_4 \to max,$$
$$\begin{cases} 2\eta_1 \qquad + 9\eta_3 + 6\eta_4 \leq 1, \\ \eta_1 + 3\eta_2 + 6\eta_3 \qquad \leq 1, \\ 4\eta_1 + 2\eta_2 + \ \eta_3 + 3\eta_4 \leq 1, \\ \eta_j \geq 0, \ \ j = 1,2,3,4. \end{cases}$$

Сведем задачу второго игрока к канонической задаче линейного программирования:

$$\varphi = \eta_1 + \eta_2 + \eta_3 + \eta_4 \to max,$$
$$\begin{cases} 2\eta_1 \qquad + 9\eta_3 + 6\eta_4 + \eta_5 \qquad = 1, \\ \eta_1 + 3\eta_2 + 6\eta_3 \qquad\qquad + \eta_6 = 1, \\ 4\eta_1 + 2\eta_2 + \ \eta_3 + 3\eta_4 \qquad + \eta_7 = 1, \\ \qquad \eta_j \geq 0, \ \ j = \overline{1,n}. \end{cases}$$

Здесь и далее через η_B обозначим вектор базисных переменных. Вектор коэффициентов целевой функции при базисных переменных обозначается c_B. Составим симплекс-таблицу, соответствующую задаче линейного программ-мирования. Очевидно, что переменные η_5, η_6, η_7 составляют допустимое базисное решение.

		c_j							Постоян-ные
		1	1	1	1	0	0	0	
c_B	η_B	η_1	η_2	η_3	η_4	η_5	η_6	η_7	

0 0 0	η_5 η_6 η_7	2 1 4	0 $\boxed{3}$ 2	9 6 1	6 0 3	1 0 0	0 1 0	0 0 1	1 1 1
$\overline{c}_j$		1	1	1	1	0	0	0	$\varphi = 0$

Относительная оценка переменной x_j равна $\overline{c}_j = c_j - c_B \cdot \overline{A}_j$, где $\overline{A}_j -$ -й столбец системы, соответствующий рассматриваемому базису.

Поскольку ненулевые относительные оценки равны, то в качестве максимальной оценки можно взять, например, $\overline{c}_2$. Следовательно, в базис будем вводить переменную η_2. Определим переменную, выводимую из базиса. Находим наименьшее отношение $\min\left\{\frac{1}{3}, \frac{1}{2}\right\} = \frac{1}{3}$. Из базиса выводим переменную η_6.

		c_j							Постоян-ные
		1	1	1	1	0	0	0	
c_B	η_B	η_1	η_2	η_3	η_4	η_5	η_6	η_7	
0 1 0	η_5 η_2 η_7	2 1/3 10/3	0 1 0	9 2 -3	6 0 $\boxed{3}$	1 0 0	0 1/3 -2/3	0 0 1	1 1/3 1/3
$\overline{c}_j$		2/3	0	-1	1	0	-1/3	0	$\varphi = 1/3$

$\overline{c}_4 -$ наибольшая положительная оценка. Следовательно, переменную η_4 вводим в базис. Наименьшее отношение $\min\left\{\frac{1}{6}, \frac{1}{9}\right\} = \frac{1}{9}$. Значит переменную η_7 выводим из базиса.

		c_j							Постоян-ные
		1	1	1	1	0	0	0	
c_B	η_B	η_1	η_2	η_3	η_4	η_5	η_6	η_7	
0 1 1	η_5 η_2 η_4	-14/3 1/3 10/9	0 1 0	15 2 -1	0 0 1	1 0 0	4/3 1/3 -2/9	-2 0 1/3	1/3 1/3 1/9
$\overline{c}_j$		-4/9	0	0	0	0	-1/9	-1/3	$\varphi = 4/9$

Положительных оценок нет. Следовательно, допустимое базисное решение $\left(0, \frac{1}{3}, 0, \frac{1}{9}, \frac{1}{3}, 0, 0\right)$ является оптимальным решением канонической задачи линейного программирования. Отсюда, оптимальное решение исходной задачи линейного программирования$-\left(0, \frac{1}{3}, 0, \frac{1}{9}\right)$. Значение целевой функции $\varphi = 4/9$.

Компоненты оптимального решения двойственной задачи определяются с помощью строки оценок последней симплекс-таблицы, а именно

$(-\bar{c}_5, -\bar{c}_6, \bar{c}_7)$. Следовательно, $\left(0, \frac{1}{9}, \frac{1}{3}\right)$ – оптимальное решение двойственной задачи, причем $f = \varphi = 4/9$. Учитывая равенство (1.4) и равенство $v = u$, получим, что $v = \frac{1}{\xi_1 + \xi_2 + \xi_3} = \frac{9}{4}$. Перейдем к первоначальным переменным задачи

$$x_i = \xi_i v, y_j = \eta_j u,$$

получим $s_x^* = \left(0, \frac{1}{4}, \frac{3}{4}\right)$, $s_y^* = \left(0, \frac{3}{4}, 0, \frac{1}{4}\right)$– оптимальные стратегии задачи теории игр. Поскольку к элементам исходной платежной матрицы мы прибавляли число 2, то цена данной игры 9/4–2=1/4.

СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Классификация систем массового обслуживания

Системы, в которых, с одной стороны, возникаю массовые запросы (требования) на выполнение каких-либо видов услуг, а с другой стороны, происходит удовлетворение этих запросов, называются *системами массового обслуживания*.

Система массового обслуживания включает в себе следующие элементы: источник требований, входящий поток требований, очередь, обслуживающее устройство (обслуживающий аппарат, канал обслуживания), выходящий поток требований.

Системы массового обслуживания классифицируют по разным признакам. К таким признакам относятся условия ожидания начала обслуживания. В соответствии с этим признаком системы подразделяются на следующие виды:

1. системы массового обслуживания с потерями (отказами);
2. системы массового обслуживания с ожиданием;
3. системы массового обслуживания с ограниченной длиной очереди;
4. системы массового обслуживания с ограниченным временем ожидания.

Системы массового обслуживания, у которых требования, поступающие в момент, когда все проборы обслуживания заняты, получают отказ и теряются, называются *системами с потерями* или *отказами*. Например, оператор сотовой связи, иногда выдающий сообщение «сеть занята».

Системы массового обслуживания, у которых возможно появление как угодно длинной очереди требований к обслуживающему устройству, называют *системами с ожиданием*.

Системы массового обслуживания, допускающие очередь, но с ограниченным числом мест в ней, называются *системами с ограниченной длиной очереди*.

Системы массового обслуживания, допускающие очередь, но с ограниченным сроком пребывания каждого требования в ней, называются *системами с ограниченным временем ожидания*.

По числу каналов или приборов системы делятся на *одноканальные* или *многоканальные*.

По месту нахождения источника требований системы массового обслуживания делятся на *разомкнутые*, когда источник находится вне системы, и *замкнутые*, когда источник находится в самой системе. К последнему виду относится, например, автопарк, в котором автомобили являются источником неисправностей, а следовательно, и требований на их обслуживание (ремонт).

Как правило, входящий поток требований является случайной величиной с некоторым законом распределения. Одной из форм классификации систем массового обслуживания является кодовая (символьная) *классификация Д. Кендалла*. В этой классификации характеристику системы записывают в виде трех, четырех или пяти символов, например A|B|S, где A− тип закона распределения входящего потока требований, B− тип закона распределения времени обслуживания, S− число каналов обслуживания.

Для экспоненциального распределения принимают символ M, для любого (произвольного) распределения− символ G. Запись M|M|3 означает, что входящий поток требований пуассоновский (простейший), время обслуживания распределено по экспоненциальному закону, в системе имеется три канала обслуживания.

Четвертый символ указывает допустимую длину очереди, а пятый− порядок отбора (приоритета) требований.

Показатели эффективности систем массового обслуживания

Показатели эффективности делятся на показатели, характеризующие качество и условия работы обслуживающей системы, и показатели, отражающие экономические особенности системы.

Показатели первой группы обычно формируют на основе полученных из расчетов значений вероятностей состояний системы. Показатели второй группы рассчитывают на основе показателей первой группы.

Среди показателей первый группы можно выделить следующие.

1. Вероятность $P_{\text{отк}}$ того, что поступающее в систему требование откажется присоединиться к очереди и теряется (вероятность отказа). Этот показатель для системы массового обслуживания с отказами равен вероятности P_m того, что в системе находится столько требований, сколько она содержит приборов (каналов) обслуживания:

$$P_{\text{отк}} = P_m,$$

где m − число каналов обслуживания.

Для системы с ограниченной длиной очереди $P_{\text{отк}}$ равно вероятности того, что в системе находится $m + l$ требований:

$$P_{\text{отк}} = P_{m+l},$$

где l − допустимая длина очереди.

Противоположным показателем является вероятность обслуживания требования

$$P_{\text{обсл}} = 1 - P_{\text{отк}}.$$

2. Среднее количество требований, ожидающих начала обслуживания:

$$M_{\text{ож}} = \sum_{n=m+1}^{m+l} (n - m)P_n,$$

где P_n − вероятность того, что в системе находится n требований.

В частности, если поток требований имеет пуассоновское распределение и закон распределения времени обслуживания экспоненциальный, то формулы для $M_{\text{ож}}$ принимают следующий вид:

для системы с ограниченной длиной очереди

$$M_{\text{ож}} = \frac{P_0 \rho^m}{m!} \sum_{n=1}^{l} n \left(\frac{\rho}{m}\right)^n,$$

где $\rho = \frac{\lambda}{\mu}$, $\lambda -$ интенсивность входящего потока требований (среднее число требований, поступающих в единицу времени), $\mu -$ интенсивность обслуживания (среднее число обслуженных требований в единицу времени);

для системы с ожиданием

$$M_{\text{ож}} = \frac{P_0 \rho^{m+1}}{m \cdot m!} \cdot \frac{1}{\left(1 - \frac{\rho}{m}\right)^2}.$$

3. Относительную q и абсолютную A пропускные способности системы вычисляют по формулам

$q = 1 - P_{\text{отк}}, \qquad A = \lambda q.$

4. Среднее число занятых обслуживанием приборов в случае экспоненциального характера потока требований и времени обслуживания

$m_{\text{з}} = \rho q.$

5. Общее количество M требований, находящихся в системе, определяют следующим образом:

для системы массового обслуживания с отказами

$M = m_{\text{з}};$

для системы массового обслуживания с ограниченной длиной очереди и ожиданием

$M = m_{\text{з}} + M_{\text{ож}}.$

6. Среднее время ожидания требованиям начала обслуживания $T_{\text{ож}}$. Если известна функция распределения вероятности времени ожидания требованиям начала обслуживания

$F(t) = P(T_{\text{ож}} < t),$

то среднее время ожидания находится как математическое ожидание случайной величины $T_{\text{ож}}$:

$$T_{\text{ож}} = M(T_{\text{ож}}) = \int_0^{\infty} t dF.$$

В частности, при показательном законе распределения требований во входящем потоке $T_{\text{ож}}$ можно определить по формуле

$$T_{\text{ож}} = \frac{M_{\text{ож}}}{\lambda}.$$

Показатели, характеризующие экономические особенности, формируют обычно в соответствии с конкретным видом системы и ее назначением.

Одним из общих экономических показателей является *экономическая эффективность*

$E = P_{обсл}\lambda cT - G_п,$

где $c -$ средний экономический эффект, полученный при обслуживании одного требования, $T -$ рассматриваемый интервал времени, $G_п -$ величина потерь в системе.

Величину потерь можно определить по следующим формулам:

для *системы с отказами*

$G_п = \left(q_{эк}m_з + q_у P_{отк}\lambda + q_{пр}m_{св}\right)T,$

где $q_{эк} -$ стоимость эксплуатации одного прибора в единицу времени, $q_у -$ стоимость убытков в результате ухода требований из системы в единицу времени, $q_{пр} -$ стоимость единицы времени простоя прибора системы, $m_{св} = m - m_з$;

для *системы с ожиданием*

$G_п = \left(q_{ож}M_{ож} + q_{эк}m_з + q_{пр}m_{св}\right)T,$

где $q_{ож} -$ стоимость потерь, связанных с простоем требований в очереди в единицу времени.

Марковские случайные процессы. Марковская цепь.

Случайный процесс, протекающий с системе S, называется *марковским процессом*, если он обладает следующим свойством: для любого момента времени t_0 вероятность каждого состояния системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем $S(t_0)$ и не зависит от того, когда и каким образом система перешла в настоящее состояние.

Состояния системы могут изменяться либо дискретно, либо непрерывно.

Случайный марковский процесс называется *процессом с дискретными состояниями*, если возможные состояния системы S_1, S_2, …,S_n можно пронумеровать, а сам процесс состоит в том, что время от времени система S скачком (мгновенно) переходит из одного состояния в другое. Примером такого процесса является процесс, протекающий в техническом устройстве. Можно представить два состояния такой системы: $S_1 -$ система работает, $S_2 -$ система вышла из строя.

Случайный марковский процесс называется *процессом с непрерывными состояниями*, если эти состояния меняются непрерывно, постепенно. Примером такого процесса является процесс движения автомобиля.

В системе с дискретными состояниями переход от состояния в состояние может происходить в определенные, фиксированные моменты времени либо в случайные моменты.

Случайный марковский процесс называется *процессом с дискретным временем*, если переходы системы из состояния в состояние возможны только в строго определенные, заранее фиксированные моменты времени t_1, t_2, … В промежутки времени между этими моментами система находится в определенным состоянии.

Случайный марковский процесс называется *процессом с непрерывным временем*, если переход системы из состояния в состояние возможен в любой заранее не известный случайный момент времени.

Так как для марковского процесса с дискретными состояниями и дискретным временем моменты t_1, t_2, …, t_k,… фиксированы, то процесс можно рассматривать как функцию целочисленного аргумента k $(k = 1,2, \dots)$ − номера шага. В этом случае переходы системы из состояния в состояние представляют собой последовательность (цепочку) событий или состояний $S_1^{(1)}, S_1^{(2)}, S_3^{(3)}, S_5^{(4)}, S_2^{(5)}, \dots$ Число в скобках обозначает номер шага, нижний индекс− номер состояния.

Случайная последовательность событий с фиксированным шагом называется *дискретной марковской цепью*, если для каждого шага вероятность перехода из любого состояния S_i в любое другое состояние S_j не зависит от того, когда и как система перешла в состояние S_i.

Если переход системы из состояния в состояние происходит в случайные моменты времени, то соответствующая цепочка состояний называется *непрерывной цепью Маркова*.

При анализе случайных процессов с дискретными состояниями используют графы состояний. Граф состояний геометрически изображает возможные состояния системы и ее возможные переходы из состояния в состояние (Рис.3.1).

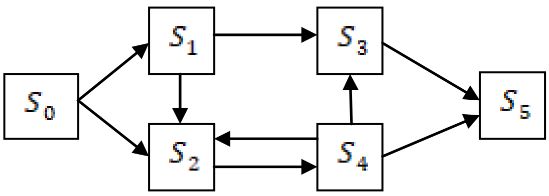


Рис.3.1. Граф состояний

Марковская непрерывная цепь называется *процессом гибели и размножения*, если граф ее состояний представляет собой цепочку, в которой каждое из промежуточных состояний связано прямой и обратной связью с каждым соседним состоянием (Рис. 3.2).

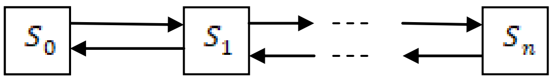


Рис.3.2. Процесс гибели и размножения

Уравнения Колмогорова для вероятностей состояний

Системы массового обслуживания эффективно описываются марковскими цепями.

Системы, представляемые в виде непрерывной цепи Маркова, обычно исследуют с помощью уравнений Колмогорова для вероятностей состояний.

Плотностью вероятности перехода λ_{ij} из состояния S_i в состояние S_j называется предел отношения вероятности этого перехода за время Δt, к длине промежутка Δt, когда последний стремится к нулю:

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t},$$

где $P_{ij}(\Delta t)$ –вероятность того, что система, находившаяся в момент времени t в состоянии S_i, за время Δt перейдет в состояние S_j.

Марковская непрерывная цепь называется *однородной*, если плотности вероятностей λ_{ij} не зависят от времени t, в противном случае она называется *неоднородной*.

Для однородных марковских непрерывных цепей, характеризующих процессы гибели и размножения, уравнения Колмогорова имеют вид

$$\frac{dP_0}{dt} = -\lambda_{01}P_0(t) + \lambda_{10}P_1(t), \tag{3.1}$$

$$\frac{dP_i}{dt} = \lambda_{i-1,i}P_{i-1}(t) - \left(\lambda_{i,i-1} + \lambda_{i,i+1}\right)P_i(t) + \lambda_{i+1,i}P_{i+1}(t), i = \overline{1,n}, \tag{3.2}$$

где $P_i(t)$ – вероятность состояния S_i, когда в системе находится i требований в момент времени t, $n + 1$ – общее число возможных состояний $S_0, S_1, \ldots, S_n$.

При гипотезе о стационарном режиме работы системы (вероятности состояний не зависят от времени) уравнения Колмогорова (3.1)-(3.2) принимают вид

$$-\lambda_{01}P_0(t) + \lambda_{10}P_1(t) = 0, \tag{3.3}$$

$$\lambda_{i-1,i}P_{i-1}(t) - \left(\lambda_{i,i-1} + \lambda_{i,i+1}\right)P_i(t) + \lambda_{i+1,i}P_{i+1}(t) = 0, i = \overline{1,n}. \tag{3.4}$$

В большинстве практических задач оказывается допустимой гипотеза о стационарном режиме работы системы. Поэтому могут быть использованы уравнения Колмогорова (3.3)-(3.4).

Математические модели систем массового обслуживания, приводимые ниже, соответствуют уравнениям Колмогорова (3.3)-(3.4) при условиях простейшего потока входящих требований и экспоненциального закона распределения времени обслуживания.

Системы массового обслуживания с отказами

Граф состояний многоканальной системы массового обслуживания с отказами при условиях простейшего потока входящих требований и экспоненциального закона распределения времени обслуживания имеет вид, изображенный на Рис. 3.3.

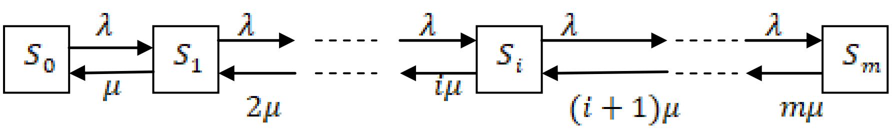


Рис. 3.3. Многоканальная система массового обслуживания с отказами

Здесь λ — интенсивность входящего потока требований, μ — производительность одного канала (прибора) обслуживания, S_0, S_1,…, S_m — состояния системы (индекс указывает число требований в системе), m — общее число каналов.

Вероятности состояний определяются по формулам

$$P_i = \frac{\rho^i}{i!} P_0, \qquad i = \overline{1,m},$$

где $\rho = \frac{\lambda}{\mu}$, а вероятность P_0 находится из выражения

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} \right)^{-1}.$$

Пример. На станции техобслуживания оборудовано три места для ремонта автомобилей. Если все места заняты, то вновь прибывший клиент обращается в другую станцию. Пусть среднее время работы с одним клиентом составляет 3 ч. Интенсивность потока прибывающих автомобилей 0,25 в час. Найти вероятность отказа клиенту и среднее число занятых мест на станции.

Имеем: Находим:

$$\mu = \frac{1}{T_{\text{обс}}} = \frac{1}{3}, \qquad \rho = \frac{\lambda}{\mu} = \lambda T_{\text{обс}} = 0{,}25 \cdot 3 = 0{,}75,$$

$$P_0 = \left(\sum_{i=0}^{3} \frac{\rho^i}{i!} \right)^{-1} = \left(1 + 0{,}75 + \frac{0{,}75^2}{2!} + \frac{0{,}75^3}{3!} \right)^{-1} = (2{,}1)^{-1},$$

$$P_{\text{отк}} = \frac{\rho^m}{m!} P_0 = \frac{0{,}75^3}{3!} \cdot \frac{1}{2{,}1} = 0{,}033,$$

$$m_{\text{з}} = \sum_{n=1}^{m} n P_n = P_0 \sum_{n=1}^{m} \frac{\rho^n}{(n-1)!} = \frac{1}{2{,}1} \left(0{,}75 + 0{,}75^2 + \frac{0{,}75^3}{2} \right) \approx 0{,}72.$$

Таким образом, $P_{\text{отк}} = 0{,}033$; $m_{\text{з}} = 0{,}72$.

Системы массового обслуживания с ограниченной длиной очереди

Граф состояний многоканальной системы массового обслуживания, имеющий m каналов с ограниченной очередью, число мест которой ограниченно величиной l, при условиях простейшего потока входящих требований и экспоненциального закона распределения времени обслуживания имеет вид, изображенный на Рис. 3.4.

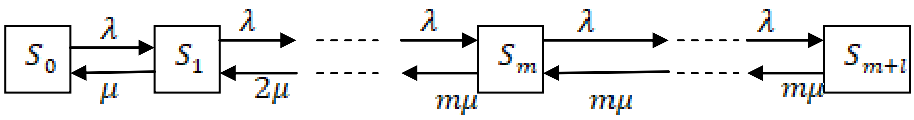


Рис. 3.4. Многоканальная система массового обслуживания с ограниченной очередью

Вероятности состояний $S_0, S_1, \ldots, S_m$ находят по формуле

$$P_i = \frac{\rho^i}{i!} P_0, \qquad i = \overline{1, m}.$$

Вероятности состояний $S_{m+1}, S_{m+2}, \ldots, S_{m+l}$ определяются с помощью формул

$$P_i = \frac{\rho^i}{m^{i-m} \cdot m!} P_0, \qquad i = \overline{m+1, m+l}.$$

Вероятность P_0 подсчитывают по формуле

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \sum_{i=m+1}^{m+l} \frac{\rho^i}{m^{i-m} \cdot m!} \right)^{-1}.$$

В большинстве практических задач отношение $\frac{\rho}{m} < 1$. Тогда P_0 находят из соотношения

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \frac{\rho^{m+1}}{m \cdot m!} \frac{1 - \left(\frac{\rho}{m}\right)^l}{1 - \frac{\rho}{m}} \right)^{-1}.$$

Пример. На автозаправочной станции установлены три колонки для выдачи бензина. Около станции находится площадка на три машины для их ожидания в очереди. На станцию прибывают в среднем две машины в минуту. Среднее время заправки машины одной машины 1 мин. Требуется определить вероятность отказа и среднюю длину очереди.

Имеем: $m = 3$, $l = 3$, $\lambda = 2$, $T_{\text{обс}} = 1$ (мин), $\mu = \frac{1}{T_{\text{обс}}} = 1$. Далее находим: $\rho = \frac{\lambda}{\mu} = \frac{2}{1} = 2, \frac{\rho}{m} = \frac{2}{3}$,

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \frac{\rho^{m+1}}{m \cdot m!} \frac{1 - \left(\frac{\rho}{m}\right)^l}{1 - \frac{\rho}{m}} \right)^{-1} =$$

$$= \left(1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3 \cdot 3!} \frac{1 - \left(\frac{2}{3}\right)^3}{1 - \frac{2}{3}} \right)^{-1} \approx 0{,}122,$$

$$P_{\text{отк}} = P_{m+l} = \frac{\rho^{m+l}}{m^l \cdot m!} P_0 = \left(\frac{\rho}{m}\right)^l \frac{\rho^m}{m!} P_0 = \left(\frac{2}{3}\right)^3 \frac{2^3}{3!} \cdot 0{,}122 = 0{,}048,$$

$$M_{\text{ож}} = \frac{P_0 \rho^m}{m!} \sum_{n=1}^{l} n \left(\frac{\rho}{m}\right)^n =$$

$$= \frac{0{,}122 \cdot 2^3}{3!} \left(\frac{2}{3} + 2 \left(\frac{2}{3}\right)^2 + 3 \left(\frac{2}{3}\right)^3 \right) = 0{,}35.$$

Таким образом, $P_{\text{отк}} = 0{,}048$, $M_{\text{ож}} = 0{,}35$ (машины).

Системы массового обслуживания с ожиданием

Граф состояний системы массового обслуживания с ожиданием аналогичен графу состояний системы с ограниченной длиной очереди при условии, что граница очереди отодвигается в бесконечность, т.е. предполагается, что в очереди может находиться неограниченное число требований. Такой граф состояний изображен на Рис. 3.5.

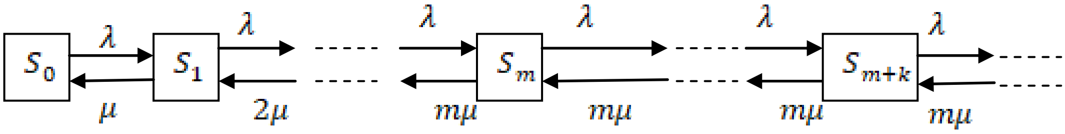


Рис. 3.5 Система массового обслуживания с ожиданием

Вероятности состояний находятся по формулам

$$P_i = \frac{\rho^i}{i!} P_0, \qquad i = \overline{1,m}.$$

$$P_i = \frac{\rho^i}{m!\, m^{i-m}} P_0, \qquad i = m+1, \dots, m+k, \dots$$

При $\frac{\rho}{m} < 1$ для определения P_0 используют формулу

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \frac{\rho^{m+1}}{m!\,(m-\rho)} \right)^{-1}.$$

Пример. В порту имеется два причала для разгрузки грузовых судов. Интенсивность потока судов равна 0,8 (судов в сутки). Среднее время разгрузки одного судна составляет 2 сут. Предполагается, что очередь ожидающих разгрузки судов может быть неограниченной длины.

Найти среднее число занятых причалов и среднее время пребывания судна в порту.

Имеем: $m = 2, \lambda = 0{,}8,\ \mu = \frac{1}{T_{\text{обс}}} = \frac{1}{2} = 0{,}5, \rho = \frac{\lambda}{\mu} = \frac{0{,}8}{0{,}5} = 1{,}6$. Находим:

$$P_0 = \left(\sum_{i=0}^{m} \frac{\rho^i}{i!} + \frac{\rho^{m+1}}{m!\,(m-\rho)} \right)^{-1} =$$

$$= \left(1 + \frac{1{,}6}{1!} + \frac{1{,}6^2}{2!} + \frac{1{,}6^3}{2!\,(2-1{,}6)} \right)^{-1} = 0{,}11,$$

$$m_з = \rho q, q = 1 \ \Rightarrow m_з = 1{,}6,$$

$$M_{\text{ож}} = \frac{P_0 \rho^{m+1}}{m \cdot m!} \cdot \frac{1}{\left(1 - \frac{\rho}{m}\right)^2} = \frac{0{,}11 \cdot 1{,}6^3}{2 \cdot 2 \cdot (1-0{,}8)^2} 2{,}8, \qquad T_{\text{ож}} = \frac{M_{\text{ож}}}{\lambda} = 3{,}5.$$

Итак: $m_з = 1{,}6$ (причалов), $T_{\text{ож}} = 3{,}5$ (сут).

Системы массового обслуживания с ограниченным временем ожидания

В системах массового обслуживания с ограниченным временем ожидания время ожидания в очереди каждого требования ограничено случайной величиной $t_{\text{ож}}$, среднее значение которой $\bar{t}_{\text{ож}}$.

Величина, обратная среднему времени ожидания, означает среднее количество требований, покидающих очередь в единицу времени, вызванное появлением в очереди одного требования: $\nu = \frac{1}{\bar{t}_{\text{ож}}}$.

При наличии в очереди k требований интенсивность потока покидающих очередь требований составляет $k\nu$.

Граф состояний такой системы изображен на Рис.3.6.

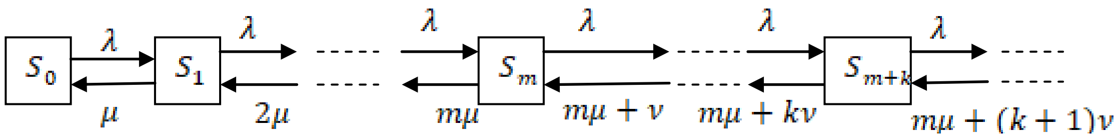


Рис.3.6. Системы массового обслуживания с ограниченным временем ожидания

Формулы для определения вероятностей состояний такой системы имеют вид

$$P_i = \frac{\rho^i}{i!}P_0, \qquad i = \overline{1,m}.$$

$$P_i = \frac{\rho^m}{m!}\cdot\frac{\lambda^k}{\prod_{j=1}^{k}(m\mu + j\nu)}, \qquad i = m+1,\dots,m+k,\dots,$$

где $\prod_{j=1}^{k}(m\mu + j\nu)$ – произведение множителей вида $m\mu + j\nu$ для каждого $j = 1,2,\dots,k$.

Вероятность P_0 определяют по формуле

$$P_0 = \left(\sum_{i=0}^{m}\frac{\rho^i}{i!} + \frac{\rho^m}{m!}\sum_{k=1}^{\infty}\frac{\lambda^k}{\prod_{j=1}^{k}(m\mu + j\nu)}\right)^{-1}.$$

В практических задачах сумму бесконечного ряда вычислить достаточно просто, так как члены ряда быстро убывают с увеличением номера.

Пример. В пункте химчистки имеется три аппарата для чистки. Интенсивность потока посетителей $\lambda = 6$ (посетителей в час). Интенсивность обслуживания посетителей одним аппаратом $\mu = 3$ (посетителей в час). Среднее количество посетителей, покидающих очередь, не дождавшись обслуживания, $\nu = 1$ (посетитель в час). Найти абсолютную пропускную способность пункта.

Имеем: $m = 3, \lambda = 6, \mu = 3, \nu = 1$. Находим: $\rho = \frac{\lambda}{\mu} = \frac{6}{3} = 2,$

$$P_0 = \left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^3}{3!}\cdot\left(\frac{6}{3\cdot 3+1} + \frac{6^2}{(3\cdot 3+1)\cdot(3\cdot 3+2\cdot 1)}\right)\right)^{-1} =$$
$$= 0{,}13.$$

Вероятность занятости всех приборов равна $P_{зан} = 1 - P_0 = 0{,}87$. тогда абсолютная пропускная способность может быть получена как произведение: $A = mP_{зан} = 3 \cdot 0{,}87 = 2{,}61$. Таким образом, $A = 2{,}61$ (посетителя в час).

Замкнутые системы массового обслуживания

В замкнутых системах массового обслуживания источник требований находится внутри системы и интенсивность потока требований зависит от состояния самой системы. Чаще всего потоком требований в такой системе является поток неисправностей от некоторой группы работающих устройств. Пусть имеется m работающих устройств, которые могут выходить из строя за счет неисправностей. Имеется так же n приборов (каналов) обслуживания этих требований. В качестве таких каналов могут выступать и люди. Обычно предполагают, что $n < m$.

Обозначим через S_0 состояние, при котором все устройства работают, а приборы обслуживания не заняты; $S_1 -$ состояние, при котором одно устройство вышло из строя и обслуживается одним прибором обслуживания; $S_n - n$ устройств не работают все приборы заняты обслуживанием; $S_m -$ все устройства не работают, из них n обслуживаются и $m - n$ ждут обслуживания. Граф состояний такой системы изображен на Рис. 3.7.

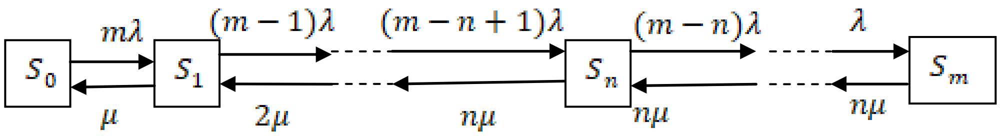


Рис. 3.7. Замкнутые системы массового обслуживания

Вероятности состояний замкнутой системы определяются по формулам:

$$P_i = \frac{\prod_{j=0}^{i-1}(m-j)}{i!}\rho^i P_0, \qquad i = \overline{1,n};$$

$$P_i = \frac{\prod_{j=0}^{i-1}(m-j)}{n!\,n^{i-n}}\rho^i P_0, \qquad i = \overline{n+1,m};$$

$$P_0 = \left(1 + \sum_{i=1}^{n} \frac{\prod_{j=0}^{i-1}(m-j)}{i!}\rho^i + \sum_{i=n+1}^{m} \frac{\prod_{j=0}^{i-1}(m-j)}{n!\,n^{i-n}}\rho^i\right)^{-1}.$$

Пример. Рабочий обслуживает группу из трех станков. Каждый станок останавливается в среднем два раза в час. Процесс наладки занимает в среднем 10 мин (1/6 часа). Определить абсолютную пропускную способность наладки рабочим станков.

Имеем: n =1, $m = 3$, $\lambda = 2$, $T_{обс} = 1/6$, $\mu = \frac{1}{T_{обс}} = 6$. Находим: $\rho = \frac{\lambda}{\mu} = \frac{1}{3}$,

$$P_0 = \left(1 + m\rho + \frac{m(m-1)}{1!\,1^1}\rho^2 + \frac{m(m-1)(m-2)}{1!\,1^2}\rho^3\right)^{-1} =$$

$$= \left(1 + 3\cdot\frac{1}{3} + 3\cdot 2\cdot\left(\frac{1}{3}\right)^2 + 3\cdot 2\cdot 1\cdot\left(\frac{1}{3}\right)^3\right)^{-1} = 0{,}346.$$

Определяем вероятность того, что рабочий будет занят обслуживанием:

$P_з = 1 - P_0 = 1 - 0{,}346 = 0{,}654.$

Если рабочий занят обслуживанием, то он обслуживает 6 станков в час. Следовательно, абсолютная пропускная способность находится как произведение:

$A = \mu P_з = 6 \cdot 0{,}654 \approx 4.$

Таким образом, $A = 4$ (станка в час).

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Понятие имитационной модели

Имитационное моделирование как эффективный метод помощи лицам, принимающим решения (ЛПР) , является мощным аппаратом исследования, пригодным для решения многих важных прикладных задач.

Основная идея имитационного моделирования состоит в том, что реальному объекту сопоставляется не слишком сложная математическая модель, а алгоритм функционирования этого объекта.

Имитационной моделью реального процесса (объекта, явления) мы будем называть программу для компьютера, реализующую упрощенную модель этого процесса вместе с алгоритмом, описывающим течение этого процесса. Когда компьютер выполняет эту программу, он имитирует «течение» реального процесса. Меняя различные входные данные программы, можно имитировать течение реального процесса в различных условиях. Таким образом, возникает возможность осуществления диалога между исследователем (ЛПР, и т.п.) и компьютером.

Исследователь задает **вопрос**: «Что произойдет с объектом, если …?»

Ответ (дает ЭВМ, проигрывая заложенную в нее имитационную модель): «В заданных условиях с объектом произойдет следующее …».

Организованный таким образом диалог человека с компьютером позволяет проводить те или иные эксперименты (они называются имитационными, вычислительными, машинными), получая при этом информацию, которая может быть отнесена к реальному процессу (объекту, явлению). Разумеется, полученная информация будет иметь тем большее отношение к реальности, чем более удачно построена имитационная модель.

Схематично процесс имитационных исследований представлен на Рис.4.1.

С возникновением современного этапа прикладных исследований, характеризующегося математизацией таких наук, как экономика, биология, социология и др., моделирование вступило в ту фазу своего развития, когда приходится иметь дело с явлениями не просто сложными, а комплексными; с явлениями, в основе которых лежат процессы различной природы, для анализа которых требуется использование разнообразных, но связанных между собой моделей. Оказалось, что имитационное моделирование, в отличие от многих других методов, может с успехом обслуживать исследования современного этапа.

Когда говорят «имитационная модель», обычно понимают, что ей свойственна большая, чем в обычных математических моделях, близость к реальному объекту, возможность модели воспроизвести широкий спектр свойств объекта, использование в модели эмпирического материала, сохраняющая всегда возможность дальнейшего уточнения модели и др.

Говоря же об имитационном моделировании, в него включают как процесс создания модели, так и ее исследование (проведения вычислительных экспериментов) с помощью компьютера.

Р. Шеннон дает следующее определение: «… имитационное моделирование есть процесс конструирования модели реальной системы и постановки экспериментов на этой модели с целью оценить (в рамках ограничений, накладываемых некоторым критерием или совокупностью критериев) различные стратегии, обеспечивающие функционирование данной системы».

Как видно из этого определения, имитационное моделирование не очень приспособлено для выяснения причин того или иного явления (оно не дает ответа на вопрос: «Почему?»). Его роль гораздо более прагматична‒ давать ответ да вопрос: «Что будет, если … ?».

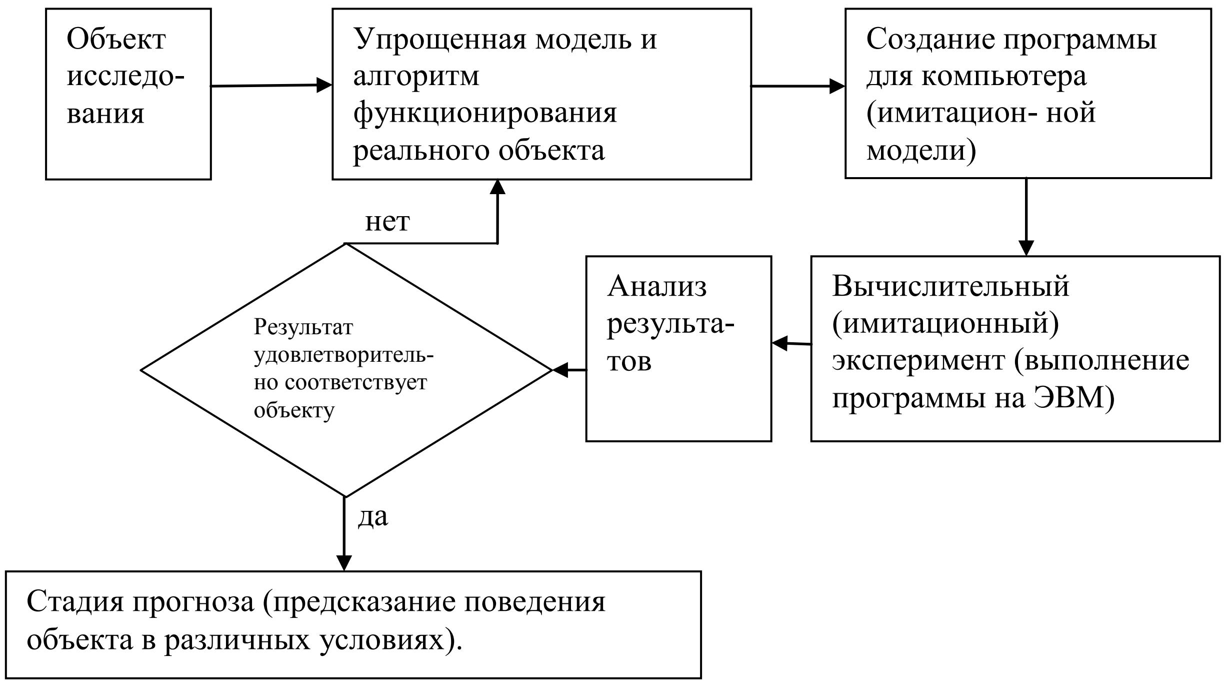


Рис. 4.1. Схема имитационного исследования

Основные этапы имитационного моделирования

Процесс построения имитационной модели во многом носит неформальный характер, алгоритмизировать его достаточно сложно.

Процесс имитационного моделирования является итеративным. При этом системный анализ чередуется и сочетается с составлением модели и алгоритма функционирования реального объекта, вычислительными экспериментами, корректировкой модели и т.п.

Имитационное моделирование состоит из следующих взаимозависимых и, зачастую, пересекающихся во времени этапов:

1. Постановка проблемы, формулировка цели моделирования;

2. Системный анализ моделируемого объекта; построение концептуальной модели;
3. Составление модели и алгоритма функционирования реального объекта; при необходимости− структуризация модели;
4. Программная реализация (написание программы для компьютера) и вычислительные эксперименты;
5. Анализ результатов экспериментов и коррекция модели, алгоритма, программы;
6. Проведение имитационных экспериментов в целях решения поставленной проблемы.

Охарактеризуем каждый из них.

1. Имитационное моделирование, как и всякое другое, должно начинаться с постановки проблемы, с понимания целей моделирования. Желательно перечислить те вопросы, ответы на которые должны быть получены в ходе вычислительных экспериментов.

Необходимость построения имитационной модели, как правило, возникает в процессе прикладных исследований конкретного объекта. Поэтому одной из важнейших особенностей этого этапа моделирования является непременное участие в разработках человека (или группы людей), который будет использовать будущую модель для решения поставленной проблемы, т.е. ЛПР. Нужно отметить, что в реальной жизни цели редко бывают четко очерченными. Это существенно затрудняет точную постановку проблемы. Поэтому упомянутый выше список вопросов следует обязательно согласовать с ЛПР. С ним же следует решать вопрос о масштабах задачи: о ее объеме, границах в пространстве, интервала моделируемого отрезка времени (долгосрочные или краткосрочные прогнозы интересуют ЛПР) и т.д.

2. С учетом целей моделирования следует выявить существенные особенности изучаемого объекта, осуществить отбор всевозможных необходимых сведений о нем и построить так называемую концептуальную модель объекта. Эта модель представляет собой неформализованное описание объекта (словесное описание, представление в виде схем, диаграмм и т.п.), являющееся основой для создания имитационной модели. Чем четче построена концептуальная модель, тем более вероятно создание хорошей имитационной модели.

Построение концептуальной модели предполагает:

a) выявление основных процессов, которые должны быть учтены при моделировании;
b) выделение тех основных характеристик объекта, которые необходимо иметь для решения исходной задачи;
c) определение множества переменных и параметров, которые влияют на динамику этих характеристик;
d) определение множества входных и выходных данных модели;
e) установление границ и законов взаимодействия объекта с окружающей средой;

f) разработку причинно следственных связей, временных отношений и гипотез, согласно которым осуществляется взаимоувязка всех перечисленных компонент в единую систему, имитационную модель.

3. Далее осуществляется переход от качественных зависимостей концептуальной модели к точному алгоритмическому описанию. Исходным пунктом при этом является задания вектора состояний модели, компоненты которого– это те характеристики изучаемого объекта, которые выделены при построении концептуальной модели как базовые, несущие в себе необходимую информацию для решения поставленной проблемы:

$$X^t = (X_1^t, X_2^t, \dots, X_n^t).$$

Вектор X^t, называемый *фазовым вектором*, описывает состояние объекта в момент времени t.

Затем учитывается воздействие на объект внешних факторов, обозначаемых ξ^t. ξ^t– k-мерный вектор $(k \le n)$, который может быть случайным.

Вектор-функция U^t размерности $m \le n$ описывает управляющие воздействия на объект. Эта вектор-функция находится в распоряжении ЛПР. В зависимости от складывающейся ситуации ЛПР выбирает управляющие воздействия на объект с целью направить поведение объекта в нужном для ЛПР направлении.

В результате состояние системы описывается вектором

$$X^t = (X_1^t, X_2^t, \dots, X_n^t, \xi^t, U^t).$$

После, а иногда и параллельно с заданием вектора состояния системы производится декомпозиция модели и выявление ее блочной структуры (если в этом есть необходимость). Чаще всего один блок описывает отдельный процесс, какую-либо подсистему или элемент исходного объекта. В результате декомпозиции модель представляется в виде комплекса взаимосвязанных подмоделей– блоков, которые взаимодействуют по определенным законам и в итоге позволяют провести имитационное исследование объекта.

Блочный принцип построения модели имеет целый ряд преимуществ, особенно ощутимых при создании сложных имитационных моделей. Прежде всего, если модель достаточно сложна, требует значительного объема памяти и машинного времени, то возможности использования такой модели для имитационного эксперимента оказываются весьма ограниченными. Единственным выходом в этом случае оказывается декомпозиция модели с тем, чтобы программы, реализующие отдельные блоки, работали последовательно и обменивались информацией по тем или иным правилам.

Однако блочная структура сложных имитационных моделей обусловлена не только возможностями вычислительной техники. Создание их требует регулярных контактов разработчика со специалистами по смежным специальностям, имеющим отношения к денной проблеме. Блочная структура позволяет полнее использовать знания специалистов в узких областях, не затрудняя взаимопонимание локальными проблемами, не имеющими отношения к данному блоку. Кроме того, именно блочный принцип дает возможность при построении модели устанавливать необходимые пропорции между подробностью моделирования и обеспеченностью информацией. Дело в том, что при исследовании объекта часто приходится сталкиваться с тем, что отдельные аспекты оказываются менее изученными по сравнению с остальными. Им соответствует меньшее количество информации. В этом случае не имеет смысла особенно скрупулезно разрабатывать «благополучный» блок, т.е. тот, в отношении которого мы располагаем обширными и точными знаниями. Ведь при работе объединенной модели точность и подробность результатов такого бока нивелируется за счет расчетов в «неблагополучном» блоке. Поэтому всегда следует соблюдать принцип равноточности боков в имитационной модели.

Поскольку блоки описывают различные подсистемы, то системное время для каждого из них может быть различным. Оно определяется, исходя из внутренних потребностей блока. Если в блоке описываются быстрые изменения, то время исчисляется не большими единицами измерения: секундами, часами, сутками; если изменения объекта происходят достаточно медленно (рост дерева, старение оборудования, глобальные изменения климата), то время исчисляется месяцами, годами, десятилетиями. Однако все результаты по отдельным блокам должны быть в итоге сведены к единому системному времени, принятому для модели в целом.

После завершения декомпозиции модели следует приступить к разработке отдельных ее блоков. Для каждого из них:

a) уточняются и конкретизируются те гипотезы, которые непосредственно относятся к процессам и элементам данного блока;
b) определяется соответствующее подмножество входных и выходных данных,
c) формализуются основные законы взаимодействия элементов блока. Т.е. происходит переход от качественных зависимостей концептуальной модели к точным количественным зависимостям (например, уравнениям) и логическим схемам взаимодействия элементов.

4. Заключительным в построении модели является объединение блоков в имитационную модель на базе стандартного или специально созданного математического обеспечения. Здесь особую роль играет выбор языка программирования: либо это будет универсальный язык типа Си, Java и т.п., либо специализированные языки имитационного моделирования типа GPSS, Симула и.т.д. Специализированные языки удобны для программирования и отличаются концептуальной направленностью. Однако такие языки требуют специальных трансляторов, которые не всегда входят в стандартное математическое обеспечение ЭВМ. В свою очередь, использование

универсальных языков, как правило, сильно увеличивает объем программы, делает код программы громоздким, хотя и допускает проведение имитационных экспериментов практически на любых компьютерах, что расширяет область практического применения модели.

Составлению компьютерных программ, реализующих всю модель в целом, предшествуют испытания и отработка различных схем взаимодействия блоков. Здесь удобно бывает рассматривать имитационную модель как совокупность автоматов с памятью и без нее, детерминированных или стохастических, а работу модели− как изучение с помощью компьютера совместного поведения автоматов в случайной или детерминированной среде.

5. Маловероятно, что построенная имитационная модель сразу же окажется удачной. Скорее всего в ней обнаружатся ошибки. Для отыскания этих ошибок используются вычислительные эксперименты как с отдельными блоками, так и с системой в целом. Причем проводятся такие эксперименты, результат которых может быть предсказан заранее. Тогда отклонение результата от прогнозируемого говорит о том, что содержатся ошибки либо в модели, либо в алгоритме, либо в программе, либо сразу в нескольких местах.

Устранение ошибок продолжается до тех пор, пока результаты экспериментов придут в соответствие с прогнозируемыми. После этого можно считать, что имитационная модель создана.

Следует отметить, что зачастую отладка программ занимает не меньше времени, чем ее первоначальное написание.

Идентификация и верификация имитационной модели

Итак, имитационная модель построена. Однако, прежде чем ее использовать, необходимо решить следующие задачи:

1. выбрать числовые значения неопределенных пока числовых параметров (идентификация);
2. убедиться, что при этих значениях параметров модель хорошо соответствует моделируемому объекту, адекватна ему (верификация, проверка адекватности).

Рассмотрим наиболее распространенную задачу идентификации. Пусть $\bar{y} = \overline{y}(t, \overline{\propto})$ – вектор выходных характеристик имитационной модели, зависящий от набора параметров $\overline{\propto} = \{\propto_1, \propto_2, \dots, \propto_m\}$. Предположим, что имеется ряд известных векторов $\overline{y}(t), t \in [t_1, t_2]$, представляющих собой результаты натурных наблюдений за моделируемым объектом. Разобьем промежуток $[t_1, t_2]$ на две части: $[t_1, \tau]$ – обучающий промежуток и $[\tau, t_2]$ – экзаменующий промежуток.

Задача идентификации состоит в том, чтобы найти такой набор параметров $\overline{\propto}$, который доставляет минимум функционалу:

$$\sum_{t=t_1}^{\tau}(\overline{y}(t)-\overline{y}(t,\overline{\alpha}))^2.$$

Для решения этой задачи могут быть использованы известные численные методы нахождения экстремума функции нескольких переменных.

Обозначим найденный в результате решения этой задачи наилучший набор параметров через $\overline{\propto^*}$. Используя в модели именно эти параметры, получим выходные характеристики в виде $\overline{y}=\overline{y}\left(t,\overline{\propto^*}\right)$. На этом решение задачи идентификации закончено.

Перейдем к задаче верификации, которая не столь формализована. Прежде всего убедимся, что на экзаменующем промежутке времени $[\tau,t_2]$ расчетная траектория $\overline{y}=\overline{y}\left(t,\overline{\propto^*}\right)$ близка к фактической $\overline{y}(t)$.

Сравнение этих двух траекторий позволяет судить об адекватности модели. Существуют специальные методы оценки близости траекторий. Один из них состоит в вычислении коэффициента несовпадения:

$$U=\frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\overline{y}(t)-\overline{y}\left(t,\overline{\propto^*}\right)\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\overline{y}^2(t)}+\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\overline{y}\left(t,\overline{\propto^*}\right)\right)^2}},\quad 0\le U\le 1.$$

Чем ближе U к нулю, тем ближе модельная траектория к фактической. В случае, когда U равен единице, модель не является адекватной и требует либо перестройки структуры, замены или уточнения гипотез, либо идентификации по более полным и достоверным данным.

Однако даже если близость траекторий имеет место, еще нет гарантии того, что модель адекватна реальному объекту. Естественно потребовать, чтобы выполнялись следующие условия:

1. машинная реализация (код программы для ЭВМ) соответствует формальной модели;
2. динамика модели соответствует динамике реального объекта;
3. результаты моделирования правильно интерпретируются.

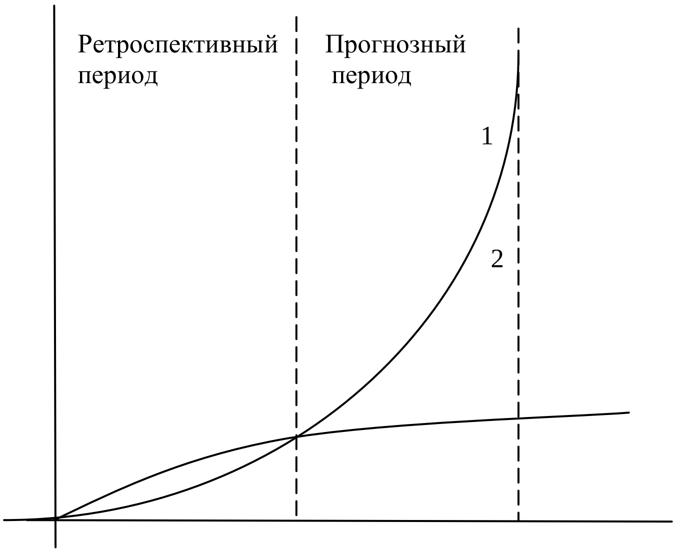
Проверка адекватности в этом смысле проводится на основе экспертного анализа и статистических методов.

Под соответствием машинной (компьютерной) и формальной моделей понимается, во-первых, идентичность их алгоритмических структур (сохраняется ли логика построения модели при машинной реализации?) и, во-вторых, совпадения областей варьирования компонент вектора состояния формальной и машинной моделей. Второе требование означает, что численные методы, которые используются для реализации модели на ЭВМ, не должны давать такую погрешность, которая выводит некоторую компоненту из области допустимых значений.

Известно, что практически любую имитационную модель можно отладить до нужной степени совпадения моделируемой и фактической траекторий.

Но при первоначальном моделировании может возникнуть ситуация, приведенная на Рис. 4.2. Модель, которая хорошо аппроксимирует фактическую траекторию на ретроспективном периоде, но дает абсолютно неверный прогноз, т.е. неадекватно отражает поведение реального объекта.

Следовательно, для того чтобы оценить адекватность модели, необходимо качественно исследовать ее динамические свойства путем проведения серии тестовых расчетов.



1− Фактическая траектория. 2− Расчетная траектория.

Рис. 4.2. Неадекватная модель

Можно выделить два основных вида тестов.

К первому относятся тесты, в которых задаются правдоподобные значения входов и управляющих воздействий. Если в этом случае расчеты по модели не противоречат известным законам реального поведения объекта (не противоречат нашим представлениям об объекте), то это говорит в пользу адекватности модели. В противном случае необходимо найти причины несогласованности и перестроить модель.

Второй тип тестов основан на использовании критических ситуаций, т.е. данных, которые не характерны для исследуемого объекта, но тем не менее могут иметь место. Эти тесты особенно важны для сложных моделей, предназначенных для долгосрочного прогнозирования, поскольку чем лучше модель описывает поведение объекта в критических ситуациях, тем более можно быть уверенным в правильности расчетов для нормальных условий.

Адекватность модели характеризуется также ее чувствительностью по отношению к изменению параметров и начальных значений вектора состояния. Если результаты исследований с помощью модели существенно изменяются при малых возмущениях параметров и незначительных отклонениях от начальных данных, другими словами, если модель является не устойчивой, то такую модель нельзя считать адекватной (при условии, конечно, что исследуемый объект обладает устойчивостью в этом смысле). Требование устойчивости тем более важно, чем менее точно могут быть определены параметры модели.

Основными методами оценки адекватности и пригодности модели являются неформальные методы (методы экспертного оценивания, т.е. учитывается мнение экспертов в соответствующих областях). Однако в некоторых случаях удается формализовать эту процедуру. В частности, весьма полезными оказываются различные статистические методы, спектральный анализ и др.

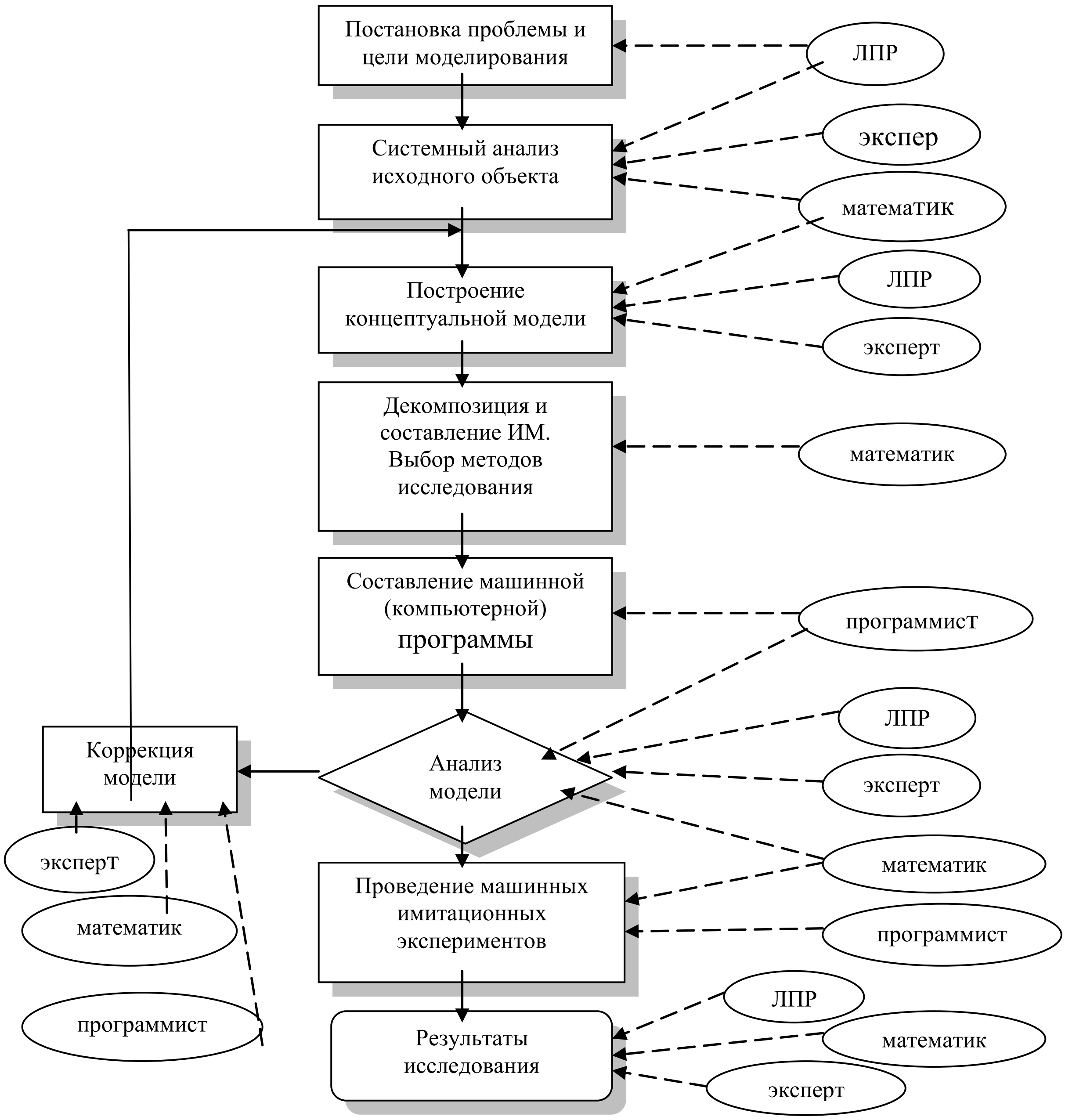


Рис. 4.3. Общая схема имитационного исследования

Итак, модель идентифицирована, верифицирована, проверена на адекватность. Если все этапы выполнены успешно, то она является готовым инструментом исследования поставленной проблемы и можно переходить к главному этапу имитационного моделирования – проведению имитационного эксперимента, которое сопровождается, с одной стороны, планированием эксперимента, а с другой – обработкой результатов эксперимента.

Планирование имитационных экспериментов представляет собой отдельную и довольно сложную задачу, о которой речь пойдет дальше.

Эксперименты делятся по цели исследования на два основных типа: дескриптивные (описательные) и оптимизационные. Эксперименты первого типа проводятся в целях исследования объекта. Другой тип включает эксперименты, направленные на выявление наилучших стратегий управления исходным объектом.

При рассмотрении этапов имитационного моделирования неоднократно подчеркивалась роль человека в имитационных исследованиях. Очевидно, что при решении крупных народнохозяйственных проблем задача имитационного моделирования становится непосильной для одного человека и к имитационным исследованиям привлекаются целые творческие коллективы, состоящие из людей различных профессий, причем место и функции каждого специалиста четко определены. В заключении этого параграфа приведем общую схему проведения имитационного исследования (рис. 4.3) с указанием типов специалистов, участвующих в разных его стадиях. Здесь эксперт− специалист в своей предметной области, ЛПР− лицо, принимающее решение.

Планирование имитационных экспериментов

Будем считать, что идентификация, верификация и все необходимые коррекции модели завершены. Теперь можно приступать к проведению имитационного эксперимента− главного этапа имитационного моделирования, включающего, кроме расчетов по модели, планирование эксперимента и обработку результатов эксперимента. Сразу же возникает вопрос о том, при каких внешних воздействиях проводить расчеты, сколько расчетов проводить для того, чтобы уверенным в верности полученного решения и т.п. все эти проблемы могут быть решены в процессе планирования эксперимента. В основном методы планирования экспериментов были разработаны для натурных экспериментов. Имитационный эксперимент сохранил черты натурного эксперимента. Основная разница между ними состоит в том, что эксперимент здесь проводится не с объектом, а с его аналогом− моделью. Поэтому многие результаты теории планирования натурных экспериментов могут быть перенесены в область имитационных исследований. Рассмотрим некоторые аспекты планирования экспериментов.

Итак, цель эксперимента− установить связь между воздействиями на модель и ее откликом на это воздействие, причем то, как формируется это отклик, при проведении имитационных экспериментов уже не интересует исследователя. Поэтому на этом этапе модель можно представить в виде

$y = f(x),$
где $x -$ воздействие на модель или фактор; $y -$ результат воздействия или реакция; $f -$ поверхность реакции.

В общем случае x и y есть вектор-функции, зависящие от времени. Любой имитационный эксперимент в этом случае может быть направлен либо на исследование поверхности реакции (например, задачи прогнозирования), либо на поиск максимума или минимума поверхности реакции в некотором пространстве факторов (задачи оптимального управления объектом и т.п.).

Факторы могут быть либо количественными, либо качественными. Здесь мы остановимся на факторах первого типа

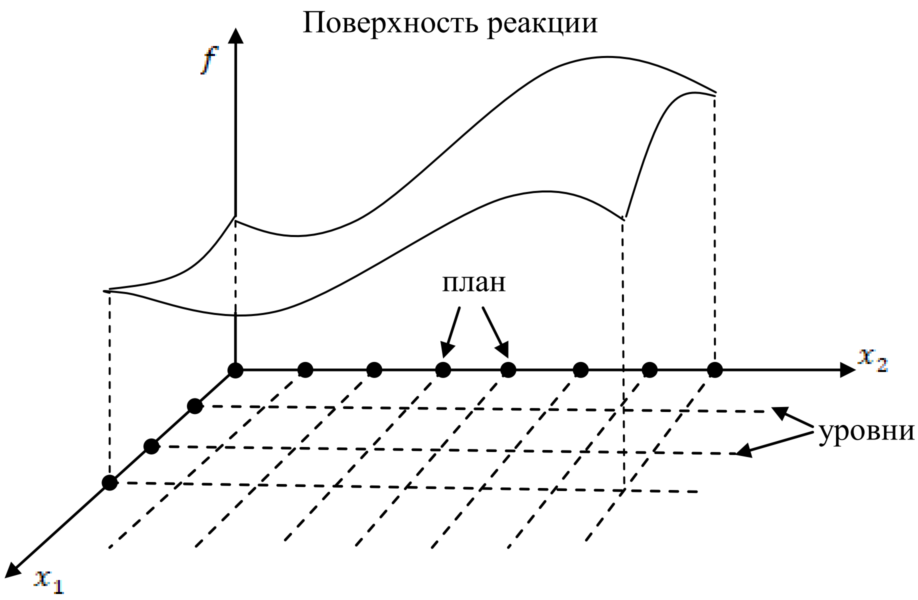


Рис. 4.4. Факторный план и поверхность реакции

Пусть, для простоты, поверхность реакции описывается функцией, зависящей от двух количественных факторов:

$f = f(x_1,x_2),$
причем x_1 и x_2 могут принимать дискретные фиксированные значения из областей X_1 и X_2. Значение фактора назовем уровнем, а совокупность всех возможных пар (x_1, x_2)– полным факторным планом (рис. 4.4).

Таким образом каждая точка плана определяет одно значение поверхности реакции. Чем больше точек плана будет рассмотрено, т.е. чем полнее построен план, тем точнее представления о поверхности реакции. Однако, несмотря на это, применение полных факторных планов ограничено и возможно лишь в случае незначительного числа факторов и их уровней.

Более часто используются неполные факторные планы, требующие меньшего числа точек плана и не приводящие при этом к ощутимым потерям информации о поверхности реакции. Здесь в основном исследуются несколько главных факторов, а неполные факторные планы применяются для «отсеивания» несущественных факторов. Процедура отсеивания состоит в последовательном построении неполных планов.

Снова вернемся к примеру двухфакторной модели

$$f = f(x_1, x_2),$$

где факторы x_1 и x_2 имеют более трех уровней каждый. Сначала имитационный эксперимент проводится для начальных значений этих факторов (x_1^0, x_2^0), а затем задается некоторое фиксированное изменение каждого уровня: δ_1 и δ_2, где

$$\delta_1 = \frac{x_1}{n_1},$$

$$\delta_2 = \frac{x_2}{n_2},$$

n_1, n_2 – число уровней x_1 и x_2 соответственно, и строятся неполные факторные планы.

Пусть $\overline{x}_1, \overline{x}_2$– средние уровни факторов x_1 и x_2. Находим четыре точки поверхности реакции

$$f_{+\delta_1}^1 = f(\overline{x}_1 + \delta_1, \overline{x}_2), \qquad f_{-\delta_1}^1 = f(\overline{x}_1 - \delta_1, \overline{x}_2),$$

$$f_{+\delta_2}^2 = f(\overline{x}_1, \overline{x}_2 + \delta_2), \qquad f_{-\delta_2}^2 = f(\overline{x}_1, \overline{x}_2 - \delta_2),$$

по которым можно судить о степени влияния каждого фактора. Так, если

$$\left|f_{+\delta_1}^1 - f_{-\delta_1}^1\right| > \left|f_{+\delta_2}^2 - f_{-\delta_2}^2\right|,$$

то x_1 является более существенным фактором, и наоборот.

В общем случае вид функции

$$f = f(x_1, x_2)$$

является громоздким и сложным в вычислительном аспекте. Поэтому еще одной типичной задачей планирования вычислительного эксперимента является аппроксимация поверхности реакции f некоторой более простой функцией φ, зависящей от тех же факторов. Как правило удается построить линейную зависимость

$$\varphi_n = \sum_{i=1}^{n} \alpha_i x_i + \alpha_0,$$

где α_i– коэффициенты линейной функции.

Пусть, например, $n = 2$. Тогда для построения линейной функции

$$\varphi_2 = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_0$$

требуется проведение имитационного эксперимента по полному двухфакторному плану.

В случае неудовлетворительной аппроксимации есть возможность строить полиномы более высокой степени. Например, полином второй степени

$$\varphi = \beta_{11}x_1^2 + \beta_{12}x_1x_2 + \beta_{22}x_2^2 + \alpha_1x_1 + \alpha_2x_2 + \alpha_0.$$

Его коэффициенты находятся за счет дополнительных экспериментов.

Процесс продолжается до тех пор, пока аппроксимация не даст удовлетворительных результатов.

Более узкой проблемой является проблема поиска экстремумов поверхности реакции, для решения которой используются известные методы оптимизации на заданном множестве значений факторов. Среди них отметим здесь метод наискорейшего спуска.

Если поверхность реакции имеет несколько локальных минимумов, то метод наискорейшего спуска не гарантирует нахождения наименьшего значения. В этом случае нужно несколько раз применять этот метод, отправляясь всякий раз от различных, отличающихся начальных условий.

Заметим, что линейная аппроксимация вблизи точки оптимума оказывается не эффективной. В окрестности очки оптимума, где поверхность реакции почти стационарна, используется аппроксимация более высокого порядка– хотя бы квадратичным полиномом

Все перечисленные методы планирования эксперимента относятся к детерминированным моделям. Для стохастических моделей однократная реализация построенного плана не позволяет получить желаемую информацию об изучаемой поверхности реакции. В этом случае необходимо несколько раз реализовывать один и тот же план с различными начальными состояниями, моделируемыми при помощи генератора случайных чисел (каждая реализация здесь называется репликой). Определение объема выборки (количества реплик) в имитационном моделировании представляет собой очень трудную, но важную задачу. С одной стороны увеличение объема ведет к увеличению затрат машинного времени и, тем самым, денежных средств. С другой стороны, чем больше количество реплик, тем более достоверна информация, полученная с помощью модели, и, тем самым, меньше возможные потери, обусловленные использованием недостоверной информации. Минимизация суммарных потеть всякий раз осуществляется с помощью методов статистического анализа и методов оптимизации.

Литература:

1. Ашманов С.А. Линейное программирование. – М.: Наука, 1981.
2. Владимирский Б.М., Горстко А.Б., Ерусалимский Я.М. Математика. Общий курс. – СПб.: Издательство «Лань», 2009.
3. Геймер Ю.Б. Введение в исследование операций. – М.: Наука, 1971.
4. Кравченко Л.В., Емелин А.А. Математическое моделирование. Учебное пособие. – Зерноград.: ФГОУ ВПО АЧГАА, 2005.
5. Мулен Э. Теория игр с примерами из математической экономики. – М.: Мир,1985.