М.А.Тынкевич

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ (ИССЛЕДОВАНИЕ ОПЕРАЦИЙ)

Учебное пособие

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«КУЗБАССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

М. А. ТЫНКЕВИЧ

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ (ИССЛЕДОВАНИЕ ОПЕРАЦИЙ)

Учебное пособие
3-е издание, исправленное и дополненное

Рекомендовано Сибирским региональным учебно-методическим центром высшего профессионального образования в качестве учебного пособия для студентов инженерно-экономических специальностей и направлений

Кемерово 2011

УДК 519.6
ББК 32.965

Данное пособие разработано на базе курса лекций по экономико-математическим моделям планирования и управления, читавшегося автором в течение многих лет как для студентов инженерно-экономических специальностей, так и специальности «Прикладная информатика».

Математическая строгость изложения в пределах возможностей студентов-нематематиков сочетается с экономической постановкой задач и алгоритмическим описанием методов. Дается обзор основных методов линейного, нелинейного и динамического программирования, теории игр, сетевого планирования и массового обслуживания.

Новое издание уделяет большое внимание задачам нелинейного программирования, расширяет возможности для самостоятельного изучения читателем отдельных разделов. Может быть полезно студентам различных специальностей при изучении методов исследования операций.

Тынкевич М. А.
Т93 Экономико-математические методы (Исследование операций): Учеб. пособие. 3-е изд., испр. и доп. / Кузбасс. гос. техн. ун-т. – Кемерово, 2011. – 222 с.
ISBN 948–5–89070–785–7

Рецензенты:
Зав. отделом экономической информатики Института экономики и ОПП СО РАН, профессор кафедры применения математических методов в экономике и планировании экономического факультета НГУ, канд. экон. наук Ю. Ш. Блам

Кафедра вычислительной математики и компьютерного моделирования механико-математического факультета Томского государственного университета (зав. кафедрой, профессор, д-р физ.-мат. наук А. В. Старченко)

Печатается по решению редакционно-издательского Совета Кузбасского государственного технического университета.

ISBN 948-5-89070-785-7

ПРЕДИСЛОВИЕ

Еще полвека назад толковый организатор крупного производства, окружив себя десятком единомышленников, был в состоянии оперативно оценивать быстро меняющуюся обстановку и принимать разумные решения. XX век с его небывалыми темпами научно-технического прогресса и усложнением хозяйственных связей породил производственные системы исключительной сложности, что, в свою очередь, потребовало создания адекватных систем управления, которые не могут уже строиться только на основании интуиции и жизненного опыта. Сегодня в цивилизованном обществе не возникает сомнений в необходимости использования для управления сложными системами методов математического моделирования, которое подчас дает решения, неожиданные не только для начинающего, но и для опытного руководителя. Сегодня традиционные лозунги, требующие повышения производительности труда или всеобщего благоденствия, остаются демагогией без конкретных решений в сфере планирования и управления, поверенных численными оценками.

Предлагаемое учебное пособие предназначено для студентов специальности «Прикладная информатика в экономике», хотя может оказаться полезным для студентов экономических специальностей и лиц с интересами в области прикладной математики.

В основе пособия лежит курс лекций, который автор читал с 1962 г. для специальности «Вычислительная математика» в Томском университете и позднее для экономических специальностей в КузГТУ.

Здесь автор пытается дать читателю, не обладающему фундаментальной математической подготовкой, представление об основных направлениях в области исследования операций, математического моделирования и методах решения соответствующих задач, хотя не всегда удается сочетать доступность изложения с математической доказательностью. Приведенные алгоритмы сопровождаются примерами их реализации и без особого труда могут быть преобразованы в компьютерные программы.

Основное внимание мы традиционно уделяем методам математического программирования, хотя затрагиваем и смежную тематику.

Автор отдает должное истории формирования исследования операций как научной дисциплины и людям, которые его создавали. По этим соображениям мы в список рекомендуемой литературы включили, наряду с современными руководствами, и первоисточники, которые могут быть полезны для читателя, заинтересованного в знакомстве с доказательной базой и примерами возможных приложений.

1. ВВЕДЕНИЕ В ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

1.1. Исследование операций и математическое моделирование

Термин «исследование операций» (*operation research*) возник в годы второй мировой войны как символ научного подхода к решению задач управления, в частности как «метод быстрого расчета программы поэтапного развертывания, подготовки и тылового снабжения» [1]. Сегодня исследование операций можно было бы определить как *совокупность методов поиска наилучших решений многообразия задач организационного управления при наличии тех или иных ограничений*.

Само по себе математическое моделирование возникло, если не с момента возникновения понятия о числе как некоторой абстракции, то с момента возникновения алгебры. Знакомясь с физикой, мы встречаемся с законами Ома или Бойля – Мариотта, которые выступают как математическое описание взаимосвязи между физическими величинами. Иоганн Кеплер уже в 1615 году в статье «Новая стереометрия винных бочек» построил математическую модель экономически выгодного соотношения между геометрическими характеристиками упомянутой тары.

Математическую модель можно определить как приближенное описание какого-то явления в природе, техническом устройстве или человеке с помощью математической символики.

Непосредственное математическое моделирование начинается с попытки выяснить факторы, которые как-то характеризуют изучаемое явление (параметры явления) и допускают возможность количественной оценки. Несомненно, что затем мы должны сформулировать ограничения, отражающие объективно существующие связи между этими факторами, представить их в виде математических соотношений (уравнений или неравенств). При этом мы добавляем в систему ограничений и свои требования к факторам, которыми в состоянии управлять.

Наконец, выбрав множество факторов (переменных величин), выявив соотношения между ними и ограничения на диапазон их значений, формулируем цель решаемой задачи – подбор среди всех допустимых значений переменных такого сочетания, при котором достигалось бы наилучшее (оптимальное) значение конкретной функции от этих переменных – баланс между подчас противоречивыми характеристиками исследуемого явления. Существенно, что цель должна быть единственной («за двумя зайцами не следует гнаться») и, например, популярные в недавнем прошлом лозунги «*максимального* удовлетворения жизненных потребностей трудящихся при *минимальных* затратах» сами по себе звучат нелепо из-за противоположности объявленных целей.

Построенная модель явления представляет ценность, если она соответствует критерию практики – наблюдается согласие между теоретическими оценками и результатами наблюдений. Замечательно, если модель при минимуме учитываемых факторов достаточно хорошо соответствует жизни. В большинстве же случаев для обеспечения близости теоретических оценок к реальности учитывают много значимых реальных факторов (особенно тех, которыми можно управлять). Но рост сложности модели требует более сложного математического аппарата ее анализа и ведет к известному «проклятию размерности», делающему выбор оптимальной политики практически неразрешимой.

Очевидно, что никакая математическая модель не может описать действительность с исчерпывающей полнотой (как утверждал известный Козьма Прутков, «никто не может объять необъятное»). Потому, согласно Р. Беллману [19], «Ученый должен идти прямой и узкой тропой между Западнями Переупрощения и Болотом Переусложнения».

Определенные сложности вносит в создаваемую модель и интерпретацию получаемых оценок учет факторов т. н. стохастической (вероятностной) природы, поскольку 100%-ная предсказуемость получаемых оценок здесь недостижима.

В области исследования физико-химических и даже некоторых биологических процессов адекватность математической модели можно проверить постановкой эксперимента. В сфере, где значимую роль играет человеческий фактор, возникает множество проблем, с величайшим трудом поддающихся формализации из-за отсутствия однозначных воззрений на природу явлений. Ограничены и возможности эксперимента – так перенос действенных методов рыночной экономики из Чикаго на российскую почву дорого обошелся ее обитателям.

Сказанное вовсе не означает невозможность математического моделирования в экономике или социологии. Многие задачи в этой сфере превосходно решаются т. н. экономико-математическими методами, решение других ограничивает диапазон рисков при принятии решений – дает стартовый материал для волевых действий человека в процессе управления, определяемых его опытом и интуицией.

Исследование операций в зависимости от типа исследуемой математической модели и методов ее анализа обычно подразделяют на относительно самостоятельные дисциплины, такие как математическое программирование, теория игр и статистических решений, сетевое планирование, теория расписаний, теория массового обслуживания, теория графов, вариационное исчисление и другие. Эта классификация достаточно условна, но общее для этих дисциплин – поиск наилучших решений на основе методов прикладной математики.

1.2. Они стояли у истоков исследования операций

В определенной мере создание математического программирования и его прикладной аспект связаны с выходом в 1939 г. работы выдающегося советского математика Леонида Витальевича Канторовича (1912−1986) «Математические методы в организации и планировании производства», где впервые была поставлена и решена задача линейного программирования. Но, как писал впоследствии Д. Данциг [1], «Л. Канторович добился многообещающих результатов, которыми в СССР пренебрегли». Его, наряду с единомышленниками (Г. Ш. Рубинштейн, В.А. Залгаллер, В. Л. Булавский, Д. Б. Юдин, Е. Г. Гольштейн и др.), обвиняли в стремлении «протащить буржуазные идеи в советскую науку» и лишь в 1959 г. Канторович стал известен на Западе благодаря публикации его книги «Экономический расчет наилучшего использования ресурсов», написанной еще в 1942 г.

Тайлинг К. Купманс, Джордж Б. Данциг, Леонид В. Канторович

В годы войны многие американские математики и экономисты были привлечены к поискам «оптимальных» решений в области экономического планирования, и к концу сороковых их интенсивный труд увенчался триумфальными итогами. Так в 1947 г. американским математиком Джорджем Б. Данцигом (1914–2005) была выдвинута идея и предложен эффективный алгоритм *симплексного метода* для решения задач *линейного программирования* (терминология Данцига). Идея по достоинству была оценена Т. К. Купмансом [1] (1910–1985), в годы войны работавшим над транспортной моделью для снабжения войск на

[1] Т. К. Купманс, совместно с Л. В. Канторовичем, был удостоен в 1975 г. Нобелевской премии по экономике.

театре военных действий и сразу разглядевшим возможность применения для общего экономического планирования.

50-е годы знаменовались фундаментальными результатами в области математического программирования и его приложений в экономике, ракетостроении и других сферах. В современных учебниках постоянно фигурируют имена талантливых математиков и экономистов того времени (Х. Хотеллинг, Г. Кун, А. Таккер, К. Эрроу, Л. Гурвиц, Р. Гомори, Л.Форд и Д. Фалкерсон, С. Гасс, Т. Саати, Г. Вагнер и др.).

Нельзя здесь не упомянуть имя Джона фон Неймана (1903–1957), одного из основателей кибернетики, одного из создателей атомной бомбы (математически доказал осуществимость взрывного способа детонации атомной бомбы) и основоположника теории автоматов (выдвинул концепцию хранения команд компьютера в его внутренней памяти), что послужило огромным толчком к развитию вычислительной техники). В 1927 г. он сформулировал известную теорему о минимаксе – основополагающий элемент теории игр, а в 1944 г. вместе с Оскаром Моргенштерном подготовил к печати монографию «Теория игр и экономическое поведение» – фундамент для дальнейшего развития теории игр и статистических решений. Ему же принадлежит формулировка фундаментальной для математического программирования теоремы о двойственности (первое строгое ее доказательство опубликовано впоследствии Таккером, Куном и Джейлом).

Джон фон-Нейман

В 1957 г. появляется монография выдающегося американского математика Ричарда. Беллмана [19], положившая начало одному из оригинальных методов исследования многошаговых процессов принятия решений – методу динамического программирования. Большой вклад в развитие методов оптимального управления динамическими системами внесла группа советских математиков во главе с Л. С. Понтрягиным.

Разумеется, отдельные результаты в области экономико-математического моделирования были получены и в предшествующие годы.

Так в 1932 г. появилась знаменитая работа Василия Леонтьева, предложившего матричную структуру, которую он назвал «*межотраслевой моделью*» американской экономики типа *«вход–выход»*. В дальнейшем она подверглась обобщению на многоальтернативность процессов производства продукта, приобрела фактор динамики и была использована в практике планирования многих стран.

Истоки современной эконометрики, тесно связанной с исследованием операций, и теории массового обслуживания можно искать уже в трудах Якоба Бернулли (1654–1705) и Иоганна Бернулли (1744–1807) по теории вероятностей.

Исключительный вклад в эту область внесли представители российской и американской науки (П. Л. Чебышев, А. А. Марков, А. Я. Хинчин, С. Н. Бернштейн, А. Н. Колмогоров, Ежи Нейман, А. Эрланг, У. Феллер, У. С. Госсет (Стьюдент), Э. Пирсон и др.).

Бесспорный вклад в методы исследования операций (даже не задумываясь о приложениях в экономике) внесли творцы комбинаторной математики (Фибоначчи, Л. Эйлер, П. Ферма и др.) и создатели многообразия методов аналитических и численных методов оптимизации.

Создание ЭВМ оказалось немаловажным стимулом развития численных методов оптимального планирования и управления, ставших обычным инструментом исследования в различных областях человеческой деятельности.

Резко увеличив вычислительные возможности человека, ЭВМ стимулировали появление методов стохастического моделирования (методов Монте-Карло), «подающих надежду на спасение» при решении задач большой размерности (при ручном счете их реализация абсолютно нереальна).

Сегодня, благодаря богатейшему программному обеспечению и дружественному интерфейсу, даже дилетант за минуты в состоянии решать многие задачи, над которыми бились тысячи исследователей в недавнем прошлом.

1.3. Математическое программирование и «проклятие размерности»

Не затрагивая пока других методов исследования операций, остановимся на задачах математического программирования.

Вообразите себя одним из руководителей крупного производства, показатели работы которого вас не устраивают. Вы хотели бы осуществить мероприятия по увеличению объема выпуска продукции, повышению качества, снижению себестоимости, обеспечению ритмичности и многие другие.

Достижение всех этих целей одновременно – некорректная задача. Нужна некоторая глобальная цель, гармонично сочетающая локальные цели на основе выбранной системы уровней значимости или требующая поддержания всех целей на каком-то предельном уровне.

Очевидно, что достижение цели зависит от множества различных факторов (количества рабочих, уровня их квалификации, фондовооруженности, запасов сырья, спроса на создаваемый продукт и т. п.). Другими словами, цель является некоторой функцией от этих факторов.

Если вам известны (из статистических оценок, опыта или здравого смысла) функциональные связи между целью и факторами, которыми можно управлять, то возникает задача поиска сочетания значений производственных факторов, обеспечивающего оптимум для поставленной цели.

Задача **поиска экстремума (максимума или минимума) некоторой функции при наличии ограничений** на значения ее переменных и составляет **общую задачу математического программирования**.

Представляет ли какие-либо трудности решение такой задачи ?

Пусть стоит задача поиска максимума функции $F(x)$ при простейших условиях $A \le x \le B$ (при каком значении x, удовлетворяющем указанным условиям, $F(x)$ принимает самое большое значение ?).

В предположении дифференцируемости $F(x)$, классическая математика предлагает взять производную от $F(x)$, приравнять нулю, решить полученное уравнение (насколько простым будет процесс решения, знаете ли вы какие-нибудь методы решения уравнений, сколько решений искать ?) и затем, например, выбрать максимальное из значений $F(x)$, вычисленных на концах интервала и в «критических» точках, попавших в этот интервал.

Если этот путь вас не устраивает, а на столе у вас стоит персональный компьютер и вы хотя бы чуть-чуть в состоянии воспользоваться вычислительными его возможностями (догадываетесь о смысле буквосочетаний Pascal, Basic, MatLab или Excel), то при желаемой точности решения в 1 % доли интервала достаточно разбить интервал $[a,b]$ на 100 частей, вычислить значения функции в образовавшихся 101 точках и найти максимальное.

Например, пытаясь найти максимальное значение функции $F(x) = x \cdot e^{-x} \ln(2+x)$ для $x \in [0,5]$, сооружаем и запускаем элементарную программу, подобную приведенной ниже программе в среде MatLab

```
a=0;  b=5;  M=0; h=(b-a)/100;
for i=0 :100  x=a+i*h; f=x*exp(-x)*log(1+x);
              if f>M  M=f; xopt=x end
 end
 xopt
 M
```

Ответ {xopt =1.6500 M =0.3088} появится за доли секунды. Поиск точки максимума вогнутой (минимума выпуклой) функции при точности 1% интервала можно выполнить в 10 раз быстрее, а точность 10^{-6} интервала достичь всего лишь за 30 вычислений $F(x)$, если воспользоваться методом, основанным на числах Фибоначчи [19] .

Напрашивается вывод, что решение задачи в случае одной переменной (конечно, если функция не принадлежит к числу осциллирующих – быстро меняющих значения) не представляет никаких принципиальных затруднений. Тем более элементарен поиск максимума выпуклой (минимума вогнутой) функции на интервале, так как здесь можно ограничиться ее вычислением на концах интервала.

Если обратиться к задаче с двумя переменными, состоящей в максимизации произвольной $F(x, y)$ при условиях $A \le x \le B$, $C \le y \le D$, то попытка воспользоваться аппаратом классической математики будет успешной лишь в случае, когда есть уверенность, что максимум достигается внутри прямоугольника, определяющего множество допустимых точек (граничных точек здесь не две, а множество).

Если использовать численное решение, то при той же точности придется разбить интервалы по x и y на 100 частей и вычислять значения функции в 101×101 точках (время счета значительно больше, хотя и не слишком ощутимо).

В случае N переменных аналогичный подход к решению потребует объем вычислений порядка 101^N. Если на вычисление одного значения тратить тысячную долю микросекунды (10^{-9} с), то при $N = 8$ время счета составит 125 дней, а при $N = 10$ – 3502 года.

Если вы надеетесь, что ниже будет изложен приемлемый простой метод решения таких задач, то позвольте вас разочаровать. Универсального метода решения задач математического программирования нет, почему и приходится изобретать оригинальные методы для отдельных классов задач.

Так для решения задач линейного программирования (задач с линейной целевой функцией и линейными ограничениями) существует универсальный симплексный метод, имеющий множество модификаций и дающий решение задачи небольшой размерности в приемлемое время даже вручную (без использования компьютера). По крайней мере, компьютерное решение подобной задачи при $N = 10$ можно реализовать за секунды.

Ниже мы постараемся дать представление о классах задач, поддающихся решению известными методами, но нет гарантии, что для конкретной задачи вам не придется изобретать оригинальные приемы поиска оптимума.

2. ОСНОВЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Как известно из элементарной алгебры, выражение типа $a_0+a_1x_1+a_2x_2+\ldots+a_nx_n$, где a_i ($i = 1 \div n$) – некоторые константы и x_i – переменные величины, принято называть *линейным*. Так зависимость $y=a+bx$ определяет величину y как линейную функцию от величины x и может быть изображена на плоскости в виде прямой линии (рис. 1). Аналогично в виде прямой на плоскости можно изобразить и линейное уравнение $\alpha x+\beta y=\gamma$ при произвольных α, β, γ (α и β не могут быть одновременно равны нулю). Линейное неравенство $\alpha x+\beta y<\gamma$ (или $\alpha x+\beta y>\gamma$) определяет множество точек (полуплоскость), лежащее по какую-то сторону от соответствующей прямой линии (точки самой прямой исключены).

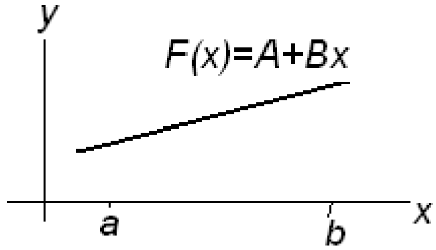


Рис. 1

В случае $n = 3$ геометрическим образом линейного уравнения $\alpha x+\beta y+\gamma z=\delta$ может служить плоскость в трехмерном пространстве, а линейному неравенству $\alpha x+\beta y +\gamma z \le \delta$ соответствуют все точки полупространства, ограниченного соответствующей плоскостью (включительно). По аналогии для $n > 3$ используют термины *гиперплоскость* и *гиперпространство*.

Задача линейного программирования (ЛП) – частный случай задач математического программирования – состоит в отыскании максимума (минимума) линейной функции при наличии линейных ограничений на ее переменные[2].

Если обратиться к поиску максимума (минимума) линейной функции одной переменной $F(X)$ в интервале $[a, b]$, то очевидно, что достаточно найти $F(a)$ и $F(b)$ и выбрать из них большее (меньшее).

2.1. Линейная программа: случай двух переменных

Рассмотрим несколько более сложную задачу.

Плановый прием в некотором вузе не превышает 5000 студентов, причем разрешен прием не более 4000 студентов своей страны и любого количества иностранных студентов. Аудиторный фонд составляет пока 2800 мест. Педагогический персонал вуза – 400 человек.

---

[2] Терминологически правильнее было бы говорить о поиске супремума (***sup**remum*) или инфимума (***inf**imum*), т. е. наибольшего или наименьшего значения, поскольку понятия экстремума (*max*, *min*) ассоциируются с возможностью дифференцирования и обращения в нуль производных функции.

По существующим нормативам для обучения 16 отечественных или 10 иностранных студентов требуется один преподаватель. Статистика показывает, что посещаемость занятий студентами составляет соответственно 40 и 80 %. Ежегодно вуз получает дотацию 2 тыс. денежных единиц на каждого студента-соотечественника и в полтора раза большую сумму с каждого иностранного студента.

Предположив, что единственной целью вуза является максимизация прибыли, попробуем выяснить наилучший план приема.

Обозначим искомую численность студентов соответственно через X и Y (очевидно, что $X \geq 0$ и $Y \geq 0$).

Ограничения на прием можно выразить условиями $X + Y \leq 5000$ и $X \leq 4000$. Ограничение по аудиторному фонду с учетом не слишком высокой посещаемости занятий приводят к соблюдению условия $0.4X + 0.8Y \leq 2800$, а норматив численности студентов на одного педагога порождает условие $X/16 + Y/10 \leq 400$.

Поскольку итоговая денежная сумма составит $2X + 3Y$ тыс. денежных единиц, то задача сведется к минимизации функции

$$L(X, Y) = 2X + 3Y$$

при условиях

$$
\begin{array}{ll}
(1) & X/16 + Y/10 \leq 400 \\
(2) & 0.4X + 0.8Y \leq 2800 \\
(3) & X + Y \leq 5000 \\
(4) & X \leq 4000 \\
(5) & X \geq 0 \\
(6) & \qquad Y \geq 0
\end{array}
$$

Другими словами, нам хочется найти значения X и Y, удовлетворяющие приведенным 6 условиям, при которых функция $L(X,Y)$ принимает самое большое значение.

Так как целевая функция $L(X, Y)$ и ограничения линейны, мы имеем дело с задачей линейного программирования. Учитывая двухмерность этой задачи, попытаемся решить ее графически.

Построив 6 прямых и выделив соответствующие полуплоскости, мы получаем множество допустимых решений (*планов*) в виде некоторого выпуклого многоугольника (рис. 2).

В какой же точке этого множества функция $L(X, Y)$ принимает самое большое значение (плоскость $Z = 2X + 3Y$ наиболее удалена по оси Z от координатной плоскости XY) ?

Если взглянуть на рис. 3, можно прийти к выводу, что такая точка не может быть внутренней точкой многоугольника планов, и наш поиск можно ограничить его вершинами или гранями.

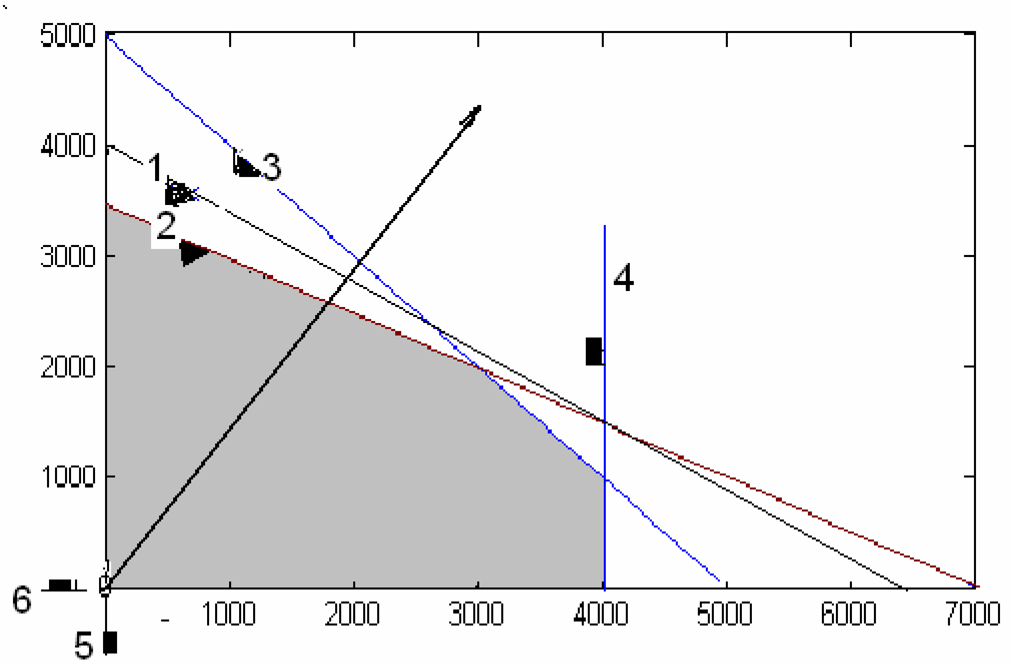


Рис. 2

Поскольку число вершин конечно (в нашем случае их только 5), достаточно найти точки пересечения граничных прямых – координаты этих вершин (решить 5 систем линейных уравнений с двумя неизвестными) и вычислить соответствующие значения $L(X,Y)$.

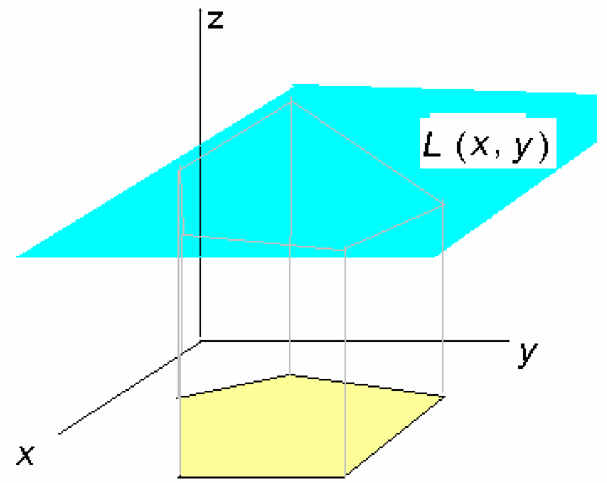


Рис. 3

Если вы хотите минимизировать затраты своей энергии на этот выбор, придется вспомнить понятие *градиента* функции в точке как вектора, составленного из частных производных функции, вычисленных в этой точке, и учесть, что *градиент указывает направление наибольшего возрастания функции в окрестности точки*. Для нелинейных функций градиент меняется от точки к точке. Например, если $f(x, y) = x^2 + xy$, то $grad\,f(x) = \{2x + y, x\}$ в различных точках меняет свою ориентацию. В случае линейной функции составляющие градиента совпадают с коэффициентами целевой функции, например $grad\{L(x, y) = 2x + 3y\} = \{2, 3\}$, т. е. градиент остается неизменным в любой точке плоскости. Опять-таки напрашивается вывод, что экстремумы линейной функции достигаются в вершинах множества планов (или на какой-то грани множества, если градиент перпендикулярен этой грани).

Для нашей задачи максимум явно достигается на пересечении прямых (2) и (3) , т.е. в точке с координатами (3000, 2000).

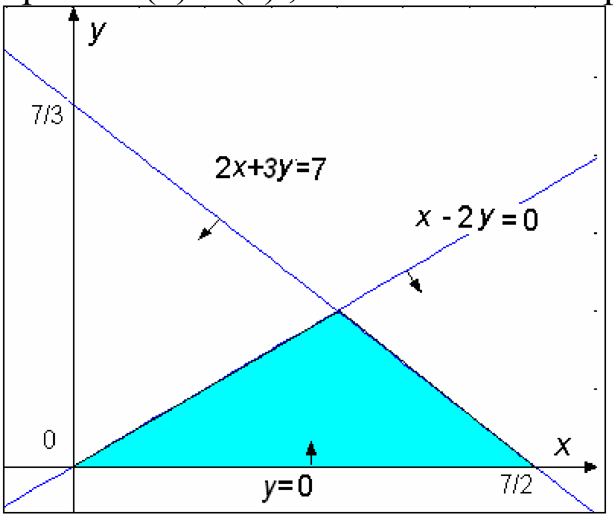


Рис. 4

$2x+3y\le 7$; $x-2y\ge 0$; $y\ge 0$

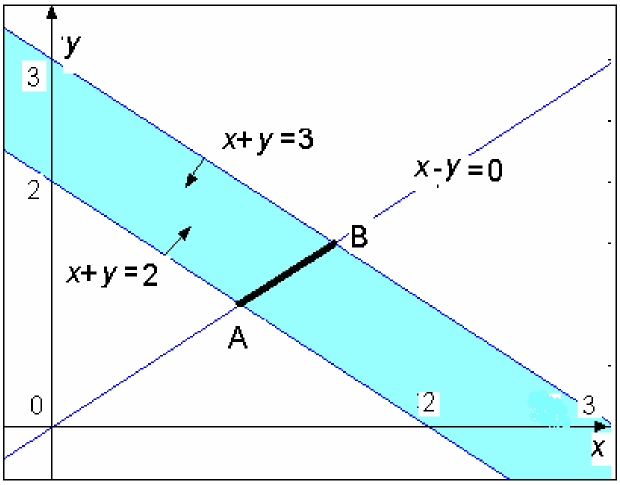


Рис. 5

$2\le x+y\le 3$; $x-y=0$

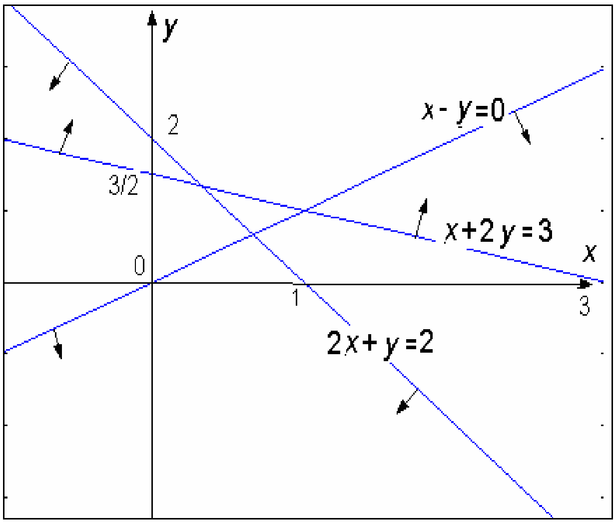


Рис.6

$2x+y\le 2$; $x+2y\ge 3$; $x-y\ge 0$

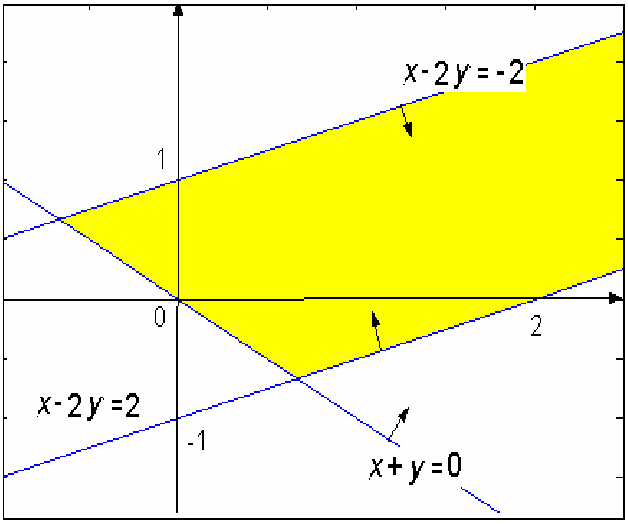


Рис.7

$x-2y\le 2$; $x-2y\le 2$; $x+y\ge 0$

Рассматривая другие примеры систем линейных ограничений, мы делаем окончательный вывод, что ***для двумерной задачи линейного программирования множество планов является выпуклым замкнутым многоугольником***.

В частности, при наличии среди ограничений уравнений, оно может оказаться отрезком (рис. 5), лучом или даже точкой). Ограничения могут оказаться противоречивыми (рис. 6).

Оптимальный план может не существовать (из-за противоречивости ограничений), оказаться единственным или представлять собой бесчисленное множество точек отрезка, служащего одной из граней множества планов (градиент перпендикулярен грани).

Множество планов может оказаться неограниченным (рис. 7). В этом случае может обнаружиться факт *неограниченности значений целевой функции по максимуму и (или) минимуму*.

Итак, в случае графического решения двумерной задачи ЛП достаточно отыскать множество планов (для этого достаточно уметь строить прямые линии, выбирать соответствующую полуплоскость и визуально выделять область, удовлетворяющую всем ограничениям), построить градиент, выбрать вершину – точку искомого экстремума и продемонстрировать умение решать систему линейных уравнений с двумя неизвестными (подстановками или по правилу Крамера).

В заключение заметим, что обнаруженные здесь свойства линейных программ переносятся и на общий случай более чем двух переменных, где прямые превращаются в плоскости, а полуплоскости – в полупространства, многоугольник – в многогранник, но использовать графические приемы решения здесь будет почти всегда нереально

2.2. Общие свойства линейных программ

Количество неизвестных величин, фигурирующих в постановке задачи, называют ее *размерностью*. Набор их значений, удовлетворяющих условиям задачи, называют *планом* (примером может служить программа производства – набор значений показателей, удовлетворяющий ограничениям по сырьевым, социальным и прочим факторам).

В роли неизвестных величин могут выступать, например, объемы выпуска продукции (обуви, приборов ночного видения или микроскопов), распределение денежных средств на ликвидацию ветхого жилья и строительство медицинских учреждений, тираж планируемого издания поэм М. Ю. Лермонтова и заговоров от сглаза. Ограничения связаны с расходом материалов, наличием льгот на отдельные виды изданий и IQ современного российского читателя, рыночной конъюнктурой или личными предпочтениями, а целевая функция $L(X)$ может определять прибыль от произведенной продукции или объем неосвоенных средств.

Общая задача линейного программирования в канонической форме состоит в нахождении вектора[3] $X=(X_1, X_2, .., X_n)$, обеспечивающего наибольшее (наименьшее) значение линейной функции

---

[3] Под термином ***вектор*** в математике понимают просто последовательность величин (вектор-строка или вектор-столбец), не ассоциируя со «стрелкой».

$$L(X) = \sum_{j=1}^{n} C_j X_j \qquad (1)$$

при условиях

$$\sum_{j=1}^{n} A_{ij} X_j = B_i \;\; (i = 1 \dots m) \;, \qquad (2)$$

$$X_j \geq 0 \qquad (j = 1 \dots n) \;. \qquad (3)$$

Остановимся на других формах записи этой задачи.

Если обозначить через A_j вектор[4], составленный из коэффициентов при X_j в (2), а через B – вектор из элементов правой части, то (2) можно записать в векторной форме

$$\sum_{j=1}^{n} A_j X_j = B \;. \qquad (4)$$

Если обозначить через X вектор неизвестных, через A матрицу коэффициентов при неизвестных в (2) и через C вектор коэффициентов линейной функции, задачу можно представить в следующем виде.

Максимизировать (минимизировать)

$$L(X) = C^T X \qquad (1a)$$

при условиях

$$A\,X = B \;, \qquad (2a)$$

$$X \geq 0 \qquad (3a)$$

(здесь T – символ транспонирования).

В дальнейшем мы покажем, что любую линейную программу можно привести к каноническому виду.

Как мы уже отмечали ранее, вектор X, удовлетворяющий ограничениям задачи, называют *планом* и совокупность таких векторов – *множеством планов*.

Если в случае двух или трех переменных мы могли ориентироваться на свои геометрические представления, то при $n > 3$ напрашивающийся по аналогии вывод о том, что *множество планов является выпуклым многогранником*, нетрудно обосновать.

В самом деле, множество планов называют *выпуклым,* если любую его точку X можно представить в виде $X = \lambda X_1 + (1-\lambda) X_2$, где $0 \leq \lambda \leq 1$; X_1 и X_2 – точки этого множества (всякая точка выпуклого множества лежит на отрезке, соединяющем какие-нибудь две точки множества).

Если X_1 и X_2 – планы задачи, то $AX_1 = B$, $X_1 \geq 0$, $AX_2 = B$, $X_2 \geq 0$.

---

[4] Здесь и далее под термином ***вектор*** мы понимаем вектор-столбец.

Для любой точки $X = \lambda X_1 + (1-\lambda)X_2$ соединяющего их отрезка, очевидно $X \geq 0$ и $AX = A(\lambda X_1 + (1-\lambda)X_2) = \lambda AX_1 + (1-\lambda) AX_2 = \lambda B + (1-\lambda) B = B$, т. е. выбранная точка является планом задачи, что доказывает выпуклость множества планов.

Множество планов называют замкнутым, если его граница принадлежит этому множеству. Например, множество $\{x + y \leq 3\}$ – замкнутое, а $\{x + y < 3\}$ – открытое. В нашем случае ограничения выступают в форме равенств или неравенств, допускающих равенство.

Точку X множества называют его вершиной (крайней точкой), если ее нельзя поместить внутри отрезка, соединяющего какие-то две точки множества, т.е. представить в виде $X = \lambda X_1 + (1-\lambda)X_2$, $0 < \lambda < 1$. В нашем случае множество планов ограничено $m+n$ гиперплоскостями и число вершин не превышает числа сочетаний из $m+n$ по n. Но *выпуклое замкнутое множество с конечным числом вершин называется выпуклым многогранником*, что подтверждает сделанные ранее интуитивные выводы.

Если допустить, что оптимальный план X_{opt} является внутренней точкой множества планов, т. е. $X_{opt} = \lambda X_1 + (1-\lambda)X_2$, то в случае задачи минимизации $L(X_{opt}) < L(X_1)$, $L(X_{opt}) < L(X_2)$. Тогда

$$L(X_{opt}) = \lambda L(X_1) + (1-\lambda)L(X_2) > \lambda L(X_{opt}) + (1-\lambda)L(X_{opt}) = L(X_{opt}),$$

что опровергает сделанное допущение. Следовательно, *оптимум линейной функции соответствует только вершинам множества планов*.

Введем используемое при дальнейшем рассмотрении без геометрических ассоциаций понятие опорного плана.

План называется опорным, если он обращает в равенство хотя бы n независимых ограничений (2) – (3). Поскольку в вершине многогранника планов пересекаются хотя бы n граничных гиперплоскостей, понятия опорного плана и вершины множества планов тождественны. Напрашивается вывод, что *оптимальный план всегда является опорным*.

Поиск опорного плана сводится к выбору из $m+n$ соответствующих уравнений с n неизвестными подсистемы n уравнений. Если эта подсистема разрешима (определитель соответствующей матрицы коэффициентов отличен от нуля) и полученное решение удовлетворяет остальным ограничениям, то это решение определяет опорный план.

Поскольку компоненты опорного плана обращают в равенство хотя бы n из имеющихся $m+n$ ограничений, то в (3) не более m ограничений будут выполняться в форме неравенств. Отсюда можно утверждать, что *число положительных компонент опорного плана не превышает m* . Если, например, решается задача с 5 неотрицательными переменными при 3 ограничениях, то в ее оптимальном плане могут быть положительными не более 3 компонент.

Система m векторов A_j при положительных компонентах опорного плана называется *б а з и с о м* этого плана. Эта подсистема соответствующих столбцов матрицы А дает матрицу с ненулевым определителем (признак *линейной независимости векторов базиса*), поскольку в противном случае система уравнений единственного решения не имеет. Знание базиса автоматически определяет соответствующий опорный план. Например, при ограничениях

$$\begin{pmatrix}3\\4\end{pmatrix}x_1+\begin{pmatrix}3\\5\end{pmatrix}x_2+\begin{pmatrix}9\\13\end{pmatrix}x_3+\begin{pmatrix}6\\8\end{pmatrix}x_4=\begin{pmatrix}9\\14\end{pmatrix}$$

за базис можно принять систему векторов A_1 и A_2 (определитель получаемой матрицы отличен от нуля), тогда в опорном плане x_3 и $x_4 = 0$, а составляющие $x_1 = 1$ и $x_2 = 2$ получить из решения системы (разложения вектора B по векторам базиса)

$$\begin{pmatrix}3\\4\end{pmatrix}x_1+\begin{pmatrix}3\\5\end{pmatrix}x_2=\begin{pmatrix}9\\14\end{pmatrix}$$

Можно принять за базис и другие системы векторов, но не A_1 и A_4 (определитель получаемой матрицы обращается в нуль и возникающие системы уравнений свойством единственности решений не обладают).

Опорный план, содержащий ровно m *положительных компонент, называется н е в ы р о ж д е н н ы м* и в противном случае – *вырожденным* (здесь m – число независимых ограничений в (2)).

2.3. Теоретические основы симплексного метода

Симплексный метод – это метод упорядоченного перебора опорных планов (упорядоченность обеспечивается монотонным изменением значения целевой функции при переходе к очередному плану).

Пусть стоит задача максимизации линейной функции

$$L(X)=\sum_{j=1}^{n} C_j X_j \qquad (1)$$

при условиях

$$\sum_{j=1}^{n} A_j X_j = B \quad , \qquad (2)$$

$$X_j \geq 0, \quad j=1,\dots,n\ . \qquad (3)$$

Предположим, что нам удалось найти опорный план X^0, в котором, например, первые m компонент отличны от нуля:

$$X^0=(X_1^{\,0}, X_2^{\,0},\dots,X_m^{\,0}, 0, \dots, 0) \qquad (4)$$

и соответствующий базис $Б = (A_1, A_2, \ldots, A_m)$.

Попытаемся выбрать другую систему базисных векторов с целью построения нового опорного плана, в котором k - я переменная ($k > m$) принимает значение $\Theta > 0$:

$$X(\Theta) = (X_1(\Theta), X_2(\Theta), \ldots, X_m(\Theta),\ 0, \ldots, \Theta, \ldots 0)\ . \tag{5}$$

Подставляя (4) в (2), имеем

$$\sum_{j=1}^{m} A_j X_j^{0} = B \quad . \tag{6}$$

Подставив (5) в (2), получаем

$$\sum_{j=1}^{m} A_j X_j(\Theta) + A_k \Theta = B \quad . \tag{7}$$

Разложим вектор A_k по векторам исходного базиса

$$A_k = \sum_{j=1}^{m} Z_{jk} A_j \quad . \tag{8}$$

Для получения коэффициентов такого разложения придется решать систему m уравнений с m неизвестными, которая имеет единственное решение (еще раз напоминаем о линейной независимости базисных векторов и ненулевом определителе). Заметим, что в ситуации, когда базисные векторы являются единичными (образуют единичную матрицу), искомые коэффициенты совпадают с компонентами исходного вектора; поэтому в дальнейшем мы будем «питать любовь» к единичному базису.

Подставляя (6) и (8) в (7), получаем

$$\sum_{j=1}^{m} A_j X_j(\Theta) + \Theta \sum_{j=1}^{m} A_j Z_{jk} = \sum_{j=1}^{m} A_j X_j^0 \quad , \tag{9}$$

откуда имеем

$$\sum_{j=1}^{m} A_j \left[X_j(\Theta) - X_j^0 + \Theta Z_{jk} \right] = 0\ . \tag{10}$$

Так как система уравнений (10) имеет единственное решение, то получаем представление первых m компонент нового плана

$$X_j(\Theta) = X_j^0 - \Theta Z_{jk} \quad , \ j=1, \ldots, m. \tag{11}$$

Естественно потребовать неотрицательность компонент нового плана. Так как нарушение неотрицательности в (11) может возникнуть лишь при $Z_{jk} > 0$, то значение Θ нужно взять не превышающим наименьшего из отношений X_j^0 к положительным Z_{jk}.

Поскольку число положительных (базисных) компонент опорного плана должно оставаться равным m, то одну из первых m (ненулевых) компонент исходного плана обращаем в нуль выбором

$$\Theta = \min_{Z_{jk}>0} \frac{X_j^0}{Z_{jk}} \quad . \qquad (12)$$

Подставляя (11) в (1), имеем

$$L\{X(\Theta)\} = \sum_{j=1}^{n} C_j X_j(\Theta) = \sum_{j=1}^{m} C_j \left(X_j^0 - \Theta Z_{jk} \right) + C_k \Theta \quad . \qquad (13)$$

Если обозначить

$$Z_k = \sum_{j=1}^{m} C_j Z_{jk} \ , \qquad (14)$$

$$\Delta_k = Z_k - C_k \ , \qquad (15)$$

то (13) примет вид

$$L\{X(\Theta)\} = L(X^0) - \Theta \Delta_k \quad . \qquad (16)$$

Из полученных соотношений напрашиваются следующие выводы.

Критерий 1 (критерий оптимальности). Если все $\Delta_k \geq 0$, выбранный план для задачи максимизации является оптимальным. Для задачи на минимум признак оптимальности – неположительность всех Δ_k.

Критерий 2. Если обнаруживается некоторое $\Delta_k < 0$ и хотя бы одно из значений $Z_{jk} > 0$, то переход к новому плану увеличит значение целевой функции.

Этот вывод с очевидностью следует из (16); в такой ситуации согласно (12) полагаем k-ю переменную равной Θ и преобразуем значения остальных (базисных) переменных в соответствии с (11).

Критерий 3. Если обнаруживается некоторое $\Delta_k < 0$, но все $Z_{jk} \leq 0$, то целевая функция не ограничена по максимуму (неограниченность по минимуму устанавливается аналогично при $\Delta_k > 0$ и всех $Z_{jk} \leq 0$).

Этот вывод следует из того, что согласно (11) компоненты нового плана сохраняют неотрицательность при любом $\Theta > 0$ (в том числе и при произвольно большом) и согласно (16) появляется возможность неограниченного изменения значения целевой функции.

Предположение о том, что базисными являются первые m компонент плана, не является принципиальным и указание диапазона по j от 1 до m в (11)–(15) можно заменить на указание о принадлежности к базису « $j \in Б$ ».

Если все опорные планы задачи являются невырожденными (число положительных компонент равно m), то Θ отлично от нуля и переход к новому плану согласно (16) изменяет значение целевой функции, что из-за ограниченности количества опорных планов гарантирует достижение экстремума за конечное число шагов. При наличии вырожденных планов возможно т. н. *зацикливание* (возврат к ранее рассмотренным планам), но на практике зацикливание никогда не возникало.

2.4. Прямой алгоритм симплексного метода

Пусть исходная задача приведена к канонической форме и начальный базис образует единичную матрицу.

Тогда базисные компоненты опорного плана совпадают с правыми частями ограничений и коэффициенты Z_{jk} разложения вектора A_k по такому базису совпадают с компонентами этого вектора.

Для единообразия описания вычислительной процедуры в дальнейшем будем пользоваться т. н. симплексными таблицами вида:

C баз	Базис плана	План X	C_1	C_2	. . .	C_m	C_{m+1}	. . .	C_k	. . .	C_n
			A_1	A_2	. . .	A_m	A_{m+1}	. . .	A_k	. . .	A_n
C_1	A_1	B_1	1	0	. . .	0	$Z_{1,m+1}$	. . .	Z_{1k}	. . .	Z_{1n}
C_2	A_2	B_2	0	1	. . .	0	$Z_{2,m+1}$	. . .	Z_{2k}	. . .	Z_{2n}
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
C_m	A_m	B_m	0	0	. . .	1	$Z_{m,m+1}$	. . .	Z_{mk}	. . .	Z_{mn}
Z_k		$L(X)$	Z_1	Z_2	. . .	Z_m	Z_{m+1}	. . .	Z_k	. . .	Z_n
Δ_k			Δ_1	Δ_2	. . .	Δ_m	Δ_{m+1}	. . .	Δ_k	. . .	Δ_n

В центральной части таблицы записываются коэффициенты при неизвестных в ограничениях, в столбце X – правая часть ограничений (базисные компоненты плана), в первой строке – коэффициенты целевой функции (линейной формы) .

В первом столбце для удобства вычислений будем заносить коэффициенты линейной формы при базисных переменных (умножение его на столбец X с суммированием дает значение $L(X)$; аналогичное умножение его на столбец A_k даст Z_k). Последняя строка получается вычитанием из предыдущей строки элементов первой строки таблицы.

Пусть стоит задача максимизации

$$L(X) = X_1+2X_2+2X_3 - X_5$$

при условиях

$$\begin{aligned} 2X_1 + X_2 + X_3 + X_4 &= 9 \\ X_1 + 2X_2 + X_3 + X_5 &= 8 \\ X_1 + 3X_2 + 2X_3 + X_6 &= 15 \end{aligned}$$

$$X_k \geq 0, \quad k = 1,\ldots,6 .$$

Здесь легко видеть наличие единичного базиса (A_4, A_5, A_6) и можно приступать к решению.

Переносим исходные данные в симплексную таблицу и получаем оценки для выбранного начального опорного плана.

C баз	Базис плана	План X	1	2	2	0	−1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	9	2	1	1	1	0	0
−1	A_5	8	1	**2**	1	0	1	0
0	A_6	15	1	3	2	0	0	1
Z_k		$L(X)$	−1	−2	−1	0	−1	0
Δ_k		−8	−2	**−4**	−3	**0**	**0**	**0**

Так как имеются отрицательные значения Δ_k, выбранный план не оптимален.

Введем в базис, например, вектор A_2 (перейти к другому опорному плану, в котором $X_2 \neq 0$).

Отыскав значение $\Theta = min\,(\frac{9}{1}, \frac{8}{2}, \frac{15}{3}) = 4$, видим, что для сохранения неотрицательности компонент будущего плана из базиса нужно вывести вектор A_5. Соответственно, выражаем X_2 из второго уравнения (второе уравнение делим на коэффициент при X_2, т. е. на 2) и исключаем X_2 из остальных уравнений (из выбранного уравнения вычитаем разрешенное, умноженное на коэффициент при X_2).

C баз	Базис плана	План X	1	2	2	0	−1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	5	3/2	0	1/2	1	−1/2	0
2	A_2	4	1/2	1	1/2	0	1/2	0
0	A_6	3	−1/2	0	**1/2**	0	−3/2	1
Z_k		$L(X)$	1	2	1	0	1	0
Δ_k		8	0	**0**	**−1**	**0**	2	**0**

Так как имеется отрицательное значение Δ_3, выбранный план не оптимален. Введем в базис вектор A_3 и, отыскивая значение $\Theta = min(\frac{5}{1/2}, \frac{4}{1/2}, \frac{3}{1/2}) = 6$, видим, что для сохранения неотрицательности компонент будущего плана из базиса нужно вывести вектор A_6. Соответственно, выражаем X_3 из третьего уравнения (третье уравнение делим на коэффициент при X_3, т. е. на 1/2) и исключаем X_3 из остальных уравнений (из выбранного уравнения вычитаем разрешенное, умноженное на

C баз	Базис плана	План X	1	2	2	0	−1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	2	**2**	0	0	1	1	−1
2	A_2	1	1	1	0	0	2	−1
2	A_3	6	−1	0	1	0	−3	2
Z_k		$L(X)$	0	2	2	0	−2	4
Δ_k		14	**−1**	**0**	**0**	**0**	−1	4

коэффициент при X_3). Так как имеются отрицательные значения Δ_k, выбранный план не оптимален.

Вводя в базис, например, вектор A_1 и отыскивая значение

$$\Theta = \min\left(\frac{2}{2}, \frac{1}{1}, -\right) = 1,$$

видим, что из базиса можно удалить A_4 или A_2 (заметьте, что мы не рассматриваем отношение к отрицательной компоненте вектора).

C баз	Базис плана	План X	1	2	2	0	-1	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	1	0	0	1/2	1/2	–1/2
2	A_2	0	0	1	0	–1/2	**3/2**	–1/2
2	A_3	7	0	0	1	1/2	–5/2	3/2
Z_k		$L(X)$	1	2	2	1/2	-3/2	3/2
Δ_k		15	**0**	**0**	**0**	1/2	**–1/2**	3/2

Перейдем к другому плану вводом в базис A_5 вместо A_2.

Так как все $\Delta_k \geq 0$, найденный план с компонентами (1, 0, 7, 0, 0, 0) оптимален и максимум значений $L(X)$ равен 15.

C баз	Базис плана	План X		2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	1	–1/3	0	2/3	0	–1/3
–1	A_5	0	0	2/3	0	–1/3	1	–1/3
2	A_3	7	0	5/3	1	–1/3	0	2/3
Z_k		$L(X)$	1	7/3	2	1/3	–1	4/3
Δ_k		15	**0**	1/3	**0**	1/3	**0**	4/3

Обратите внимание на то, что последние два плана вырожденные (число положительных компонент меньше 3) и переход между ними не сопровождался изменением значения $L(X)$.

Полезно заметить, что здесь обращаются в нуль значения Δ_k только для базисных векторов и, следовательно, отсутствует возможность существования других оптимальных планов (с тем же значением целевой функции).

Рассмотрим другой пример, состоящий в максимизации

$$L(X) = 2X_1 - X_2$$

при условиях

$$-X_1 + X_2 + X_3 = 1$$
$$3X_1 - 5X_2 + X_4 = 6$$
$$X_k \geq 0,\ k = 1, \ldots, 4$$

C баз	Базис плана	План X	2	–1	0	0
			A_1	A_2	A_3	A_4
0	A_3	1	–1	1	1	0
0	A_4	6	**3**	–5	0	1
Z_k		$L(X)$	0	0	0	0

C баз	Базис плана	План X	2	–1	0	0
			A_1	A_2	A_3	A_4
0	A_3	3	0	**–2/3**	1	1/3
2	A_1	2	1	**–5/3**	0	1/3
Z_k		$L(X)$	2	–10/3	0	2/3

Δ_k	0	**–2**	1	**0**	**0**

Δ_k	6	**0**	**–7/3**	**0**	2/3

После перехода от начального плана к очередному обнаруживаем $\Delta_2 < 0$, найденный план не оптимален. Попытка ввода в базис вектора A_2 обнаруживает, что коэффициенты его разложения по базису неположительны и, соответственно, целевая функция не ограничена сверху (может принимать произвольно большие значения).

2.5. Приведение задачи к канонической форме

Прежде чем приступать к решению линейной программы симплексным методом, необходимо привести задачу к канонической форме. В приведенных выше примерах задачи были поставлены в канонической форме и с очевидностью отыскивался начальный базис и опорный план.

Рассмотрим более общий случай.

При приведении линейной программы к каноническому виду используют следующие приемы.

Если на некоторую переменную X_k отсутствуют условия неотрицательности, то ее заменяют разностью двух неотрицательных переменных

$$X_k = X'_k - X''_k\ ,\ X'_k \geq 0,\ X'' \geq 0.$$

Если же на некоторую переменную стоит условие неположительности, производится замена $X_k = -X'_k\ ,\ X'_k \geq 0$.

Если некоторое из основных ограничений допускает неравенство, то вводят неотрицательную т. н. *ослабляющую* (*свободную*, *дополнительную*) переменную, уравновешивающую разность между левой и правой частями ограничения.

Например, ограничения

$$X_1 + 2X_2 + 3X_3 \geq 7$$
$$X_1 - X_2 + X_3 \leq 2$$

преобразуются к виду

$$X_1 + 2X_2 + 3X_3 - X_4 = 7,\quad X_4 \geq 0$$
$$X_1 - X_2 + X_3 + X_5 = 2\ ,\quad X_5 \geq 0$$

Если среди основных ограничений присутствует ограничение с отрицательной правой частью, из соображений удобства последующего выбора начального опорного плана это ограничение умножают на -1.

2.6. Выбор начального опорного плана

Пусть задача приведена к канонической форме и компоненты вектора правой части неотрицательны. Если в системе векторов коэффициентов при переменных (матрице A) обнаруживается подсистема, образующая единичную подматрицу, то эти векторы образуют базис опорного плана и вектор правой части определяет базисные компоненты этого плана.

Если такой единичной подматрицы не обнаруживается, то либо придется перебирать все подсистемы m уравнений с m неизвестными в надежде обнаружить неотрицательные решения, либо прибегнуть к методу искусственного базиса.

В последнем случае в ограничения добавляют неотрицательные, т. н. *искусственные переменные* так, чтобы возникла единичная подматрица коэффициентов, и эти переменные включают в целевую функцию с коэффициентом +М для задачи минимизации и с коэффициентом – М для задачи минимизации (М > 0 – большое число).

Полученная М-задача решается до получения оптимального плана.

Если в оптимальном плане М-задачи значения искусственных переменных равны нулю, то значения остальных компонент образуют оптимальный план исходной задачи.

Если в оптимальном плане М-задачи значение хотя бы одной из искусственных переменных отлично от нуля, то исходная задача не имеет ни одного плана (*ее ограничения противоречивы*).

Пусть стоит задача максимизации линейной функции

$$L(X)=X_1+2X_2-X_3$$

при условиях

$$\begin{aligned} 2X_1+X_2-X_3&\le 7\\ 4X_1-X_2+4X_3&\ge 6\\ -X_1+X_2+X_3&\le 2\\ X_k\ge 0,\quad k&=1,2,3. \end{aligned}$$

Приводим задачу к канонической форме, вводя три ослабляющих переменных, и получаем задачу в следующем виде.

Максимизировать

$$L(X)=X_1+2X_2-X_3$$

при условиях

$$\begin{aligned} 2X_1+X_2-X_3+X_4&=7\\ 4X_1-X_2+4X_3-X_5&=6\\ -X_1+X_2+X_3+X_6&=2\\ X_k\ge 0,\ k&=1,\dots,6. \end{aligned}$$

При попытке искать начальный единичный базис обнаруживаем лишь два единичных вектора A_4 и A_6. Вводим во второе уравнение искусственную переменную $X_7\ge 0$ и включаем ее в линейную форму с коэффициентом $-$M, получая задачу:

максимизировать

$$L(X) = X_1 + 2X_2 - X_3 - M X_7$$

при условиях

$$\begin{aligned} 2X_1 + X_2 - X_3 + X_4 &= 7 \\ 4X_1 - X_2 + 4X_3 - X_5 + X_7 &= 6 \\ -X_1 + X_2 + X_3 + X_6 &= 2 \end{aligned}$$

$$X_k \geq 0,\ k=1,\ldots,7.$$

Решаем поставленную задачу прямым алгоритмом симплекс-метода (не забывайте, что M > 0 $-$ очень большое число).

C баз	Базис плана	План X	1	2	-1	0	0	0	$-$M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
0	A_4	7	2	1	-1	1	0	0	0
$-$M	A_7	6	**4**	-1	4	0	-1	0	1
0	A_6	2	-1	1	1	0	0	1	0
Z_k		-6M	$-$4M	M	$-$4M	0	M	0	$-$M
Δ_k			**$-$4M$-$1**	M$-$2	$-$4M$-$1	**0**	M	**0**	**0**

C баз	Базис плана	План X	1	2	-1	0	0	0	$-$M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
0	A_4	4	0	**3/2**	-3	1	1/2	0	$-1/2$
1	A_1	3/2	**1**	$-1/4$	1	0	$-1/4$	0	1/4
0	A_6	7/2	0	3/4	**2**	0	$-1/4$	1	1/4
Z_k		3/2	1	$-1/4$	1	0	$-1/4$	0	1/4
Δ_k			**0**	**$-$9/4**	2	**0**	$-1/4$	**0**	M+1/4

C баз	Базис плана	План X	1	2	-1	0	0	0	$-$M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
2	A_2	8/3	0	1	-2	2/3	1/3	0	$-1/3$
1	A_1	13/6	1	0	1/2	1/6	$-1/6$	0	1/6
0	A_6	3/2	0	0	**7/2**	$-1/2$	$-1/2$	1	1/2
Z_k		15/2	1	2	$-7/2$	3/2	1/2	0	$-1/2$
Δ_k			**0**	**0**	**$-$5/2**	3/2	1/2	**0**	M-1/2

C баз	Базис плана	План X	1	2	-1	0	0	0	$-$M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
2	A_2	74/21	0	1	0	8/21	1/21	4/7	$-1/21$
1	A_1	41/21	1	0	0	5/21	$-2/21$	$-1/7$	2/21

–1	A_3	3/7	0	0	1	–1/7	–1/7	2/7	1/7
Z_k		60/7	1	2	–1	8/7	1/2	5/7	–1/7
Δ_k			**0**	**0**	**0**	8/7	1/2	**5/7**	M-1/7

Поскольку искусственная переменная обратилась в нуль, мы получили оптимальный план исходной задачи

$$X_{opt} = (\frac{41}{21}, \frac{74}{21}, \frac{3}{7}) \ , \ L_{max} = 60/7 \ .$$

Рассмотрим еще один пример. Минимизировать

$$L(X) = 2X_1 - X_2 + X_3$$

при условиях

$$X_1 + X_2 + X_3 \le 3$$
$$-3X_1 + X_2 - X_3 \ge 6$$
$$X_k \ge 0, k=1, 2, 3 \ .$$

Приводим задачу к канонической форме вводом ослабляющих переменных X_4 и X_5 и для поиска начального базиса добавим искусственную переменную X_6.

С баз	Базис плана	План X	2	–1	1	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	3	1	**1**	1	1	0	0
M	A_6	6	–3	1	– 1	0	–1	1
Z_k		6M	–3M	M	–M	0	–M	M
Δ_k			–3M–2	**M+1**	–M–1	**0**	–M	**0**

С баз	Базис плана	План X	2	–1	1	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6
– 1	A_2	3	1	1	1	1	0	0
M	A_6	3	–4	0	–2	– 1	–1	1
Z_k		3M-3	–4M–1	–1	–2M–1	–M–1	–M	M
Δ_k			–4M–3	**0**	–2M–2	–M–1	–M	**0**

Так как все $\Delta_k < 0$, то найденный план оптимален. Однако значение искусственной переменной $X_6 = 3 > 0$, что свидетельствует о противоречивости ограничений поставленной задачи.

Здесь мы рассмотрели примеры практически всех ситуаций, возникающих при решении линейной программы симплексным методом.

Число «шагов» симплексного процесса имеет порядок $m \div 2m$, что существенно меньше возможного количества опорных планов, определяемого числом сочетаний из $m + n$ по n.

Замечание. *Если среди основных ограничений $L > 1$ условий представлены отношением $\ge$, то для уменьшения числа искусственных пе-*

ременных после приведения к канонической форме можно выбрать уравнение с максимальной правой частью, а другие заменить результатом их вычитания из выбранного. В итоге получим $L-1$ единичных векторов и потребность в единственной искусственной переменной) .

2.7. Двойственность в линейном программировании

Пусть стоит задача максимизации

$$L(X) = C^T X \quad (1)$$

при условиях

$$A\,X = B \quad (2)$$

$$X \geq 0 \quad (3)$$

Задача минимизации

$$\tilde{L}\ (Y) = B^T Y \quad (4)$$

при условии

$$A^T Y \geq C \quad (5)$$

называется *сопряженной* к задаче (1)–(3), и обе задачи образуют пару двойственных задач.

Например, для задачи максимизации

$$L(X) = X_1 + 3X_2 - X_3$$

при условиях

$$\begin{aligned} 4X_1 - 7X_2 + X_3 &= 5 \\ 6X_1 + 8X_2 - 2X_3 &= 9 \\ X_1 \geq 0,\ X_2 \geq 0,\ X_3 &\geq 0 \end{aligned}$$

сопряженная задача состоит в минимизации

$$\tilde{L}\ (Y) = 5Y_1 + 9Y_2$$

при условиях

$$\begin{aligned} 4Y_1 + 6Y_2 &\geq 1 \\ -7Y_1 + 8Y_2 &\geq 3 \\ Y_1 - 2Y_2 &\geq -1 \end{aligned}$$

Из (1)–(5) приведением к канонической форме и путем простых преобразований можно получить другую пару двойственных задач.

Минимизировать	Максимизировать
$L(X) = C^T X$	$\tilde{L}\,(Y) = B^T Y$
при	при
$A\,X = B$	$A^T\,Y \leq C$
$X \geq 0$	

Пусть стоит задача максимизации

$$L(X) = X_1 + 3X_2 - X_3$$

при условиях

$$4X_1 - 7X_2 + X_3 \leq 5$$

$$6X_1 + 8X_2 - 2X_3 \le 9$$
$$X_1 \ge 0, \quad X_2 \ge 0, \quad X_3 \ge 0$$

Приводя ее к каноническому виду, получаем задачу максимизации $L(X) = X_1 + 3X_2 - X_3$ при условиях

$$4X_1 - 7X_2 + X_3 + X_4 = 5$$
$$6X_1 + 8X_2 - 2X_3 + X_5 = 9$$
$$X_1 \ge 0, \quad X_2 \ge 0, \quad X_3 \ge 0, \quad X_4 \ge 0, \quad X_5 \ge 0$$

Тогда сопряженная задача состоит в минимизации

$$\widetilde{L}(Y) = 5Y_1 + 9Y_2$$

при условиях

$$4Y_1 + 6Y_2 \ge 1$$
$$-7Y_1 + 8Y_2 \ge 3$$
$$Y_1 - 2Y_2 \ge -1$$
$$Y_1 \ge 0, \quad Y_2 \ge 0$$

Из этого примера можно усмотреть и т. н. *симметричную пару* двойственных задач.

Максимизировать	Минимизировать
$L(X) = C^T X$	$\widetilde{L}(Y) = B^T Y$
при	при
$AX \le B$	$A^T Y \ge C$
$X \ge 0$	

2.7.1. Первая теорема двойственности

Обратимся к первичной в нашем рассмотрении паре задач.

Максимизировать	Минимизировать
$L(X) = C^T X$	$\widetilde{L}(Y) = B^T Y$
при	при
$AX = B$	$A^T Y \ge C$
$X \ge 0$	

Установим ряд частных выводов.

1. Пусть X и Y – некоторые планы задач. Тогда с учетом матричного соотношения $(AB)^T = B^T A^T$

$$L(X) = C^T X \le (A^T Y)^T X = Y^T A X = Y^T B = B^T Y = \widetilde{L}(Y),$$

т. е. для любых планов $L(X) \le \widetilde{L}(Y)$.

2. Пусть для планов X^* и Y^* имеет место равенство $C^T X^* = B^T Y^*$. Тогда $C^T X \le B^T Y^* = C^T X^*$, т. е. при $X = X^*$ достигается максимум $L(X)$. С учетом $B^T Y \ge C^T X^* = B^T Y^*$ делаем вывод об оптимальности планов X^* и Y^*.

3. Пусть линейная функция сопряженной задачи $\widetilde{L}(Y)$ не ограни-

чена снизу. Тогда существует план Y такой, что $B^T Y < -M$, где $M > 0$ большое число. Если бы существовал хотя бы один план X исходной задачи, то $C^T X \le B^T Y < -M$, чего при конечных значениях X не может быть.

Отсюда можно получить следующее утверждение.

Если одна из двойственных задач разрешима, то разрешима и сопряженная задача. При этом для оптимальных планов X^* и Y^* этих задач имеет место равенство значений целевых линейных функций

$$C^T X^* = B^T Y^*.$$

Если целевая линейная функция одной из задач не ограничена, то ограничения сопряженной задачи противоречивы. Если в одной из задач противоречивы ограничения, то в другой задаче либо не ограничена целевая функция, либо противоречивы ограничения.

2.7.2. Вторая теорема двойственности

Условия $X_j \ge 0$ и $\sum_{i=1}^{m} A_{ij} Y_i \ge C_j$ называют п а р о й двойственных условий. Условие называют с в о б о д н ы м на некотором плане, если оно выполняется как строгое неравенство (в противном случае его называют з а к р е п л е н н ы м).

Без доказательства предлагаем следующую теорему.

Если пара двойственных задач разрешима, то для их оптимальных планов в каждой паре двойственных условий одно является свободным и другое – закрепленным.

<u>Пример 1.</u> Примем за исходную задачу максимизации

$$5X_1+3X_2$$

при условиях

$$\begin{aligned}
&(1)\quad X_1+2X_2 \le 4\\
&(2)\quad X_1 - X_2 \ge 2\\
&(3)\quad X_1 \ge 0\\
&(4)\quad X_2 \ge 0
\end{aligned}$$

После преобразования к канонической форме получаем задачу максимизации

$$5X_1+3X_2$$

при условиях

$$\begin{aligned}
&X_1+2X_2 + X_3 = 4\\
&X_1 - X_2 - X_4 = 2\\
&(3)\quad X_1 \ge 0\\
&(4)\quad X_2 \ge 0
\end{aligned}$$

$$(1)\qquad X_3 \ge 0$$
$$(2)\qquad X_4 \ge 0\ .$$

Сопряженная задача состоит в минимизации

$$4Y_1+2Y_2$$

при условиях

$$(3)\qquad Y_1+Y_2 \ge 5$$
$$(4)\qquad 2Y_1-Y_2 \ge 3$$
$$(1)\qquad Y_1 \ge 0$$
$$(2)\qquad -Y_2 \ge 0.$$

Таким образом установлены следующие пары условий:

$$(1)\qquad X_1+2X_2 \le 4\ ,\qquad Y_1 \ge 0\ ;$$
$$(2)\qquad X_1-X_2 \ge 2\ ,\qquad -Y_2 \ge 0\ ;$$
$$(3)\qquad X_1 \ge 0,\qquad Y_1+Y_2 \ge 5\ ;$$
$$(4)\qquad X_2 \ge 0,\qquad 2Y_1-Y_2 \ge 3$$

Если графически или симплексной процедурой удастся найти оптимальный план одной из задач, например X_{opt} = (4,0), то по теореме оптимальный план сопряженной задачи отвечает условиям:

$$(1)\ \boxed{X_1+2X_2 = 4,\quad Y_1 \ge 0}\ ;\qquad (3)\ \boxed{X_1 > 0\ ,\ Y_1+Y_2 = 5}\ ;$$
$$(2)\ \boxed{X_1-X_2 > 2,}\ -\boxed{Y_2= 0};\qquad (4)\ \boxed{X_2 = 0\ ,\ 2Y_1-Y_2 \ge 3}\ .$$

Из пары равенств $-Y_2 = 0$ и $Y_1 + Y_2 = 5$ получаем решение Y = (5,0), удовлетворяющее всем остальным условиям и, следовательно, являющееся оптимальным планом сопряженной задачи. Подтверждением тому является и равенство значений линейной формы $5 \cdot 4 + 3 \cdot 0 = 4 \cdot 5 + 2 \cdot 0$.

Заметим, что если бы мы ошиблись в выборе X_{opt} , то получили бы противоречивые условия для Y_{opt}. Соответственно, вторая теорема двойственности выступает основой проверки планов на оптимальность.

<u>Пример 2</u>. Минимизировать

$$3X_1+4X_2$$

при условиях

$$(1)\qquad X_1 + 2X_2 \ge 5$$
$$(2)\qquad 2X_1 + X_2 \ge 5$$
$$(3)\qquad 4X_1 + 3X_2 \le 20$$
$$(4)\qquad 2X_1 - X_2 = 2$$
$$(5)\qquad X_2 \ge 0\ .$$

После приведения к канонической форме минимизируем

$$3X_1' - 3X_1'' + 4X_2$$

при условиях

$$X_1' - X_2'' + 2X_2 - X_3 \qquad = 5 \Big| \quad \Big|$$

$$\begin{array}{ll} & 2X_1' - 2X_1'' + X_2 - X_4 = 5 \\ & 4X_1' - 4X_1'' + 3X_2 + X_5 = 20 \\ (4) & 2X_1' - 2X_1'' - X_2 = 2 \\ & X_1' \ge 0 \quad X_1'' \ge 0 \\ (5) & X_2 \ge 0 \\ (1) & X_3 \ge 0 \\ (2) & X_4 \ge 0 \\ (3) & X_5 \ge 0 \end{array}$$

Сопряженная задача сведется к максимизации

$$5Y_1 + 5Y_2 + 20Y_3 + 2Y_4$$

при условиях

$$\begin{array}{ll} & Y_1 + 2Y_2 + 4Y_3 + 2Y_4 \le 3 \\ & -Y_1 - 2Y_2 - 4Y_3 - 2Y_4 \le -3 \\ (5) & 2Y_1 + Y_2 + 3Y_3 - Y_4 \le 4 \\ (1) & -Y_1 \le 0 \\ (2) & -Y_2 \le 0 \\ (3) & Y_3 \le 0 \end{array}$$

(первые два неравенства обращаются в одно равенство).

Таким образом, мы получили пару двойственных задач :

минимизировать		максимизировать
$3X_1 + 4X_2$		$5Y_1 + 5Y_2 + 20Y_3 + 2Y_4$
при условиях		при условиях
(1) $X_1 + 2X_2 \ge 5$	$\leftrightarrow$	$-Y_1 \le 0$
(2) $2X_1 + X_2 \ge 5$	$\leftrightarrow$	$-Y_2 \le 0$
(3) $4X_1 + 3X_2 \le 20$	$\leftrightarrow$	$Y_3 \le 0$
(4) $2X_1 - X_2 = 2$		
(5) $X_2 \ge 0$	$\leftrightarrow$	$2Y_1 + Y_2 + 3Y_3 - Y_4 \le 4$
		$Y_1 + 2Y_2 + 4Y_3 + 2Y_4 = 3$

Отсюда можно сделать два полезных вывода:

– если в исходной задаче нет условия неотрицательности на некоторую переменную, то построенное по коэффициентам при этой переменной ограничение сопряженной задачи выполняется равенством;

– если некоторое (основное) ограничение исходной задачи задано равенством, то двойственного условия сопряженной задачи не существует.

Пусть каким-то путем (например, графически) удалось найти оптимальный план исходной задачи $X_{opt} = (9/5, 8/5)$. Подставив его в ограничения (1), (2), (3) и (5) этой задачи, получаем выполнение соотношений в форме = , > , < , >. Если оптимальный план найден правильно, то для компонент оптимального плана сопряженной задачи

$$Y_1 \geq 0\ ,\ -Y_2 = 0\ ,\ Y_3 = 0,$$
$$2Y_1 + Y_2 + 3Y_3 - Y_4 = 4 \quad \text{и} \quad Y_1 + 2\,Y_2 + 4\,Y_3 + 2\,Y_4 = 3\ ,$$

откуда $Y_{opt} = (\ 1/5, 0, 0, 2/5\ )$.

Если некоторая задача решается прямым алгоритмом симплексного метода, то *решение сопряженной задачи можно видеть в строке Z конечной симплексной таблицы в позициях, соответствующих начальному единичному базису*.

Решим симплексным методом сопряженную задачу. После приведения ее к канонической форме вводом ослабляющей переменной Y_5 и выполнения замен $Y_3 = -Y_3'$, $Y_4 = Y_4' - Y_4''$ получаем задачу максимизации

$$5\,Y_1 + 5\,Y_2 - 20\,Y_3' + 2\,Y_4' - 2\,Y_4''$$

при условиях

$$Y_1 + 2\,Y_2 - 4\,Y_3' + 2\,Y_4' - 2\,Y_4'' = 3$$
$$2\,Y_1 + Y_2 - 3\,Y_3' - Y_4' + Y_4'' + Y_5 = 4\ ,$$
$$\text{все } Y_k \geq 0.$$

C баз	Базис плана	План Y	5	5	–20	2	–2	0	–M
			Y_1	Y_2	Y_3'	Y_4'	Y_4''	Y_5	Y_6
–M	Y_6	3	1	2	–4	**2**	–2	0	1
0	Y_5	4	2	1	–3	–1	1	1	0
Z_k		–3M	–M	–2M	4M	–2M	2M	0	–M
Δ_k			–M–5	–2M–5	4M+20	**–2M–2**	2M+2	**0**	**0**
2	Y_4'	3/2	1/2	1	–2	1	–1	0	1/2
0	Y_5	11/2	**5/2**	2	–5	0	0	1	1/2
Z_k		3	1	2	–4	2	–2	0	1
Δ_k			**–4**	–3	16	**0**	0	**0**	M+1
2	Y_4'	2/5	0	3/5	–1	1	–1	–1/5	2/5
5	Y_1	11/5	1	4/5	–2	0	0	2/5	1/5
Z_k		**59/5**	5	26/5	–12	2	–2	**8/5**	**9/5**
Δ_k			**0**	1/5	8	**0**	0	8/5	M+9/5

Как мы предупреждали выше, в последних двух позициях строки Z_k таблицы (исходный единичный базис – векторы Y_6 и Y_5) имеем решение задачи, двойственной к решаемой, – $X = (9/5,\ 8/5)$.

В этом можно убедиться, если учесть, что при симплексных преобразованиях (обычных тождественных преобразованиях для системы линейных алгебраических уравнений) на месте исходной единичной подматрицы получается матрица, обратная для матрицы, составленной из базисных векторов.

Базисные компоненты плана равны произведению этой обратной матрицы на исходный вектор правой части системы, а решение сопряженной задачи – произведению коэффициентов линейной функции при базисных переменных на эту матрицу. Так здесь, умножая обратную матрицу

$$\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2/5 & -1/5 \\ 1/5 & 4/5 \end{bmatrix}$$

(см. исходный базис Y_6, Y_5) на соответствующие векторы

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2/5 \\ 11/5 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 5 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}^{-1} = (8/5,\ 9/5)$$

Нетрудно заметить, что разумнее из двух задач решать симплексным методом ту, которая после приведения к канонической форме содержит наименьшее число основных (типа $AX = B$) ограничений.

К постановке двойственных задач часто приходится прибегать при решении больших линейных программ со специфическими матрицами коэффициентов. В дальнейшем мы столкнемся с примерами такого рода при рассмотрении т. н. транспортных, распределительных и других задач.

2.7.3. Экономическая интерпретация симметрической пары двойственных задач

Представьте себе, что вы решаете некоторую производственную задачу, сформулированную в виде задачи линейного программирования.

Например, решая задачу выпуска продукции при ограниченных ресурсах (видах сырья), вы ввели серию обозначений:

X_j – искомый объем производства j-го вида продукта ($j = 1 \div n$),

B_i – запас i-го вида сырья ($i = 1 \div m$),

A_{ij} – затраты i-го сырья на создание единицы j-го продукта,

C_j – стоимость единицы j-го продукта.

В результате некоторых размышлений вы пришли к задаче поиска значений X_j, удовлетворяющих условиям (расход сырья на производство продукции не превышает его запасов)

$$\sum_{j=1}^{n} A_{ij} X_j \le B_i \quad (i = 1 \ldots m),$$

$$X_j \ge 0 \qquad (j = 1 \,..\, n)$$

и к тому же обеспечивающих максимальное значение выручки от продажи продукции

$$L(X) = \sum_{j=1}^{n} C_j X_j$$

Подставив конкретные данные (расценки на готовую продукцию, объем запасов сырья, расходные нормы и т. д.), в итоге кропотливого труда получаете оптимальную производственную программу.

Событие радостное, но всякое полученное знание требует своего развития, заставляет задуматься, в частности, о разумности принятых расценок на продукцию, о соответствии их тем ценам, по которым вы приобретали сырье.

Обозначив через Y_i ***объективную оценку стоимости*** *единицы i -го сырья*, осознаем, что совокупная стоимость ресурсов, затраченных на производство единицы j-го продукта, должна быть не меньше объявленной стоимости, то есть

$$\sum_{i=1}^{m} A_{ij} Y_i \ge C_j \quad , j = 1 \, .. \, n$$

$$Y_i \ge 0 \qquad (i = 1 \, .. \, m),$$

и общая стоимость затраченного сырья должна быть минимальной

$$\widetilde{L}(Y) = \sum_{i=1}^{m} B_i Y_i \to min$$

Высказанные соображения приводят к целесообразности решения т. н. *симметричной пары двойственных задач*:

максимизировать	минимизировать
$L(X) = C^T X$	$\widetilde{L}(Y) = B^T Y$
при $A X \le B$	при $A^T Y \ge C$
$X \ge 0$	$Y \ge 0$

Напомним, что в приведенной матричной записи C, B, X, Y выступают как векторы-столбцы (T – знак транспонирования).

Здесь известная первая теорема двойственности требует равенства стоимости затраченных ресурсов и объявленной стоимости произведенной продукции.

Что касается второй теоремы о соотношениях в парах двойственных условий для оптимальных планов, то обнаружение неравенства $\sum_{j=1}^{n} A_{ij} X_j < B_i$ при некотором i говорит о том, что i-е сырье не является лимитирующим и его объективная стоимость $Y_i = 0$.

Таким образом, в рассмотренной постановке *двойственная задача является математической формулировкой объективной оценки всех производственных факторов.*

2.7.4. Постоптимальный анализ и устойчивость решений

Получив оптимальный план (программу производства продукции на очередной период), вы не гарантированы от того, что завтра изменятся цены на сырье, часть какого-то сырья окажется похищенной или изменится технология производства (последнее встречается в жизни гораздо реже остальных случаев). Останется ли найденный вами план, по своей структуре, оптимальным или потребуется искать таковой заново?

Другими словами, обладает ли найденное оптимальное решение устойчивостью при изменении значений параметров задачи (исходных данных)? В каких диапазонах можно менять значения тех или иных параметров при сохранении оптимальности найденного плана?

Пусть нам удалось получить решение задачи максимизации функции $L(X) = C^T X$ при условиях $AX \le B$, $X \ge 0$, то есть найден оптимальный план и (самое существенное!) соответствующая ему базисная система векторов условий $Б_{опт}$. Как мы уже замечали выше, если задача решалась каким-то алгоритмом симплексного метода, то в последней симплексной таблице на месте исходного единичного базиса получается матрица $Б_{опт}^{-1}$, обратная к матрице, составленной из векторов упомянутой системы[5].

Можно показать, что ненулевые (базисные) составляющие $X_{опт}$ определяются как произведение $Б_{опт}^{-1} B$, а компоненты сопряженной задачи как $C_{опт}^T Б_{опт}^{-1}$, где $C_{опт}^T$ – коэффициенты целевой функции $L(X)$ при базисных переменных.

В векторно-матричной записи высказанные замечания представимы в виде

$$X_{опт} = Б_{опт}^{-1} B \;,\; Y_{опт}^T = C_{опт}^T \, Б_{опт}^{-1}$$

Отыскав

[5] Квадратная матрица, символически обозначенная как A^{-1}, называется *обратной* по отношению к квадратной матрице A, если их произведение дает единичную матрицу $A^{-1}A = AA^{-1} = E$. Обратная матрица существует, если строки (или столбцы) исходной матрицы линейно независимы, то есть если определитель отличен от нуля. Здесь существование обратной матрицы несомненно, поскольку базис – система линейно независимых векторов.

$$L(X_{опт}) = C^T_{опт} X_{опт} = C^T_{опт}\ Б^{-1}_{опт} B$$

$$\widetilde{L}(Y_{опт}) = B^T Y_{опт} = B^T (C^T_{опт}\ Б^{-1}_{опт})^T =$$

$$= B^T (Б^{-1}_{опт})^T C_{опт} = (C^T_{опт}\ Б^{-1}_{опт} B)^T$$

и убедившись в равенстве правых частей этих выражений, можете лишний раз увидеть выполнение первой теоремы двойственности.

Например, решая задачу максимизации

$$L(X) = 3\,x_1 + 2\,x_2$$

при условиях

$$x_1 + 5\,x_2 \le 7$$
$$5\,x_1 + x_2 \le 8$$
$$x_1,\ x_2 \ge 0$$

получаем последовательность симплексных таблиц, где матрица оптимального базиса состоит из векторов коэффициентов при x_2 и x_1:

C баз	Базисные переменные	План *X*	3	2	0	0
			x_1	x_2	x_3	x_4
0	x_3	7	1	5	1	0
0	x_4	8	5	1	0	1
z_k		0	0	0	0	0
Δ_k			–3	–2	·	·
0	x_3	27/5	0	24/5	1	–1/5
3	x_1	8/5	1	1/5	0	1/5
z_k		24/5	3	3/5	0	3/5
Δ_k			·	–7/5	·	3/5
2	x_2	**9/8**	0	1	**5/24**	**–1/24**
3	x_1	**11/8**	1	0	**–1/24**	**5/24**
z_k		**51/8**	3	2	**7/24**	**13/24**
Δ_k			·	·	7/24	13/24

Поскольку здесь базис оптимального плана образован векторами коэффициентов при x_2 и x_1, а начальный базис – векторами коэффициентов при x_3 и x_4, то

$$\acute{A}_{\hat{i}\ddot{i}\grave{o}} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}, \qquad Б^{-1}_{опт} = \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix}.$$

Можете убедиться, что оценки

$$Б^{-1}_{опт} B = \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 9/8 \\ 11/8 \end{bmatrix} = [x_2, x_1]^T{}_{опт},$$

$$C^T_{опт}\,Б^{-1}_{опт} = = [2 \quad 3]\cdot\begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} = \begin{bmatrix} \frac{7}{24} & \frac{13}{24} \end{bmatrix} = Y_{опт}$$

соответствуют выводам, фигурирующим в итоговой симплексной таблице.

Из приведенного примера видно, что знание базиса оптимального плана автоматически дает возможность найти компоненты этого плана и решение сопряженной задачи (не забывайте обращать внимание на порядок следования базисных переменных).

Вообразите теперь, что компоненты вектора B подверглись изменению: все b_i изменились на ε_i .

Останется ли базис оптимального плана неизменным и решение можно искать как $X_{опт} = Б^{-1}_{опт}B$ при любом B ?

Очевидно, что это справедливо, если компоненты соответствующего оптимального плана останутся неотрицательными

$$X_{опт} = Б^{-1}_{опт}(B+\varepsilon) \ge 0 \;.$$

Для нашего примера

$$X_{опт} = Б^{-1}_{опт}(B+\varepsilon) =$$

$$= \begin{bmatrix} \frac{5}{24} & -\frac{1}{24} \\ -\frac{1}{24} & \frac{5}{24} \end{bmatrix} \cdot \begin{bmatrix} 7+\varepsilon_1 \\ 8+\varepsilon_2 \end{bmatrix} = \begin{bmatrix} 9/8 \\ 11/8 \end{bmatrix} + \begin{bmatrix} \frac{5}{24}\varepsilon_1 - \frac{1}{24}\varepsilon_2 \\ -\frac{1}{24}\varepsilon_1 + \frac{5}{24}\varepsilon_2 \end{bmatrix} \ge 0,$$

то есть условие сохранности базиса определяется решением системы неравенств:

$$\begin{matrix} \frac{9}{8} + \frac{5}{24}\varepsilon_1 - \frac{1}{24}\varepsilon_2 \ge 0 \\ \frac{11}{8} - \frac{1}{24}\varepsilon_1 + \frac{5}{24}\varepsilon_2 \ge 0 \end{matrix} \quad \equiv \quad \begin{matrix} -5\varepsilon_1 + \varepsilon_2 \le 27 \\ \varepsilon_1 - 5\varepsilon_2 \le 33 \end{matrix}$$

Если вам известны значения ожидаемых корректур ε_i, то достаточно проверить для них выполнение полученных неравенств. Если же ограничиться вариантами, где меняется единственное из значений b_i при сохранности остальных, то вынести суждение можно без особого труда.

Так, если принять $\varepsilon_2 = 0$, то получается система неравенств с одной переменной, решение которой доступно школьнику:

$$\frac{9}{8} + \frac{5}{24}\varepsilon_1 \ge 0$$

$$\frac{11}{8} - \frac{1}{24}\varepsilon_1 \ge 0$$

отсюда имеем $-27/5 \le \varepsilon_i \le 33$, то есть при изменении ресурса b_1 в диапазоне [7–27/5=8/5; 7+33=40] базис оптимального плана остается неизменным и для выяснения диапазона изменения компонент оптимального плана достаточно искать $X_{опт} = Б_{опт}^{-1} B$ при предельных значениях b_1.

Аналогично можно сделать выводы при корректуре коэффициентов целевой функции. Так при корректуре коэффициента при x_1 имеем

$$C_{опт}^{T}\, Б_{опт}^{-1} = [2 \quad , 3+\Delta]\cdot \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} = \left[\frac{7}{24} - \frac{\Delta}{24}; \quad \frac{13}{24} + \frac{5\Delta}{24}\right] \geq 0,$$

откуда допустимый диапазон корректуры $-13/5 \leq \Delta \leq 7$.

Знание $Б_{опт}^{-1}$ может облегчить анализ ситуации, когда появляются новые виды продукции со своими характеристиками, определяемые вектором A_k.

Не пересчитывая симплексную таблицу заново, можно получить соответствующий ее столбец умножением $Б_{опт}^{-1}$ на A_k.

2.8. Параметрическое линейное программирование

Параметрическое программирование представляет собой один из разделов математического программирования, изучающий задачи, в которых целевая функция или ограничения зависят от одного или нескольких параметров.

Необходимость рассмотрения подобных задач обусловлена различными причинами, основная из которых связана с тем, что исходные данные для численного решения любой реальной задачи оптимизации практически всегда определяются приближенно или могут изменяться под влиянием каких-то факторов, что может существенно сказаться на оптимальности выбираемой программы (плана) действий. В роли таких факторов могут выступать время, температура, курс доллара по отношению к рублю, цена на сырье, удаленность от поставщиков и др. Соответственно, чтобы быть готовым к изменениям ситуации (исходных данных), при решении оптимизационной задачи разумно указывать не конкретные данные, а *диапазон возможного изменения данных*.

С математической точки зрения параметрическое программирование выступает как одно из средств *анализа чувствительности решения к вариации исходных данных*, оценки *устойчивости* решения.

В отличие от *постоптимального анализа*, где мы пытаемся выяснить диапазоны вариации характеристик задачи, при которых найденный оптимальный план остается оптимальным, здесь мы включаем некоторые параметры в математическую модель и строим оптимальные решения как функции от них.

Сразу оговоримся, что универсальных методов анализа устойчивости решений произвольной задачи математического программирования нет и в случае множественности параметров или нелинейных связей трудности такого анализа даже с помощью современных компьютеров исключительны.

Рассмотрим задачу параметрического линейного программирования, в которой **только коэффициенты целевой функции линейно зависят от некоторого единственного параметра** λ :

отыскать максимум (или минимум) функции

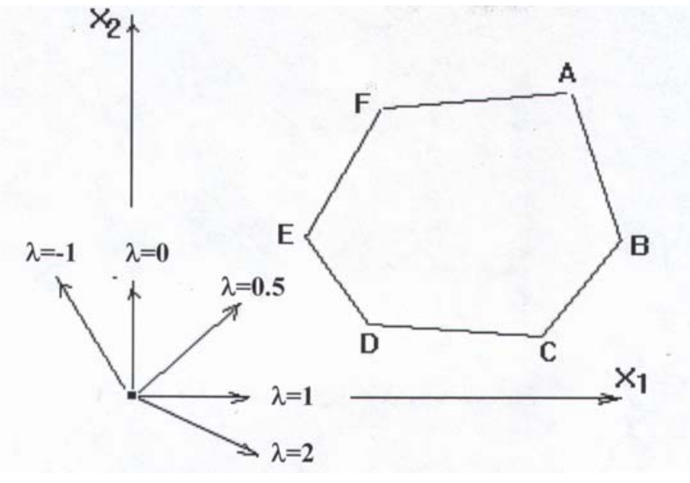
$$L(X, \lambda) = \sum_{j=1}^{n} (C_j + D_j\lambda)X_j$$

при условиях

$$\sum_{j=1}^{n} A_j X_j = B, \quad X_j \geq 0, \quad j = 1 \ldots n; \quad \lambda_1 \leq \lambda \leq \lambda_2.$$

Если обратиться к геометрической интерпретации задачи, то можно заметить, что градиент целевой функции зависит от параметра. Например, для целевой функции $L(X, \lambda) = \lambda X_1 + (1-\lambda)X_2$ при различных значениях параметра λ градиент определяет различные направления роста функции.

Нетрудно видеть (см. рисунок), что, если при некотором значении параметра максимум достигается в вершине А, то небольшая вариация этого значения несколько изменит направление градиента, но не изменит положения точки максимума. Отсюда напрашивается вывод, что некоторый план, оптимальный при $\lambda = \lambda_0$, оптимален и в окрестности λ_0, т. е. при $\alpha \leq \lambda \leq \beta$, где $\lambda_0 \in [\alpha, \beta]$.



Можно заметить, что при градиенте, ставшем перпендикулярным некоторой стороне многоугольника планов, имеем два разных оптимальных опорных плана с одним и тем же значением линейной формы, откуда можно утверждать непрерывность экстремума линейной формы по λ .

В случае неограниченности множества планов можно утверждать, что *если линейная форма не ограничена при каком-то* $\lambda = \lambda_0$ *, то она не ограничена при всех* λ *, больших или меньших* λ_0.

Процедура решения задач параметрического линейного программирования в случае зависимости от параметра коэффициентов целевой функции незначительно отличается от обычного симплексного метода (примеры решения подобных задач приведены ниже).

В случае **зависимости от параметра компонент вектора правых частей ограничений,** т. е. при поиске экстремума функции

$$L(X) = \sum_{j=1}^{n} C_j \; X_j$$

при условиях

$$\sum_{j=1}^{n} A_j X_j = B + \lambda D \;, \quad X_j \geq 0 \;, \quad j = 1 \ldots n \;; \qquad \lambda_1 \leq \lambda \leq \lambda_2 \,,$$

во избежание сложностей, связанных с требованием сохранения неотрицательности компонент плана при любых λ (сохранения неотрицательности правой части системы уравнений при всех ее тождественных преобразованиях), достаточно поставить и решить сопряженную задачу, воспользовавшись вышеупомянутым алгоритмом решения задач параметрического линейного программирования при зависимости от параметра коэффициентов целевой функции, и с помощью известных двойственных соотношений отыскать решение исходной задачи.

Пример 1. Рассмотрим задачу минимизации

$$L(X, \lambda) = \lambda X_1 - \lambda X_2 - X_3 + X_4$$

при условиях

$$3X_1 - 3X_2 - X_3 + X_4 \geq 5;$$
$$2X_1 - 2X_2 + X_3 - X_4 \leq 3;$$
$$X_k \geq 0 \;, k = 1 \,..\, 4 \;; \quad -\infty < \lambda < \infty.$$

Как обычно, приводим задачу к канонической форме и с использованием метода искусственного базиса отыскиваем начальный опорный план $X^0 = (0, 0, 0, 0, 0, 3, 5)$ с $L(X^0,\lambda) = 5$M.

C баз	Базис 1	План X	λ	$-\lambda$	-1	1	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
M	A_7	**5**	3	–3	–1	**1**	–1	**0**	**1**
0	A_6	**3**	2	–2	1	–1	0	**1**	**0**
Δ_k		5M	3M–λ	–3M+λ	–M+1	**M–1**	–M	0	0

Так как определяющую роль на этом шаге решения играет величина M, превышающая все величины задачи, то не обращаем внимания на

λ и, обнаружив невыполнение критерия оптимальности для начального опорного плана, вводим в базис A_4 вместо A_7 (переходим к следующему опорному плану):

C баз	Базис 2	План X	λ	$-\lambda$	-1	1	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_4	**5**	3	–3	–1	**1**	–1	**0**
0	A_6	**8**	**5**	–5	0	**0**	–1	**1**
Δ_k		5	**3 –λ**	–3+λ	0	0	–1	0

Полученный опорный план $X^1 = (0, 0, 0, 5, 0, 8)$ с $L(X^1, \lambda) = 5$ будет оптимальным, если все значения Δ_k неположительны, т. е.

$$\begin{cases} \Delta_1 \equiv 3 - \lambda \le 0 \\ \Delta_2 \equiv -3 + \lambda \le 0 \end{cases} \qquad \begin{cases} \lambda \ge 3 \\ \lambda \le 3 \end{cases}.$$

Решаем систему двух линейных неравенств и обнаруживаем, что найденный *план* X^1 *оптимален при* $\lambda = 3$.

Исследуем оставшиеся из заданного диапазона значения λ.

Пусть $\lambda > 3$. Тогда $\Delta_2 > 0$ и вектор A_2 подлежит вводу в базис, но в силу неположительности его компонент приходим к выводу, что *при* $\lambda > 3$ *целевая функция не ограничена снизу.*

Пусть $\lambda < 3$. Тогда $\Delta_1 > 0$ и в базис вводится вектор A_1:

C баз	Базис 3	План X	λ	$-\lambda$	-1	1	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_4	**1/5**	**0**	0	–1	**1**	–2/5	–3/5
λ	A_1	**8/5**	**1**	–1	0	0	–1/5	1/5
Δ_k		(8λ+1)/5	0	0	0	0	– **(λ+2)/5**	(λ–3)/5

Полученный опорный план является оптимальным, если все значения Δ_k неположительны, т. е.

$$\begin{array}{l|l} \Delta_5 = -(\lambda+2)/5 \le 0 & \lambda \ge -2 \\ \Delta_6 = (\lambda - 3)/5 \le 0 & \lambda \le 3 \end{array}$$

Очевидно, что найденный *план* $X = (8/5, 0, 0, 1/5)$ с $L(X, \lambda) = (8\lambda+1)/5$ *оптимален при* $-2 \le \lambda \le 3$.

Пусть $\lambda < -2$. Тогда $\Delta_5 > 0$ и вектор A_5 подлежит вводу в базис; в силу неположительности его компонент приходим к выводу, что *при* $\lambda < -2$ *целевая функция не ограничена снизу.*

Таким образом, мы получили решение задачи:

$$L_{min}(X,\lambda)=\begin{cases}\to -\infty, & \lambda< -2;\\ (8\lambda+1)/5, & -2\le \lambda\le 3; \quad X_{opt}=(8/5\ ,0\ ,\ 0\ ,\ 1/5\ )\\ 5, & \lambda=3\ ; \quad X_{opt}=(\ 0\ ,\ 0\ ,\ 0,\ 5\ )\\ \to -\infty, & \lambda> 3\ .\end{cases}$$

Пример 2. Рассмотрим задачу минимизации

$$L(X,\lambda) = (2-\lambda)\, X_1 - 3\lambda\, X_2 + (\lambda - 3)X_3$$

при условиях

$$X_1 + X_2 + X_3 \le 5$$
$$3\, X_1 - X_2 - 2\, X_3 \le 6$$
$$X_1 + 2\, X_2 + 2\, X_3 \le 8$$
$$X_k \ge 0\ ,\quad k = 1\ ..\ 3; \qquad -\infty < \lambda < \infty\ .$$

C баз	Базис 1	План X	$2-\lambda$	-3λ	$\lambda-3$	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	5	1	1	1	1	0	0
0	A_5	6	3	–1	–2	0	1	0
0	A_6	8	1	2	**2**	0	0	1
Δ_k		0	$\lambda-2$	3λ	$\boxed{\mathbf{3-\lambda}}$	0	0	0

Находим начальный опорный план задачи X^0 = (0, 0, 0, 5, 6, 8) с $L(X^0, \lambda) = 0$, который был бы оптимален при выполнении условий:

$$\Delta_1 = \lambda - 2 \le 0\ , \qquad \Delta_2 = 3\lambda \le 0, \qquad \Delta_3 = 3 - \lambda \le 0\ .$$

Однако попытка решения этой системы трех линейных неравенств обнаруживает её противоречивость ($\lambda \le 2$, $\lambda \le 0$, $\lambda \ge 3$).

Пусть $\lambda < 3$. Тогда $\Delta_3 > 0$ и вводим в базис A_3:

C баз	Базис 2	План X	$2-\lambda$	-3λ	$\lambda-3$	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	1	1/2	0	0	1	0	–1/2
0	A_5	14	4	1	0	0	1	1
$\lambda-3$	A_3	4	1/2	**1**	1	0	0	1/2
Δ_k		$4(\lambda-3)$	$(3\lambda-7)/2$	$\boxed{\mathbf{4\lambda-3}}$	0	0	0	$(\lambda-3)/2$

Полученный опорный план оптимален, если

$$\Delta_1 = (3\lambda-7)/2 \le 0\ , \qquad \Delta_2 = 4\lambda-3 \le 0, \qquad \Delta_6 = (\lambda-3)/2 \le 0\ .$$

Решение этой системы неравенств обнаруживает *оптимальность плана* X=(0, 0, 4) с $L(X, \lambda) = 4(\lambda-3)$ *при* $\lambda \le 3/4$.

Пусть $\lambda > 3/4$. Тогда $\Delta_2 > 0$ и вводим в базис A_2:

C баз	Базис 3	План $X_{баз}$	$2-\lambda$	-3λ	$\lambda-3$	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6

0	A_4	1	1/2	0	0	1	0	–1/2
0	A_5	10	7/2	0	–1	0	1	1/2
-3λ	A_2	4	1/2	1	1	0	0	1/2
Δ_k		-12λ	$-(\lambda+4)/2$	0	$-4\lambda+3$	0	0	$-3\lambda/2$

Полученный план оптимален, если

$$\Delta_1 = -(\lambda+4)/2 \le 0, \quad \Delta_3 = -4\lambda+3 \le 0, \quad \Delta_6 = -3\lambda/2 \le 0 .$$

Решение системы трех неравенств обнаруживает, что *план* $X = (0, 4, 0)$ с $L(X, \lambda) = -12\lambda$ *оптимален при всех* $\lambda \ge 3/4$.

Таким образом, рассмотрен весь диапазон значений λ. Задача решена:

$$L_{min}(X,\lambda) = \begin{cases} 4(\lambda-3), \ \lambda \le \dfrac{3}{4}; & X_{opt} = (0, 0, 4) \\ -12\lambda, \ \lambda \ge \dfrac{3}{4}; & X_{opt} = (0, 4, 0). \end{cases}$$

<u>Пример 3</u>. Рассмотрим задачу максимизации

$$L(X, \lambda) = X_1 - X_2 - 2X_3$$

при условиях

$$X_1 + X_2 + X_3 \le 3 + \lambda;$$
$$2X_1 - X_2 + X_3 \le 5 - \lambda;$$
$$X_k \ge 0, \quad k = 1 .. 3; \qquad -\infty < \lambda < \infty.$$

Чтобы решить эту задачу, достаточно решить двойственную к ней задачу, имеющую вид:

минимизировать

$$L(Y, \lambda) = (3+\lambda)Y_1 + (5-\lambda)Y_2$$

при условиях

$$Y_1 + 2Y_2 \ge 1;$$
$$Y_1 - Y_2 \ge -1;$$
$$Y_1 + Y_2 \ge -2;$$
$$Y_1, Y_2 \ge 0;$$
$$-\infty < \lambda < \infty .$$

Приводим двойственную задачу к канонической форме (умножив предварительно второе и третье неравенства на –1) и начинаем обычное решение обычным симплексным методом. Заметьте, что указанное умножение тождественно смене знака у переменных x_2 и x_3 исходной задачи.

C баз	Базис 1	План Y	$3+\lambda$	$5-\lambda$	0	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6

M	A_6	1	**1**	2	–1	0	0	1
0	A_4	1	–1	1	0	1	0	0
0	A_5	2	–1	–1	0	0	1	0
Δ_k		M	**M –3–λ**	2M–5+λ	–M	0	0	0

C баз	Базис 2	План Y	3+λ	5–λ	0	0	0	**M**
			A_1	A_2	A_3	A_4	A_5	A_6
3+λ	A_1	1	1	**2**	–1	0	0	1
0	A_4	2	0	3	–1	1	0	1
0	A_5	3	0	1	–1	0	1	1
Z_k		3+λ	3+λ	6+2λ	– (3+λ)	**0**	**0**	**3+λ**
Δ_k			0	1+3λ	– (3+λ)	0	0	–M+..

Найденный план Y= (1, 0) оптимален, если $\Delta_2 = (1+3\lambda) \le 0$ и $\Delta_3 = -(3+\lambda) \le 0$, т. е. *при* $-3 \le \lambda \le -1/3$ $Y_{opt} = (1, 0)$.

В строке Z_k (в позициях 6, 4 и 5 в соответствии с начальным базисом) находим решение прямой задачи: X_{opt} =(3+λ , 0 , 0), $L(X_{opt}) = 3+\lambda$.

Пусть $\lambda <-3$. Попытка ввода в базис вектора A_3 обнаруживает, что в этом случае *целевая функция решаемой (двойственной) задачи не ограничена снизу и, следовательно, ограничения исходной задачи противоречивы*.

В случае $\lambda >-1/3$ имеем:

C баз	Базис 3	План Y	3+λ	5–λ	0	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6
5–λ	A_2	1/2	1/2	1	–1/2	0	0	1/2
0	A_4	1/2	–3/2	0	**1/2**	1	0	–1/2
0	A_5	5/2	–1/2	0	–1/2	0	1	1/2
$\mathbf{Z_k}$		(5–λ)/2	(5–λ)/2	5–λ	– (5–λ)/2	**0**	**0**	**(5–λ)/2**
Δ_k			– (3λ+1)/2	0	**– (5–λ)/2**	0	0	–M+...

Решив систему неравенств

$$\Delta_1 = -(3\lambda+1)/2 \le 0,$$
$$\Delta_3 = -(5-\lambda)/2 \le 0,$$

обнаруживаем, что *при* $-1/3 \le \lambda \le 5$ $Y_{opt} = (0, 1/2)$, $X_{opt} = ((5-\lambda)/2, 0, 0)$, $L(X_{opt}) = (5-\lambda)/2$.

Продолжаем решение задачи при $\lambda > 5$. Получаем:

C баз	Базис 4	План Y	3+λ	5–λ	0	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6

$5-\lambda$	A_2	1	–1	1	0	1	0	0
0	A_3	1	–3	0	1	2	0	–1
0	A_5	3	–2	0	0	1	1	0
$\mathbf{Z_k}$		$5-\lambda$	$-(5-\lambda)$	$5-\lambda$	0	$\boldsymbol{5-\lambda}$	**0**	**0**
Δ_k			–8	0	0	$5-\lambda$	0	–M

Видим, что

при $\lambda \geq 5$ $Y_{opt} = (0, 1)$, $X_{opt}=(0, -5+\lambda, 0)$, $L(X_{opt}) = 5-\lambda$.

Интервал значений параметра λ исчерпан, выявлены четыре интервала устойчивости оптимальных решений задачи.

Диапазон λ	Сопряженная задача	Исходная задача
$\lambda < -3$	$L(Y,\lambda) \to -\infty$	*ограничения противоречивы*
$-3 \leq \lambda \leq -1/3$	$Y_{opt} = (1, 0)$	$X_{opt} = (3+\lambda , 0 , 0)$, $L(X_{opt}) = 3+\lambda$
$-1/3 \leq \lambda \leq 5$	$Y_{opt} =(0, 1/2)$	$X_{opt} = ((5-\lambda)/2, 0, 0)$, $L(X_{opt}) = (5-\lambda)/2$
$\lambda \geq 5$	$Y_{opt} = (0, 1)$	$X_{opt} = (0, -5+\lambda, 0)$, $L(X_{opt}) = 5-\lambda$

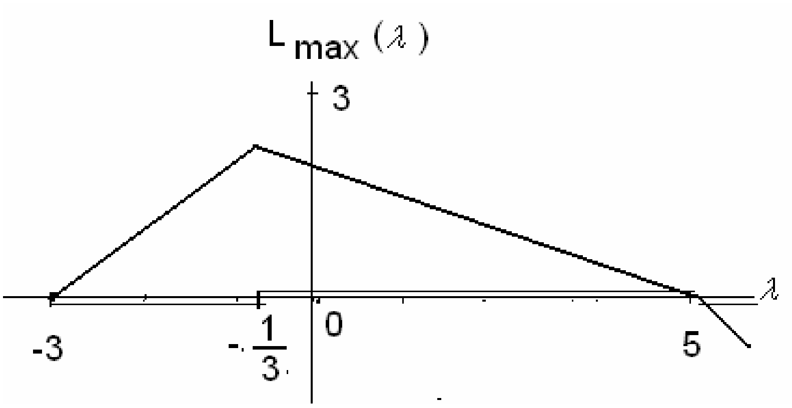


Рис. 8. График максимальных значений функции $L(\lambda)$

Увы, в случае *зависимости от параметра компонент матрицы ограничений* подобного простого *универсального* подхода к решению не существует.

Из рассмотренных примеров видно, что иногда на начальных этапах возникают противоречивые системы неравенств. Во избежание этого можно начинать процесс с поиска оптимального плана для λ_{min} (или λ_{max}). Отыскав таковой, можно приступать к последующему анализу подынтервалов оптимальности. Такой подход является естественным и при программной реализации задачи.

Использованный здесь подход применим и в случае нелинейной зависимости от параметра (например, квадратичной), но здесь возникают системы нелинейных неравенств, решение которых часто представляет достаточно сложную, нестандартную задачу.

Иногда рассмотренные приемы проверки на оптимальность могут оказаться небесполезными в ситуации нескольких параметров, но решение систем линейных неравенств относительно нескольких переменных – не из тех задач, которые можно доверить даже самой умной машине.

3. ЦЕЛОЧИСЛЕННОЕ ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

3.1. Постановка задачи

Задача линейного целочисленного программирования формулируется следующим образом:

найти такое решение (план)

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

при котором линейная функция

$$L(X) = \sum_{j=1}^{n} c_j x_j$$

принимает максимальное (*или минимальное*) *значение при ограничениях*

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

$$x_j \geq 0, \ j = 1, 2, \dots n$$

$$x_j \text{ - целые числа}, j = 1, 2, \dots n.$$

Заметим, что последнее условие может накладываться не на все неизвестные и тогда говорят о *частично целочисленных* задачах.

Где же могут возникнуть такие задачи?

Естественно, при решении многих задач оптимального планирования искомые величины могут быть только целочисленными. Такие задачи связаны с определением численности работников в структурных подразделениях предприятия, количества станков в заводском цехе, объемом производства или транспортировки дорогой крупногабаритной, неделимой продукции (подводных лодок, вертолетов, угольных комбайнов) и т. д. Пытаться округлять получаемые оценки в большую или меньшую сторону неразумно: или мы превысим свои ресурсные возможности, или не используем их в должной мере (излишки иногда можно употребить с пользой на что-нибудь еще).

Иногда встречаются задачи, где некоторая величина x_s может принимать лишь какие-то дискретные значения, например, +1, 0 или –1. Если провести замену $x_s = 1 \cdot z_1 + 0 \cdot z_2 - 1 \cdot z_3$ и наложить требования

$$z_1 + z_2 + z_3 = 1, \quad z_1, \ z_2, \ z_3 \geq 0 \text{ –целые числа},$$

то получим эквивалентную задачу целочисленного программирования. К подобным относятся, например, задачи с фиксированными размерами производимых или закупаемых партий каких-то продуктов.

Достаточно часто возникают задачи с так называемыми *булевыми переменными*, решениями которых являются суждения типа «да-нет». Если значению «да» сопоставить единицу, а «нет» – нуль, то добавлением условий { $0 \le x_s \le 1$, x_s *–целое* } мы опять-таки получаем целочисленную программу. Линейность целевой функции и ограничений при этом сохраняется.

Встречаются задачи, где целевая функция содержит нелинейные элементы типа

$$C(x) = \begin{cases} A + B \cdot x, & \text{если } x > 0 \\ 0, & \text{если } x = 0 \end{cases}$$

(здесь величина B определяет непосредственные затраты на единицу производимой продукции, тогда как A – затраты на подготовку производства, отсутствующие в периоды производственного простоя).

Если ввести новую переменную $0 \le \delta \le 1$ и потребовать ее целочисленности, то заменой $C(x) = A \cdot \delta + B \cdot x$ и дополнительным условием $x \le M \cdot \delta$, где M – верхняя граница для x или попросту очень большое число, получаем решение целочисленной линейной программы целочисленную линейную программу с числом переменных на 1 большим.

Часто задачи целочисленного программирования называют задачами дискретного, или диофантова[6] программирования.

3.2. Метод последовательных отсечений (метод Гомори)

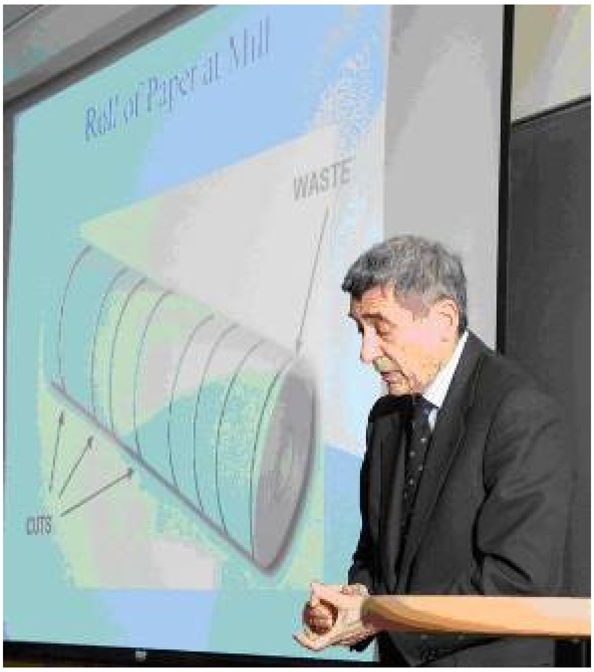
Сущность упомянутого метода заключается в последовательном построении дополнительных ограничений, «отсекающих» найденные нецелочисленные решения (нецелочисленные оптимальные планы) задачи, но «не отсекающих» ни одного целочисленного плана. Мы пользуемся термином «отсекает» в смысле «делает неприемлемым», «делает недопустимым», «делает не удовлетворяющим новой системе условий».

Хотя идея такого подхода принадлежит Дж.Данцигу[7] и Р. Гомори[8],

---

[6] Диофант Александри́йский (III век н. э.) – древнегреческий математик. Основное его произведение «Арифметика», содержащее множество интересных задач и методов их решения, по существу заложило основы символьной математики (алгебры), в дальнейшем развитые Ф. Виета и П. Ферма. Одна из главных задач этой книги – поиск целочисленных (рациональных, по Диофанту) решений т. н. неопределённых уравнений.

[7] Джордж Бернард Данциг (1914 –2005)– видный американский математик, предложил симплексный метод в современной форме и, вместе с Л. В.Канторовичем, является создателем *линейного программирования* (термин введен в 1949 г.).



Р. Гомори

достаточно эффективная ее реализация с гарантией конечности процесса отсечений справедливо называется методом Гомори.

Алгоритм решения задачи линейного целочисленного программмирования этим методом определяется следующими действиями.

1. Решаем поставленную задачу обычным симплексным методом без учета условия целочисленности. Естественно, что если все компоненты оптимального плана окажутся целочисленными, то он дает решение поставленной задачи. Также очевидно, что и обнаружение неразрешимости задачи при отказе от требования целочисленности свидетельствует о неразрешимости и целочисленной программы.

2. Если же среди компонент найденного оптимального решения есть нецелые, то к ограничениям задачи добавляем новое ограничение, обладающее следующими свойствами:

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

Эта идея впервые была представлена в форме дополнительного ограничения:

$$\sum_{j \notin \text{Базису}} X_j \ge 1 \tag{1}$$

(сумма небазисных компонент оптимального плана должна быть отлична от нуля, т. е. хотя бы одна из небазисных компонент должна быть ненулевой). В самом деле, оптимальный план с нулевыми значениями небазисных компонент этому условию не удовлетворяет, что подтверждает разумность отсечения этого плана от исходного множества.

К сожалению, для абсолютного большинства задач, где значения целочисленных переменных меняются в большом диапазоне, скорость

[8] Ральф Гомори (род. 7 мая 1929 г.). Окончил Кембриджский университет, в 1954 году получил докторскую степень по математике в Принстонском университете. Вернувшись на работу в Принстон после службы во флоте (1954–1957), занялся математическими методами исследования операций и создал первые алгоритмы отсечения планов (*cutting-plane algorithms*). С 1959 года работал в IBM, до 2007 года возглавлял исследовательские работы этого гиганта компьютерного мира.

сходимости процесса таких отсечений мала.

Позднее Р. Гомори предлагает другой вариант – выбирать компоненту оптимального плана ***с наибольшей дробной частью*** и по соответствующей этой компоненте k-й строке симплексной таблицы (k-му уравнению) строить ограничение вида

$$f_k = \sum_{j \notin \text{базису}} f_{kj} x_j - S^*, \quad S^* \geq 0, \text{ (эквивалент } \sum_{j \notin \text{базису}} f_{kj} x_j \geq f_k \text{)} \qquad (2)$$

где $f_k = x_j - [x_j]$; $f_{kj} = z_{kj} - [z_{kj}]$; S^* – новая переменная; $[x_j]$, $[z_{kj}]$ - целые части соответствующих величин.

Обратите внимание на то, что [2.57] =2, но [– 2.57] = – 3 и соответственно значения f_{kj} равны 0.57 и 0.43.

Так, если выбранное уравнение имеет вид

$$3.14 = -2.57\, x_1 + x_2 + 3.12 x_3 - 4.12\, x_4$$

то получаем новое (дополнительное) ограничение

$$0.14 = 0.43 x_1 + 0.12\, x_3 + 0.88\, x_4 - S, \; S \geq 0,$$

которое присоединяем к имеющимся в симплексной таблице, тем самым получая расширенную задачу.

Поскольку для продолжения ее решения опять возникает вопрос о выборе начального опорного плана, то для включения в число базисных рекомендуем ту переменную, для которой величина $|\,\Delta_j / f_{kj}\,|$ минимальна. Если для этой переменной величина θ (см. симплексный метод) достигается по дополнительной строке, то эту строку (уравнение) разрешаем относительно выбранной переменной и исключаем эту переменную из остальных уравнений (получен опорный план для расширенной задачи).

Если же величина θ не соответствует дополнительной строке, то приходится вводить искусственную переменную и надеяться, что в дальнейшем эта переменная покинет базис (этого не случится, если новое множество планов окажется пустым – новая система ограничений противоречива).

Упомянутые действия повторяются до получения целочисленного решения или установления неразрешимости задачи.

Существует доказательство конечности этого процесса, но оценить заранее число таких шагов невозможно. Все зависит от количества нецелочисленных планов задачи и ее размерности (хотя многие задачи даже малой размерности решаются весьма долго).

Необходимо сделать некоторые замечания.

1) Если для дробного x_j обнаружится целочисленность всех коэффициентов соответствующего уравнения, то задача не имеет целочисленного решения.

2) Если дополнительная переменная S^* вошла в число базисных (появились более жесткие условия, ранее введенное становится лишним), то соответствующие ей строку и столбец можно удалить. Соблюдение этого правила сохраняет размерность решаемой задачи в разумных пределах – число уравнений не превысит $m+n$ – суммы числа уравнений и переменных в исходной задаче).

Рассмотрим высказанные соображения на несложном примере (здесь мы ограничились двумя переменными, чтобы иметь возможность геометрической интерпретации).

3.3. Пример решения задач методом Гомори

Пусть для приобретения нового оборудования предприятие выделяет 19 ден. ед. Оборудование должно быть размещено на площади, не превышающей 16 м^2. Предприятие может заказать оборудование двух видов: машины типа «А» стоимостью 2 ден.ед., требующие производственную площадь 4 м^2 и обеспечивающие производительность за смену 8 т продукции, и машины типа «В» стоимостью 5 ден. ед., занимающие площадь 1 м^2 и обеспечивающие производительность за смену 6 т продукции.

Требуется составить оптимальный план приобретения оборудования, обеспечивающий максимальную общую производительность.

Обозначив через x_1, x_2 количество приобретаемых машин соответственно типа «А» и «В» и через L – их общую производительность, получаем математическая модель задачи:

$$max\ L = 8x_1 + 6x_2$$

при ограничениях:

$$2x_1 + 5x_2 \le 19$$
$$4x_1 + \ \ x_2 \le 16$$
$$x_1 \ge 0,\ \ x_2 \ge 0$$
$$x_1, x_2 \ -\textit{целые числа}$$

Решаем задачу симплексным методом без учета целочисленности (предполагаем, что читатель освоился с технологией этого метода).

C_0	$Б_0$	X_0	8	6	0	0
			x_1	x_2	x_3	x_4
0	x_3	19	2	5	1	0
0	x_4	16	4	1	0	1
z_j		0	0	0	0	0
Δ_j			–8	–6	0	0

C_1	$Б_1$	X_1	8	6	0	0
			x_1	x_2	x_3	x_4
0	x_3	11	0	9/2	1	–1/2
8	x_1	4	1	1/4	0	1/4
z_j		32	8	2	0	2
Δ_j			0	–4	0	2

C_2	$Б_2$	X_2	8	6	0	0

			x_1	x_2	x_3	x_4
6	x_2	22/9	0	1	2/9	–1/9
8	x_1	61/18	1	0	–1/18	5/18
Δ_j		376/9	0	0	8/9	14/9

Получен оптимальный, но нецелочисленный план

$$X_{опт} = (61/18;\ 22/9),\ \ L_{max} = 376/9.$$

Так как максимальной дробной частью обладает компонента плана x_2, относительно которой разрешено первое уравнение (4/9 > 7/18), дополнительное ограничение записываем по первой строке, получая

$4/9 = 2/9x_3 + 8/9x_4 – S_1,\ \ S_1 \geq 0$ – *первое ограничение Гомори*

(подставьте найденный оптимальный план в это условие и убедитесь в его «отсечении»). Составленное уравнение дописываем к уже имеющимся в симплексной таблице.

C_2	$Б_2$	X_2	8	6	0	0	0
			x_1	x_2	x_3	x_4	S_1
6	x_2	22/9	о	1	2/9	–1/9	0
8	x_1	61/18	1	о	–1/18	5/18	0
Δ_j		376/9	0	0	8/9	14/9	·
		4/9	0	0	2/9	8/9	–1

Теперь имеем новую задачу линейного программирования, в которой 3 ограничения на 5 неотрицательных переменных. Для получения опорного плана этой задачи необходимо найти третью базисную переменную. Для этого определяем: $min\dfrac{\Delta_j}{f_{kj}} = min\left(\dfrac{8/9}{2/9};\dfrac{14/9}{8/9}\right)$ =7/4 и предлагаем для ввода в базис переменную x_4.

Отыскав величину $\theta = min\left(-\cdot\dfrac{61/18}{5/18},\dfrac{4/9}{8/9}\right)$ и обнаружив ее соответствие дополнительной строке (третьему уравнению), не прибегая к искусственной переменной, выражаем x_4 из этого уравнения и получаем опорный план расширенной задачи.

C_3	$Б_3$	X_3	8	6	0	0	0	0
			x_1	x_2	x_3	x_4	S_1	S_2
6	x_2	5/2	0	1	1/4	0	–1/8	0
8	x_1	13/4	1	0	–1/8	0	5/16	0
0	x_4	1/2	0	0	1/4	1	–9/8	0
Δ_j		41	0	0	1/2	0	7/4	·
		1/2	0	0	1/4	0	7/8	–1

Найденный план оптимален, но нецелочисленный. Строим новое ограничение Гомори.

Так как максимальная дробная часть среди компонент плана равна 1/2, записываем дополнительное ограничение по первой строке (можно и по третьей).

$1/2 = 1/4x_3 + 7/8S_1 - S_2,\ S_2 \geq 0$ – - *второе ограничение Гомори*

Это ограничение добавляем в последнюю симплексную таблицу и получаем задачу с 4 ограничениями.

Определяя переменную, вводимую в базис, из совпадения $min\left(\frac{1/2}{1/4};\ \frac{7/4}{7/8}\right) = 2$ видим свободу выбора между x_3 и S_1. Выбрав S_1 и обнаружив соответствие $\theta = min\left(-;\frac{13/4}{5/16};-;\ \frac{1/2}{7/8}\right) = 4/7$ дополнительному ограничению, получаем

C_4	$Б_4$	X_4	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	S_1	S_2	S_3
6	x_2	18/7	0	1	2/7	0	о	–1/7	0
8	x_1	43/14	1	0	–3/14	0	о	5/14	0
0	x_4	8/7	0	0	4/7	1	о	–9/7	0
0	S_1	4/7	0	0	0	0	1	–8/7	0
Δ_j		40	0	0	0	0	0	2	·
		4/7	о	о	2/7	о	·	6/7	–1

Получаем новый оптимальный нецелочисленный план. Учитывая сделанное выше замечание 2, вычеркиваем строку и столбец, соответствующие переменной S_1.

В полученном плане максимальную дробную часть имеет компонента x_2, поэтому записываем дополнительное ограничение по первой строке :

$4/7 = 2/7x_3 + 6/7S_2 - S_3,\ S_3 \geq 0$ – *третье ограничение Гомори.*

Отыскав $min\left(\frac{0}{2/7};\ \frac{2}{6/7}\right) = 0$, вводим в базис переменную x_3. Минимальное значение θ = 2, что соответствует дополнительной строке, и после очередных симплексных преобразований получаем:

C_5	$Б_5$	X_5	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	S_2	S_3	S_4
6	x_2	2	0	1	0	0	–1	1	0
8	x_1	7/2	1	0	0	0	1	–3/4	0
0	x_4	0	0	0	0	1	–3	2	0
0	x_3	2	0	0	1	0	3	–7/2	0
Δ_j		40	0	0	0	0	2	0	·
		1/2	0	0	0	0	0	1/4	–1

Обнаружив оптимальность, но нецелочисленность найденного плана, строим

$1/2 = 1/4S_3 - S_4$, $S_4 \geq 0$ — *четвертое ограничение Гомори.*

Выявив намерение ввести в базис переменную S_3, обнаруживаем, что $\theta = min\left(\frac{2}{1}; -;\ \frac{0}{2}; -; \frac{1/2}{1/4}\right) = 0$ соответствует не дополнительному уравнению, а третьему.

К сожалению приходится прибегнуть к искусственной переменной (обозначаем ее x_5)

C_5'	$Б_5'$	X_5'	8	6	0	0	0	0	0	–M
			x_1	x_2	x_3	x_4	S_2	S_3	S_4	x_5
6	x_2	2	0	1	0	0	–1	1	0	0
8	x_1	7/2	1	0	0	0	1	–3/4	0	0
о	x_4	0	0	0	0	1	–3	**2**	0	0
о	x_3	2	0	0	1	0	3	–7/2	0	0
–M	x_5	1/2	0	0	0	0	0	1/4	–1	1
Δ_j		40–M/2	0	0	0	0	2	–M/4	M	0

C_6	$Б_6$	X_6	8	6	0	0	0	0	0	–M
			x_1	x_2	x_3	x_4	S_2	S_3	S_4	x_5
6	x_2	2	0	1	о	–1/2	1/2	0	0	0
8	x_1	7/2	1	0	0	3/8	–1/8	0	0	0
о	S_3	0	0	0	0	1/2	–3/2	1	0	0
о	x_3	2	0	0	1	7/4	–9/4	0	0	0
-M	x_5	1/2	0	0	0	–1/8	**3/8**	0	–1	1
Δ_j		40- M/2	0	0	0	M/8	2–3M/8	0	M	0

C_7	$Б_7$	X_7	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	S_2	S_4	S_5
6	x_2	4/3	0	1	о	–1/3	0	4/3	0
8	x_1	11/3	1	0	0	1/3	0	–1/3	0
о	x_3	5	0	0	1	1	0	–6	0
о	S_2	4/3	0	0	0	–1/3	1	–8/3	0
Δ_j		112/3	0	0	0	2/3	0	16/3	·
		2/3	0	0	0	1/3	0	2/3	–1

И опять нецелочисленный оптимальный план. На основе второго уравнения строим

$2/3 = 1/3x_4 + 2/3S_4 - S_5$, $S_5 \geq 0$ *– пятое ограничение Гомори.*

Выбрав для ввода в базис переменную x_4 и обнаружив соответствие $\theta = min\left(-;\ \frac{11/3}{1/3};\ \frac{5}{1};\ \frac{2/3}{1/3}\right) = 2$ новому уравнению, получаем

C_8	$Б_8$	X_8	8	6	0	0	0	0
			x_1	x_2	x_3	x_4	S_4	S_5
6	x_2	2	0	1	0	0	2	–1
8	x_1	3	1	0	0	0	–1	1
0	x_3	3	0	0	1	0	–8	3
0	x_4	2	0	0	0	1	2	–3
Δ_j		36	0	0	0	0	4	2

Получен оптимальный целочисленный план $X_8 = (3, 2, 3, 2)$; $L_{max} = 36$.

В чем же результат такого длительного решения?

Экономическая интерпретация: согласно полученному решению, предприятию необходимо закупить 3 машины типа «А» и 2 машины типа «В». При этом будет достигнута максимальная производительность работы оборудования, равная 36 т продукции за смену. Полученную экономию денежных средств в размере $x_3 = 3$ ден. ед. можно будет направить на какие-либо иные цели, например, на премирование рабочих, которые будут заниматься отладкой полученного оборудования. На излишнюю площадь $x_4 = 2$ м^2 можно поставить ящик с цветами, если рабочие не страдают аллергией на цветы.

Геометрическую интерпретацию процедуры выполненных отсечений начинаем с построения множества планов (см. рисунок). Точка 1 определяет оптимальный нецелочисленный план.

Из условий задачи имеем:

$$x_3 = 19 - 2x_1 - 5x_2 \ ; \quad x_4 = 16 - 4x_1 - x_2 \ .$$

Подставляем эти выражения в первое ограничение Гомори

$$2/9x_3 + 8/9x_4 - S_1 = 4/9, \ S_1 \geq 0$$

и после преобразований получаем $S_1 = 18 - 4x_1 - 2x_2 \geq 0$.

Полученное ограничение $4x_1 + 2x_2 \leq 18$ отсекает от множества планов область, содержащую точку 1. Новый оптимальный нецелочисленный план – точка 2.

Подставляя выражения $x_3 = 19 - 2x_1 - 5x_2$ и $S_1 = 18 - 4x_1 - 2x_2$ во второе ограничение Гомори :

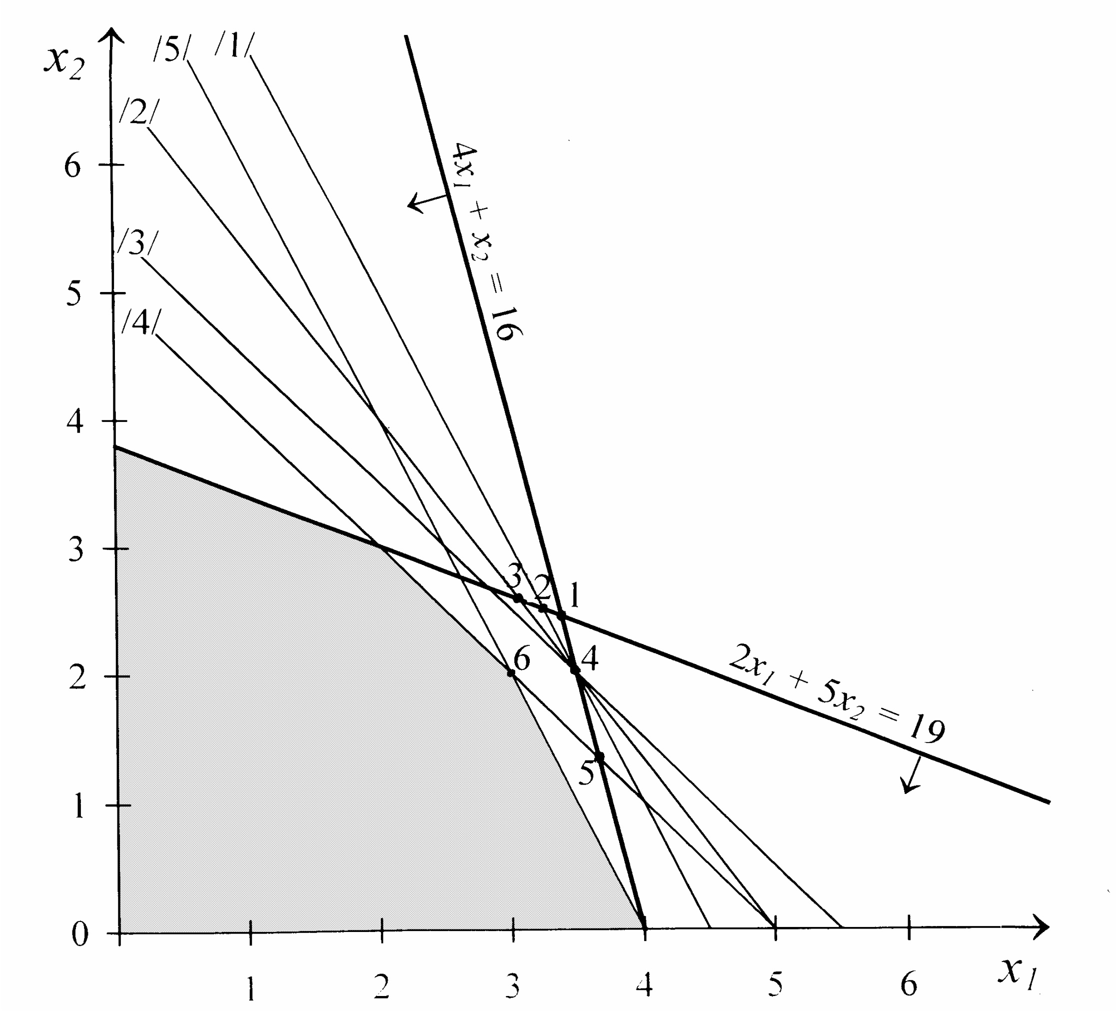
$$1/4x_3 + 7/8\,S_1 - S_2 = 1/2, \quad S_2 \geq 0.$$

Получаем $S_2 = 20 - 4x_1 - 3x_2 \geq 0$ ($4x_1 + 3x_2 \leq 20$). Это ограничение отсекает от множества планов область, содержащую точку 2. Новый оптимальный нецелочисленный план – точка 3.

Аналогичные подстановки в третье ограничение Гомори :

$$2/7x_3 + 6/7S_2 - S_3 = 4/7, \quad S_3 \geq 0$$

дают $S_3 = 22 - 4x_1 - 4x_2 \geq 0$. На рисунке видим, что условие $4x_1 + 4x_2 \leq 22$ отсекает от множества планов область, содержащую точку 3. Новый оптимальный нецелочисленный план - точка 4.



Подстановки в четвертое ограничение Гомори :

$$1/4\ S_3 - S_4 = 1/2, S_4 \geq 0$$

дают $S_4 = 5 - x_1 - x_2 \geq 0$, и наблюдаем, как условие $\boldsymbol{x_1 + x_2 \leq 5}$ делает недопустимой точку 4. Новый оптимальный нецелочисленный план – точка 5.

Наконец, пятое ограничение Гомори :

$$1/3\ x_4 + 2/3\ S_4 - S_5 = 2/3, \ S_5 \geq 0$$

преобразуем к эквивалентному виду $S_5 = 8 - 2x_1 - x_2 \geq 0$ и наблюдаем, как условие $2x_1 + x_2 \leq 8$ отсекает от множества планов область, содержащую точку 5.

В итоге устанавливаем оптимальность и целочисленность плана –

точки 6 с координатами (3, 2).

3.4. Метод ветвей и границ

Этот метод с красивым названием идеологически гораздо проще метода Гомори, хотя и использует идею отсечения оптимального нецелочисленного плана.

Получив нецелочисленный оптимальный план задачи, в котором какая-то составляющая x_k оказалась нецелочисленной (x_k =A), мы ставим две задачи на основе ограничений решенной задачи и дополнительных условий $x_k \le$ [A] и $x_k \ge$ [A] + 1 соответственно. Например, если в найденном оптимальном плане x_3 =4.7 , то в новых задачах будут условия $x_3 \le 4$ и $x_3 \ge 5$ (недопустимость найденного плана в будущем очевидна). Каждую из полученных задач решаем до получения оптимального плана и нового разветвления.

В общем случае процесс ветвления продолжается до обнаружения противоречивости ограничений или получения целочисленных решений, из которых остается выбрать наилучшее.

В задачах небольшой размерности или при малом числе вариантов значений переменных метод вполне эффективен и часто используется в приложениях. Но компьютерная его реализация едва ли доставит удовольствие программисту, а использование разработок других ненадежных авторов чревато возможностью несоответствия рекламы и факта.

Для иллюстрации метода возьмем пример из предыдущего раздела:

$$max\ L = 8x_1 + 6x_2$$

при ограничениях:

$$2x_1 + 5x_2 \le 19$$
$$4x_1 + \ \ x_2 \le 16$$
$$x_1 \ge 0, x_2 \ge 0 \ - \ \textit{целые числа}$$

Решая эту задачу симплексным методом, мы получаем оптимальный план $X_{опт} = (61/18; 22/9)$, $L_{max} = 376/9 = 41\frac{7}{9}$.

Если вам не по душе нецелочисленность x_1,, ставим две задачи:

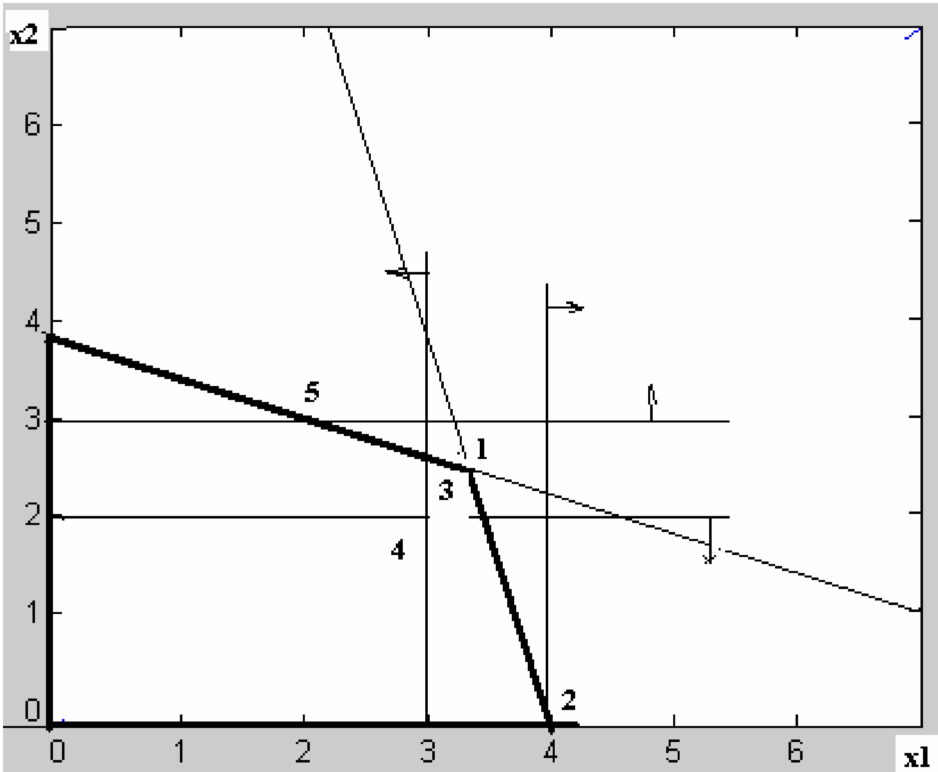
A1	A2
$max\ L = 8x_1 + 6x_2$	$max\ L = 8x_1 + 6x_2$
при ограничениях:	при ограничениях:
$2x_1 + 5x_2 \le 19$	$2x_1 + 5x_2 \le 19$
$4x_1 + \ x_2 \le 16$	$4x_1 + \ x_2 \le 16$
$x_1 \le 3$	$x_1 \ge 4$
$x_1 \ge 0, x_2 \ge 0$	$x_1 \ge 0, x_2 \ge 0$

x_1, x_2 – *целые числа*	x_1, x_2 – *целые числа*

Задача А2 обладает (повезло !) единственным планом (4,0), для которого $L_{max} = 32$.

Задача А1 решается до получения оптимального плана (3, 13/5) и появляются две ветви:

А11	А12
$max\ L = 8x_1 + 6x_2$	$max\ L = 8x_1 + 6x_2$
при ограничениях:	при ограничениях:
$2x_1 + 5x_2 \le 19$	$2x_1 + 5x_2 \le 19$
$4x_1 + x_2 \le 16$	$4x_1 + x_2 \le 16$
$x_1 \le 3,\ x_2 \le 2$	$x_1 \le 3,\ x_2 \ge 3$
$x_1 \ge 0,\ x_2 \ge 0$ - *целые числа*	$x_1 \ge 0,\ x_2 \ge 0$ - *целые числа*



На приведенном рисунке точка 1 определяет решение исходной задачи, точки 3 и 2 соответственно – решение задач А1 и А2, точки 4 и 5 – решение задач А11 и А12.

Заметим, что существует много задач, где требование целочисленности накладывается не на все переменные. В этом случае ограничение Гомори немного модифицируется

$$f_k = \sum_{j \notin Б} f_{kj} x_j - S^*\ ,\ \ S^* \ge 0 \qquad (3)$$

где

$$f_{kj}=\begin{cases} A_{kj} & j\notin Y, \quad A_{kj}\ge 0\\ -f_k/(1-f_k)/A_{kj} & j\notin Y, \quad A_{kj}<0\\ f_{kj} & j\in Y, \quad f_{kj}\le f_k\\ -f_k/(1-f_k)/(1-f_{k\,j}) & j\notin Y, \quad f_{kj}>f_k\end{cases}$$

(здесь A_{kj} – коэффициенты выбранного уравнения, Y – множество целочисленных переменных), модификация метода ветвей и границ не нуждается в комментариях.

Например, решая задачу минимизации линейной формы

$$2X_1+2X_2+5X_3$$

при условиях :

$$\begin{aligned} 3X_1+2X_2+X_3&\ge 5\\ -X_1+X_2+4X_3&\ge 7\\ X_1-X_2+2X_3&\ge 4\\ X_1,X_2,X_3&\ge 0\\ X_1,X_2&- \text{ целые}\end{aligned}$$

мы приходим к оптимальному плану, определяемому таблицей вида

C баз	Баз 4	План X	2 X_1	2 X_2	5 X_3	0 X_4	0 X_5	0 X_6
2	X_2	13/30	0	1	0	–1/5	–1/6	13/30
5	X_3	55/30	0	0	1	0	–1/6	–1/6
2	X_1	23/30	1	0	0	–1/5	1/6	–7/30
Z_j		7.9	2	2	5	–4/5	–5/6	–13/30
Δ_j			0	0	0	–4/5	–5/6	–13/30

Так как наибольшую дробную часть имеет (среди целочисленных по условиям задачи) компонента X_1, то выбираем третье уравнение и строим дополнительное ограничение в виде

$$\frac{23}{30}=-\frac{23/30}{1-23/30}\left(-\frac{1}{5}\right)X_4+\frac{1}{6}X_5-\frac{23/30}{1-23/30}\left(-\frac{7}{30}\right)X_6-S_1\,,\ S_1\ge 0,$$

т. е.

$$\frac{23}{30}=\frac{23}{35}X_4+\frac{1}{6}X_5+\frac{23}{30}X_6-S_1\quad,\quad S_1\ge 0\ .$$

3.5. Задачи, приводимые к целочисленным

Как мы уже отмечали ранее, существует ряд задач, по своей природе не требующих целочисленности искомого плана, но решаемых, в частности, методами целочисленного линейного программирования. К ним относятся задачи с дискретным множеством планов, задачи с разрывной целевой функцией, задачи с кусочно-линейными функциями, задачи с относительным произволом при выборе ограничений и другие.

Так ситуацию, где некоторая переменная X_k может принимать конечное множество конкретных значений $T_1, T_2, \ldots, T_m$ заменой

$$X_k = \alpha_1 T_1 + \alpha_2 T_2 + \ldots + \alpha_m T_m \ ,$$

где $\alpha_1 + \alpha_2 + \ldots + \alpha_m = 1$, $\alpha_j \ge 0$ – целые (j=1 .. m), приводим к целочисленной, линейной по α_j программе (одно из значений α_j равно 1, а остальные равны нулю).

Задачу с разрывной целевой функцией

$$C(X_k) = \begin{cases} A + BX_k \text{ при } X_k > 0 \\ \quad 0 \qquad \text{при } X_k = 0 \end{cases}$$

путем замены $C(X_k) = \alpha A + BX_k$, где $0 \le \alpha \le 1$ – целое, $X_k \le M \cdot \alpha$, M – верхняя граница для X_k, сводим к целочисленной линейной программе с α =0 (X_k=0) или с α =1 (X_k>0).

В ряде задач не требуется одновременного выполнения всех ограничений (требование И заменяется на ИЛИ). Пусть, например, стоит задача поиска экстремума $L(X)$ при условиях неотрицательности X и справедливости хотя бы одного из m условий

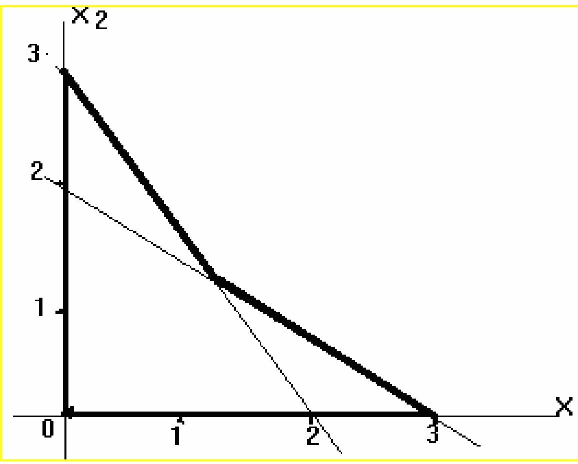
$$\sum_{j=1}^{n} A_{ij} X_j \le B_i \ , \quad i = 1,..,m \ .$$

Если найти верхние оценки M_1, M_2, ..., M_m для значений $\sum_{j=1}^{n} A_{ij} X_j - B_i$ и ограничения задачи записать в форме:

$$\sum_{j=1}^{n} A_{ij} X_j - B_i - (1 - \alpha_i) M_i \ \le \ 0, \quad i = 1,..,m$$

$$\alpha_1 + \alpha_2 + \ldots + \alpha_m = 1 \ , \ \alpha_i \ge 0 \text{ – целые},$$

то получение некоторого α_k =1 дает выполнение k-го ограничения исходной задачи.



Так для задачи с ограничениями

$$2\,X_1 + \ 3\,X_2 \le 6$$

или

$$3\,X_1 + \ 2\,X_2 \ \le \ 6 \ ,$$

$$X_1 \ \ge \ 0 \ , \ X_2 \ \ge \ 0$$

множество планов не выпукло и традиционная симплексная процедура недопустима. Заменой

$$2X_1 + 3X_2 - 6 - M\alpha \le 0$$

$$3X_1 + 2X_2 - 6 - M(1-\alpha) \le 0,$$

где $0 \le \alpha \le 1$ – целое, M > 6, получаем обычную частично целочисленную линейную программу.

4. ЗАДАЧИ ТРАНСПОРТНОГО ТИПА

Сам термин «транспортная задача» уже говорит о необходимости решения проблемы, возникающей при организации перемещения тех или иных продуктов с помощью некоторых транспортных средств (транспортировка угля из угольных центров Кузбасса по существующей сети железных дорог к потребителям, доставка ночной выпечки от хлебокомбината до булочных города, перекачка нефти (газа, воды) по трубопроводам от источников к потребителям и др.).

Решение подобных задач может преследовать самые разные цели. Это может быть желание проехать *кратчайшим путем* на автомашине от «солнечного Магадана» до «Гренадской волости в Испании» или в *кратчайший срок* доставить отопительные батареи из Новокузнецка и Магнитогорска до Приморья, куда «неожиданно» в декабре пришла зима. Это может быть поставка грузов с *минимальными затратами* на эту процедуру или организация полетов бомбардировщиков от аэродромов для *наилучшего подавления* каких-либо целей. Это может быть выбор для назначения на вакантные должности среди множества претендентов с тем, чтобы достичь *наибольшего общего эффекта* назначении (разумеется, мы сможем решить эту задачу лишь при наличии независимой экспертизы соответствия каждого претендента каждой из должностей).

В таких условиях возникает естественный вопрос: «***от кого, кому и сколько***», чтобы, как говорила пресса в недавние времена, были «обеспечены жизненные потребности трудящихся». Поиск ответа на этот вопрос (поиск какого-нибудь плана перевозок) едва ли вызовет особые затруднения. Однако у нас может возникнуть естественное желание, чтобы найденный план отвечал одному из двух критериев – минимуму денежных затрат или возможности реализации в кратчайшие сроки, что плохо сочетается (весьма редко сочетается «дешево и быстро», ибо за скорость надо доплачивать, а при плохой организации получается подчас «и медленно, и дорого»).

Некоторые задачи, например, поиска кратчайшего или самого длинного пути между некоторыми пунктами сети хороших автодорог по какому-либо критерию, обычно решаются достаточно просто на базе метода динамического программирования.

Расчеты, связанные с течением жидкостей в сети трубопроводов, уже несопоставимо сложнее и, наряду со стремлением к оптимизации, требуют учета физики процесса и решения соответствующих уравнений математической физики.

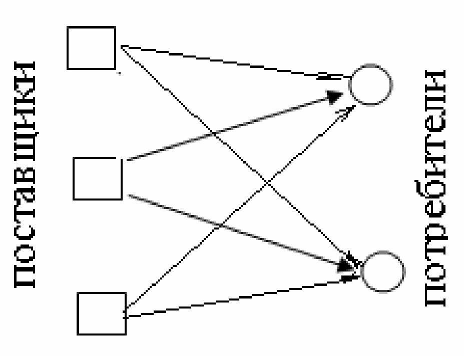
Несмотря на внешнюю простоту постановки многих задач транспортировки привычных грузов, их решение часто исключительно сложное. Так значительные сложности возникают при наличии каких-то промежуточных пунктов в транспортной сети или ограничений на «объемы» поставок, но тем не менее для подобных задач существуют универсальные методы решения (например, венгерский метод).

Несоизмеримо сложнее решать т. н. многопродуктовые задачи, где в одной автомашине везут ящики с пивом, мешки с мукой и садовый инструмент, причем для разных получателей.

4.1. Классическая транспортная задача

4.1.1. Постановка задачи и свойства решений

Здесь предполагается наличие поставщиков (производителей, источников) некоторой продукции и заинтересованных в ней ее потребителей. Примечательно, что поставки происходят напрямую без промежуточных пунктов (как показано на рисунке).



Пусть имеется m поставщиков некоторого продукта в количествах A_i (i = 1 .. m) и n потребителей этого продукта в количествах B_j (j = 1 .. n). Пусть при этом соблюдается условие баланса

$$\sum_{i=1}^{m} A_i = \sum_{j=1}^{n} B_j \tag{1}$$

(т. н. *закрытая модель* производства-потребления).

Пусть известны стоимость C_{ij} перевозки единицы продукта от i-го поставщика к j-му потребителю. Если обозначить через X_{ij} соответствующие объемы перевозок, то задача сведется к нахождению плана перевозок, минимизирующего суммарную их стоимость:

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} \tag{2}$$

при условиях

$$\sum_{j=1}^{n} X_{ij} = A_i \qquad (i = 1 .. m); \qquad (3)$$

$$\sum_{i=1}^{m} X_{ij} = B_j \qquad (j= 1 .. n); \qquad (4)$$

$$X_{ij} \geq 0 \qquad (i = 1 .. m \ ; \ j= 1 .. n) . \qquad (5)$$

Полученная задача представляет частный случай общей линейной программы с $m + n + mn$ ограничениями на mn переменных. Однако едва ли вас соблазнит перспектива решения задачи ЛП с 10 ограничениями и 24 неотрицательными переменными (при $m = 4$, $n = 6$) прямым алгоритмом симплексного метода.

Убедимся, что эта задача не из числа тех, решение которых требует гения А. Эйнштейна или М. В. Келдыша (читателю известны эти имена?), но едва ли элементарна для выпускника начальной школы.

Отметим некоторые «приятные» особенности этой задачи, выделяющие ее среди произвольных линейных программ.

1. Транспортная задача имеет непротиворечивые ограничения.

В подтверждение этого заявления достаточно взять набор значений

$$X_{ij} = \frac{A_i B_j}{\sum_{i=1}^{m} A_i} \quad (i = 1 .. m \ ; \ j= 1 .. n) \qquad (6)$$

и подстановкой в (3)-(5) с учетом (1) убедиться, что он является планом (удовлетворяет ограничениям).

2. Линейная форма транспортной задачи ограничена как сверху, так и снизу, что является следствием ограниченности множества планов

$$0 \leq X_{ij} \leq \ min\,(A_i \,, B_j \,) \,.$$

Отсюда следует, что *транспортная задача имеет оптимальный план*.

3. Поскольку равенства (3)–(4) с учетом (1) линейно зависимы достаточно по отдельности их сложить и следствием получить (1), то одним из них можно пренебречь. Следовательно, *опорный план транспортной задачи имеет не более $m + n - 1$ положительных компонент* (если их число равно $m + n - 1$, то его называют *невырожденным*).

4. Приятной *особенностью опорных планов транспортной задачи является их целочисленность при целых значениях A_i и B_j* (на всех этапах симплексного преобразования коэффициенты при неизвестных равны 1, –1 или 0).

5. Если условие баланса (1) нарушено (*открытая модель*), то задачу можно свести к закрытой модели вводом фиктивного (воображаемого) m + 1-го поставщика или фиктивного n+1-го потребителя с объемом поставки (потребления) соответственно $\sum_{j=1}^{n} B_j - \sum_{i=1}^{m} A_i > 0$

или $\sum_{i=1}^{m} A_i - \sum_{j=1}^{n} B_j > 0$ с нулевыми стоимостями перевозок.

Если обозначить через U_i (i = 1 .. m) и V_j (j = 1 .. n) двойственные переменные, сопряженная задача будет состоять в максимизации

$$L(U,V) = \sum_{i=1}^{m} A_i U_i + \sum_{j=1}^{n} B_j V_j \tag{7}$$

при условиях

$$U_i + V_j \le C_{ij} \quad (i = 1 .. m ,\ j = 1 .. n). \tag{8}$$

В этом вы можете убедиться, если поставите задачу, сопряженную задаче минимизации (m = 2, n = 3)

$$C_{11}X_{11} + C_{12}X_{12} + C_{13}X_{13} + C_{21}X_{21} + C_{22}X_{22} + C_{23}X_{23}$$

при

$$\begin{array}{cccccccc|c}
X_{11}+ & X_{12}+ & X_{13} & & & & = & A_1 & U_1 \\
 & & & X_{21}+ & X_{22}+ & X_{23} & = & A_2 & U_2 \\
X_{11}+ & & & X_{21} & & & = & B_1 & V_1 \\
 & X_{12}+ & & & X_{22} & & = & B_2 & V_2 \\
 & & X_{13}+ & & & X_{23} & = & B_3 & V_3 \\
X_{11}, & X_{12}, & X_{13}, & X_{21}, & X_{22}, & X_{23} & \ge & 0 &
\end{array}$$

Заметим, что многие задачи можно описать в терминах классической транспортной задачи или ее аналога с требованием максимума «затрат».

В последнем случае сопряженная задача будет состоять в минимизации (7) при условиях $U_i + V_j \ge C_{ij}$ (i =1 .. m , j =1 .. n).

4.1.2. Выбор начального опорного плана

Идея всех методов поиска начального опорного плана заключается в установке максимального возможного объема перевозки по маршруту, т. е. минимума из предложения и спроса $X_{ij} = min(A_i , B_j)$.

Самым простейшим из методов поиска начального опорного плана является "*метод северо-западного угла*", отправной точкой для которого является левый верхний угол таблицы.

Пусть m= 3, A=(17, 33, 20); n= 5, B=(14, 14, 14, 14, 14). Так как баланс здесь соблюдается, приступаем к поиску начального опорного плана в табличной форме (для удобства перенесем в "оцифровку" таблицы значения A, B). Имеем $X_{11} = min$ (17, 14) = 14, $X_{21} = X_{31}$ =0.

Продолжаем поиск с угла незаполненной части таблицы:

$$X^0 = \begin{array}{|ccccc|c|} 14 & 3 & \circ & \circ & \circ & 17 \\ \circ & 11 & 14 & 8 & \circ & 33 \\ \circ & \circ & \circ & 6 & 4 & 20 \\ \hline 14 & 14 & 14 & 14 & 14 & B\backslash A \end{array}$$

$X_{12} = min\,(17–14, 14) = 3$, $X_{13} = X_{14} = X_{15} = 0$, $X_{22} = min(33, 14–3) = 11$, $X_{32} = 0$, $X_{23} = min\,(33–11, 14) = 14$, $X_{33} = 0$,

$X_{24} = min(33–11–14, 14) = 8$, $X_{25} = 0$; $X_{34} = 14 - 8 = 6$, $X_{35} = 20 - 6 = 14$.

Можно придумать и другие модификации этого метода. В частности, *метод минимального элемента матрицы стоимостей* определяется правилом первоочередного выбора перевозки на самом "дешевом" маршруте.

Пусть значения A, B те же, что и в рассмотренном примере, а стоимости перевозок определяются матрицей C. Так как минимальная стоимость соответствует C_{22}, то начинаем поиск компонент опорного плана с элемента $X_{22} = \min(33, 14) = 14$, $X_{12} = X_{32} = 0$; затем

$$C = \begin{array}{|ccccc|} 6 & 4 & 2 & 4 & 6 \\ 5 & 1 & 6 & 8 & 5 \\ 8 & 5 & 7 & 2 & 3 \end{array}$$

$X_{13} = min\,(17, 14) = 14$, $X_{23} = X_{33} = 0$; $X_{34} = min(20, 14) = 14$, $X_{14} = X_{24} = 0$;

$$X^0 = \begin{array}{|ccccc|c|} \circ & \circ & 14 & \circ & 3 & 17 \\ 14 & 14 & \circ & \circ & 5 & 33 \\ \circ & \circ & \circ & 14 & 6 & 20 \\ \hline 14 & 14 & 14 & 14 & 14 & B\backslash A \end{array}$$

$X_{35} = min(20–14, 14) = 6$, $X_{31} = 0$; $X_{21} = min(33–14, 14) = 14$, $X_{11} = 0$;

наконец, $X_{15} = 17 - 14 = 3$; $X_{25} = 33 - 14 - 14 = 5$.

Какой из найденных планов ближе к оптимальному? Ответить на этот вопрос можно только вычислив суммарную стоимость перевозок.

Какой из методов поиска опорного плана лучше? Однозначного ответа нет: при машинной реализации метод северо-западного угла проще по своему алгоритму и имеет ряд преимуществ, связанных с проблемой вырожденности. При ручном решении в большинстве случаев предпочтение следует отдать методу минимального элемента матрицы.

Заметьте, что оба найденные планы являются невырожденными (содержат $m + n - 1 = 3 + 5 - 1$ положительных компонент).

При решении задач симплексным методом мы автоматически выделяли систему базисных векторов и базисных компонент плана. При поиске опорного плана транспортной задачи базисные нулевые компоненты в случае вырожденного плана теряются среди небазисных. Чтобы этого не происходило, в дальнейшем небазисные компоненты плана будем выделять «точками» вместо нулей.

Пусть $m = n = 4$, все A_i, B_j равны 1; дана матрица стоимостей C и найден по минимальной стоимости опорный план X. Среди нулевых компонент

$$C = \begin{pmatrix} 5 & 4 & 6 & 5 \\ 3 & 2 & 3 & 4 \\ 2 & 3 & 4 & 3 \\ 1 & 5 & 2 & 2 \end{pmatrix} \qquad X = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

надо выделить для включения в число базисных три компоненты так, чтобы не возникало замкнутой цепочки по базису (это требование возникает из условия независимости системы базисных векторов). Так из приведенных ниже вариантов приемлем только четвертый.

0	**0**	1	
	1		
			1
1	**0**		

		1	
	1		**0**
	0		**1**
1			0

		1	
	1		
0			**1**
1		0	**0**

		1	
	1	0	
0			1
1	0		

Чтобы при выборе начального опорного плана в «большой» задаче не искать варианты без цепочек, используют ε-прием: увеличивают все A_i на ε (неотрицательное число, много меньшее значений A_i и B_j), а одно из значений B_j – на $m\varepsilon$, затем спокойно ищут начальный план упомянутыми ранее методами; далее при решении вручную берут ε =0 , выделив все базисные компоненты.

Так для приведенной задачи

$$C = \begin{array}{|cccc|} \hline 5 & 4 & 6 & 5 \\ 3 & 2 & 3 & 4 \\ 2 & 3 & 4 & 3 \\ 1 & 5 & 2 & 2 \\ \hline \end{array} \quad X = \begin{array}{|cccc|c} \hline & & 1-2\varepsilon & 3\varepsilon & 1+\varepsilon \\ & 1 & \varepsilon & & 1+\varepsilon \\ & & & 1+\varepsilon & 1+\varepsilon \\ 1 & & \varepsilon & & 1+\varepsilon \\ \hline 1 & 1 & 1 & 1+4\varepsilon & \end{array} \approx \begin{array}{|cccc|} \hline & & 1 & 0 \\ & 1 & 0 & \\ & & & 1 \\ 1 & & 0 & \\ \hline \end{array}$$

В заключение отметим, что в случае транспортной задачи максимизации поиск начального плана идет точно так же, но предпочтение отдается максимальным «стоимостям».

4.1.3. Метод Д. Данцига последовательного улучшения плана

Данный метод представляет компактную форму обычной симплексной процедуры.

Пусть найден некоторый начальный опорный план. Согласно второй теореме двойственности, этот план будет оптимальным, если для его базисных компонент $X_{ij} > 0$ условия сопряженной задачи выполняются в виде $U_i + V_j = C_{ij}$ и для остальных компонент – в виде $U_i + V_j \le C_{ij}$ или $U_i + V_j \ge C_{ij}$ (соответственно для транспортной задачи минимизации или максимизации).

Поэтому $m + n-1$ базисным компонентам выбранного плана сопоставляется система $m + n-1$ уравнений $U_i + V_j = C_{ij}$ с $m + n$

неизвестными, которая решается с точностью до константы (например, U_1=0). Если все $\Delta_{ij} = U_i + V_j - C_{ij} \le 0$, то выбранный план оптимален.

Если же нашлось $\Delta_{kp} > 0$, то план не оптимален и подвергается перестройке к плану с ненулевым значением соответствующей компоненты. Полагаем $X_{kp} = \theta > 0$ и ищем т. н. *минимизирующую цепочку по базису* так, чтобы новый набор значений X_{ij} удовлетворял требованиям предложения и спроса. Затем выбираем максимальное допустимое θ, сохраняя неотрицательность компонент нового плана.

Пример. Для предложенных условий отыскиваем начальный опорный план по минимальной стоимости:

$$C = \begin{array}{|cccc|c} 1 & 2 & 3 & 4 & 8 \\ 2 & 6 & 7 & 9 & 10 \\ 1 & 2 & 5 & 8 & 2 \\ \hline 9 & 3 & 4 & 4 & B \backslash A \end{array} \qquad X^o = \begin{array}{|cccc|c} 8 & & & & 8 \\ & 2 & 4 & 4 & 10 \\ 1 & 1 & & & 2 \\ \hline 9 & 3 & 4 & 4 & B \backslash A \end{array}$$

Для проверки этого невырожденного плана на оптимальность строим систему 6 уравнений с 7 неизвестными

$$U_1 + V_1 = 1\,, \quad U_2 + V_2 = 6\,, \quad U_2 + V_3 = 7\,,$$
$$U_2 + V_4 = 9\,, \quad U_3 + V_1 = 1\,, \quad U_3 + V_2 = 2\,,$$

которую можно решить при $U_1 = 0$, и затем находим $\Delta_{ij} = U_i + V_j - C_{ij}$.

Построение системы и ее решение удобнее выполнять в табличной форме (значения правых частей этой системы заносятся в таблицу, отыскиваются решения и их суммы).

$$U_i + V_j = \begin{array}{c|cccc} U \backslash V & 1 & 2 & 3 & 5 \\ \hline 0 & \mathbf{1} & 2 & 3 & 5 \\ 4 & 5 & \mathbf{6} & \mathbf{7} & \mathbf{9} \\ 0 & \mathbf{1} & \mathbf{2} & 3 & 5 \end{array} \qquad \Delta_{ij} = \begin{array}{|cccc|} \circ & 0 & 0 & 1 \\ 3 & \circ & \circ & \circ \\ \circ & \circ & -2 & -3 \end{array}$$

Так как $\Delta_{21} > 0$, то план X^0 не оптимален. Перестраиваем его, полагая $X_{21} = \theta$:

$$X^1 = \begin{array}{|cccc|} 8 & & & \\ \theta & 2-\theta & 4 & 4 \\ 1-\theta & 1+\theta & & \end{array} \;\overset{\theta=1}{=}\; \begin{array}{|cccc|} 8 & & & \\ 1 & 1 & 4 & 4 \\ & 2 & & \end{array}$$

Проверяем этот план на оптимальность:

$$U_i + V_j = \begin{array}{c|cccc} U \backslash V & 1 & 5 & 6 & 8 \\ \hline 0 & \mathbf{1} & 5 & 6 & 8 \\ 1 & 2 & \mathbf{6} & \mathbf{7} & \mathbf{9} \\ -3 & -2 & \mathbf{2} & 3 & 5 \end{array} \qquad \Delta_{ij} = \begin{array}{|cccc|} \circ & 3 & 3 & 4 \\ \circ & \circ & \circ & \circ \\ -3 & \circ & -2 & -3 \end{array}$$

Так как $\Delta_{14} > 0$, то план X^1 не оптимален. Перестраиваем его, полагая $X_{14} = \theta$:

$$\begin{array}{|cccc|} 8-\theta & & & \theta \end{array} \qquad \begin{array}{|cccc|} 4 & & & 4 \end{array}$$

$$X^2 = \begin{vmatrix} 1+\theta & 1 & 4 & 4-\theta \\ & 2 & & \end{vmatrix} \quad \theta = 4 \quad \begin{vmatrix} 5 & 1 & 4 & \\ & 2 & & \end{vmatrix}$$

Проверка плана X^2 опровергает факт его оптимальности из-за $\Delta_{12} > 0$.

$$U_i+V_j = \begin{array}{c|cccc} U\backslash V & 1 & 5 & 6 & 4 \\ \hline 0 & \mathbf{1} & 5 & 6 & \mathbf{4} \\ 1 & \mathbf{2} & \mathbf{6} & \mathbf{7} & 5 \\ -3 & -2 & \mathbf{2} & 3 & 1 \end{array} \qquad \Delta_{ij} = \begin{vmatrix} \circ & 3 & 3 & \circ \\ \circ & \circ & \circ & -4 \\ -3 & \circ & -2 & -7 \end{vmatrix}$$

Перестраиваем его, полагая $X_{12} = \theta$:

$$X^3 = \begin{vmatrix} 4-\theta & \theta & & 4 \\ 5+\theta & 1-\theta & 4 & \\ & 2 & & \end{vmatrix} \quad \theta=1 \; = \quad \begin{vmatrix} 3 & 1 & & 4 \\ 6 & & 4 & \\ & 2 & & \end{vmatrix}$$

Проверка плана X^3 опровергает факт его оптимальности из-за $\Delta_{14} > 0$.

$$U_i+V_j = \begin{array}{c|cccc} U\backslash V & 1 & 2 & 6 & 4 \\ \hline 0 & \mathbf{1} & \mathbf{2} & 6 & \mathbf{4} \\ 1 & \mathbf{2} & 3 & \mathbf{7} & 5 \\ 0 & 1 & \mathbf{2} & 6 & 4 \end{array} \qquad \Delta_{ij} = \begin{vmatrix} \circ & \circ & 3 & \circ \\ \circ & -3 & \circ & -4 \\ 0 & \circ & 1 & -4 \end{vmatrix}$$

Перестраиваем его, полагая $X_{13} = \theta$:

$$X^4 = \begin{vmatrix} 3-\theta & 1 & \theta & 4 \\ 6+\theta & & 4-\theta & \\ & 2 & & \end{vmatrix} \quad \theta=3 \; = \quad \begin{vmatrix} & 1 & 3 & 4 \\ 9 & & 1 & \\ & 2 & & \end{vmatrix}$$

Проверяем этот план на оптимальность:

$$U_i+V_j = \begin{array}{c|cccc} U\backslash V & -2 & 2 & 3 & 4 \\ \hline 0 & -2 & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ 4 & \mathbf{2} & 6 & \mathbf{7} & 8 \\ 0 & -2 & \mathbf{2} & 3 & 4 \end{array} \qquad \Delta_{ij} = \begin{vmatrix} -3 & \circ & \circ & \circ \\ \circ & 0 & \circ & -1 \\ -3 & \circ & -2 & -4 \end{vmatrix}$$

Так как теперь все $\Delta_{ij} \le 0$, то план X^4 оптимален, но существуют и другие оптимальные планы (для небазисного $X_{22} = 0$ имеем $\Delta_{22} = 0$, что дает возможность получения другого плана с тем же значением целевой функции); например,

$$X^5 = \begin{vmatrix} & 1-\theta & 3+\theta & 4 \\ 9 & \theta & 1-\theta & \\ & 2 & & \end{vmatrix} \quad \theta=1 \; = \quad \begin{vmatrix} & 0 & 4 & 4 \\ 9 & 1 & & \\ & 2 & & \end{vmatrix}$$

Обратите, кстати, внимание на вырожденность плана X^3 .

4.1. 4. Задача о назначении персонала

Пусть имеется m категорий претендентов в количестве A_i ($i = 1 .. m$) и n групп вакантных должностей по B_j ($j = 1 .. n$) в каждой.

Известны оценки C_{ij} использования претендента i -й категории на должности из j -й группы.

Задача поиска распределения с максимальной суммарной эффективностью приводит к задаче, отличающейся от транспортной лишь требованием максимизации целевой функции.

Возьмем для примера $m = 3$, $n = 3$, $A = (6, 3, 1)$, $B = (2, 2, 2)$ и приведенную ниже матрицу эффективностей C.

Так как условие баланса нарушено за счет избытка претендентов, вводим фиктивную группу должностей с нулевой эффективностью, после чего отыскиваем начальный опорный план по правилу предпочтения максимальному значению из элементов матрицы C.

$$C = \begin{array}{|ccc|} 4 & 6 & 1 \\ 5 & 6 & 3 \\ 9 & 8 & 8 \end{array} \quad C' = \begin{array}{|cccc|} 4 & 6 & 1 & 0 \\ 5 & 6 & 3 & 0 \\ 9 & 8 & 8 & 0 \end{array} \quad X^0 = \begin{array}{|cccc|c} & 2 & 0 & 4 & 6 \\ 1 & & 2 & & 3 \\ 1 & & & & 1 \\ \hline 2 & 2 & 2 & 4 & B \backslash A \end{array}$$

Проверка на оптимальность дает

$U_i + V_j =$

$U \backslash V$	0	6	1	0
0	0	**6**	**1**	**0**
2	**2**	8	**3**	2
9	**9**	15	10	9

$$\Delta_{ij} = \begin{array}{|cccc|} -4 & \circ & \circ & \circ \\ \circ & 2 & \circ & -1 \\ \circ & 7 & 2 & 1 \end{array}$$

Так как $\Delta_{11} < 0$, то план X^0 не оптимален. Перестраиваем его, полагая $X_{11} = \theta$:

$$X^1 = \begin{array}{|cccc|} \theta & 2 & 0-\theta & 4 \\ 1-\theta & & 2+\theta & \\ 1 & & & \end{array} \quad \theta = 0 \quad \begin{array}{|cccc|} 0 & 2 & & 4 \\ 1 & & 2 & \\ 1 & & & \end{array}$$

Проверяем этот план на оптимальность:

$U_i + V_j =$

$U \backslash V$	4	6	2	0
0	**4**	**6**	2	**0**
1	**5**	7	**3**	1
5	**1**	11	7	5

$$\Delta_{ij} = \begin{array}{|cccc|} \circ & \circ & 1 & \circ \\ \circ & 1 & \circ & 1 \\ \circ & 3 & -1 & 5 \end{array}$$

Так как $\Delta_{31} < 0$, перестраиваем план X^1, полагая $X_{31} = \theta$:

$$X^2 = \begin{array}{|cccc|} 0 & 2 & & 4 \\ 1+\theta & & 2-\theta & \\ 1-\theta & & \theta & \end{array} \quad \theta = 1 \quad \begin{array}{|cccc|} 0 & 2 & & 4 \\ 2 & & 1 & \\ & & 1 & \end{array}$$

И снова проверка плана на оптимальность:

$U_i + V_j =$

$U \backslash V$	4	5	2	0
0	**4**	**6**	2	**0**
1	**5**	6	**3**	1
6	10	11	**8**	6

$$\Delta_{ij} = \begin{array}{|cccc|} \circ & \circ & 1 & \circ \\ \circ & 0 & \circ & 0 \\ 1 & 3 & \circ & 6 \end{array}$$

Так как все $\Delta_{ij} \geq 0$, план X^2 оптимален (есть и другие оптимальные планы).

В отечественной литературе при рассмотрении транспортных задач часто упоминается *метод потенциалов*. Здесь вместо условий $U_i + V_j \le C_{ij}$ выбираются эквивалентные условия в форме $-U_i + V_j \le C_{ij}$, выполняющиеся для положительных перевозок оптимального плана в форме равенства, и промежуточные оценки оформляются последовательностью сводных таблиц (в одной клетке 3 или 4 показателя).

Нами рассмотрена классическая однопродуктовая транспортная задача с прямыми связями между "поставщиками" и "потребителями" и методика ее решения. На практике же чаще приходится иметь дело с транспортировкой в сети с промежуточными узлами и ограничениями на пропускные способности коммуникаций. Такого рода *задача в сетевой постановке* решается более сложными методами, но и эти методы (в частности, известный *венгерский метод*) базируются на соотношениях двойственности.

Транспортная задача с критерием оптимальности не по затратам, а по времени выполнения комплекса перевозок решается опять-таки посредством постановки сопряженной задачи с привлечением алгоритма поиска *максимального потока в транспортной сети* (этот же алгоритм используется в упомянутом венгерском методе).

Обе упомянутые задачи будут нами рассмотрены ниже.

4.2. Распределительные задачи

Указанная группа задач выступает своеобразным обобщением транспортной задачи и возникает при закреплении ресурсов по видам работ, при закреплении транспортных средств за маршрутами, при составлении топливно-энергетических балансов, при планировании военных операций с максимальным поражением целей, при распределении заказов между фирмами и т. п.

Рассмотрим одну из такого рода задач.

Имеется m типов автомашин в количествах A_i ($i = 1 .. m$), которые должны быть закреплены за n маршрутами с объемами перевозок, равными B_j ($j = 1 .. n$). Известно, что расходы на эксплуатацию и объем перевозок одной машины i-го типа на j-м маршруте равны C_{ij} и L_{ij}. Требуется найти распределение с минимальными эксплуатационными затратами.

Поставленная задача сводится к минимизации

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} \qquad (1)$$

при условиях

$$\sum_{j=1}^{n} X_{ij} \le A_i \quad (i = 1 .. m) \; , \tag{2}$$

$$\sum_{i=1}^{m} L_{ij} X_{ij} = B_j \quad (j= 1 .. n), \tag{3}$$

$$X_{ij} \; \ge \; 0 \; (i = 1 .. m \; , \; j = 1 .. n) \, . \tag{4}$$

Для приведения к канонической форме вводим в (2) ослабляющие переменные (эквивалентно добавлению фиктивного n+1 -го маршрута с нулевыми значениями $C_{i\ n+1}$). Сопоставим этому маршруту значения $L_{i\ n+1} = 1$ и добавим n+1–е ограничение в (3) с каким-то значением B_{n+1}, требуя минимизации

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n+1} C_{ij} X_{ij} \tag{5}$$

при условиях

$$\sum_{j=1}^{n+1} X_{ij} = A_i \qquad (\; i = 1 .. m), \tag{6}$$

$$\sum_{i=1}^{m} L_{ij} X_{ij} = B_j \qquad (j= 1 .. n+1), \tag{7}$$

$$X_{ij} \; \ge \; 0 \quad (i = 1 .. m \; , \; j = 1 .. n+1) \, . \tag{8}$$

Сопряженная задача состоит в максимизации

$$\widetilde{L}(U,V) = \sum_{i=1}^{m} A_i \, U_i + \sum_{j=1}^{n+1} B_j V_j \tag{9}$$

при условиях

$$U_i + L_{ij} V_j \le C_{ij} \quad \text{при всех } \; i \, , j \, . \tag{10}$$

Рассмотрим алгоритм У. Х. Малкова, использующий идею А. Фергюсона и Д. Данцига.

Решение начинается с построения начального опорного плана задачи (5)–(8) с использованием правила предпочтения минимальным затратам C_{ij} , максимальной производительности L_{ij} или максимальному производительности на единицу затрат

$$R_{ij} = \frac{L_{ij}}{C_{ij}} \quad . \tag{11}$$

Если для транспортной задачи использовалось $X_{i\,j} = min\ (A_i\, ,\, B_j\,)$, то здесь этот прием заменяется на

$$X_{ij} = min\ (\, A_i \, , \, B_j \, / \, L_{ij} \,) \;\; . \tag{12}$$

Число базисных компонент в найденном плане равно m+n. Если план оказывается вырожденным, то к числу базисных можно отнести и нулевые значения (не допуская замкнутой *вне фиктивного маршрута* цепочки по базису). Для автоматического устранения вырожденности используют ε-метод: увеличить все A_i на ε и некоторое B_j на $m\varepsilon$.

После получения опорного плана производится его проверка на оптимальность: согласно второй теореме двойственности базисным X_{ij} сопоставляется система $m+n$ уравнений с $m+n+1$ неизвестными $U_i + L_{ij} V_j = C_{ij}$, которая решается с точностью до константы (например, $V_{n+1} = 0$) и осуществляется проверка выполнения (10).

При обнаружении $\Delta_{kp} = U_k + L_{kp} V_p - C_{kp} > 0$, для соответствующего элемента X_{kp} отыскиваем так называемую *минимизирующую цепочку* с одним или двумя выходами на фиктивный столбец. Величины коррекции ΔX_{ij} для компонент нового плана полагаем равными $\gamma\theta_{ij}$.

По выделенной цепочке строим систему уравнений для сохранения баланса по строкам и столбцам относительно величин коррекции компонент нового плана (для строк $-\theta_{kp} + \theta_{kj} = 0$, для столбцов $-L_{kp}\theta_{kp} + L_{ip}\theta_{ip} = 0$). Решение полученной однородной системы находим с точностью до постоянного множителя. Приняв $\theta_{kp} = 1$, ищем остальные θ_{ij} и для неотрицательности компонент нового плана берем

$$\gamma = \min_{\theta_{ij} < 0} \frac{X_{ij}}{|\theta_{ij}|}\,. \tag{13}$$

Очевидно, что при таком выборе соответствующая компонента плана обратится в нуль и будет выведена из базиса.

<u>Пример.</u> Пусть $m = 4$, $n = 5$, $A = (10, 50, 15, 30)$, $B = (60, 175, 400, 100, 100)$ и матрицы стоимостей и производительностей (с добавлением фиктивного маршрута) имеют вид:

$$C = \left|\begin{matrix} 1 & 1 & 4 & 5 & 2 \\ 10 & 1 & 5 & 2 & 4 \\ 10 & 2 & 2 & 5 & 1 \\ 5 & 4 & 10 & 5 & 5 \end{matrix}\right|\left|\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}\right| \qquad L = \left|\begin{matrix} 5 & 10 & 20 & 15 & 20 \\ 2 & 5 & 5 & 10 & 4 \\ 30 & 8 & 20 & 25 & 5 \\ 15 & 20 & 10 & 20 & 10 \end{matrix}\right|\left|\begin{matrix} 1 \\ 1 \\ 1 \\ 1 \end{matrix}\right|$$

Для поиска начального опорного плана определим значения R_{ij} согласно (11) и затем руководствуемся (12):

$R =$

1	5	5	3	10
1/5	5	1	5	1
3	4	10	5	5
3	5	1	4	2

$X^0 =$

	10					10
	15	20	10		5	50
		5				15
4				10	16	30
60	175	400	100	100	...	$B\|A$

$X_{12} = min(10, 175/10) = 10$, $\quad X_{11} = X_{13} = X_{14} = X_{15} = X_{16} = 0$,

$X_{33} = min(15, 400/20) = 15$, $\quad X_{31} = X_{32} = X_{34} = X_{35} = X_{36} = 0$,

$X_{22} = min(50, (175–10·10)/5) = 15,\quad X_{42} = 0,$
$X_{24} = min(50–15, 100/10) = 10,\quad X_{44} = 0,$
$X_{41} = min(30, 60/15) = 4,\quad X_{21} = 0,$
$X_{45} = min(30–4, 100/10) =10,\quad X_{25} = 0,$
$X_{23} = min(50–15–10, (400–15·20)/5) = 20,\quad X_{43} = 0,$
$X_{26} = 50–15–20–10 = 5,\quad X_{46} = 30–4–10 =16 .$

.

Строим систему уравнений
$U_1 + 10\,V_2 = 1$, $U_2 + 10\,V_4 = 2$, $U_4 + 15\,V_1 = 5$, $U_2 + 5\,V_2 = 1$, $U_2 + V_6 = 0$, $U_4 + 10\,V_5 = 5$, $U_2 + 5\,V_3 = 5$, $U_3 + 20\,V_3 = 2$, $U_4 + V_6 = 0$ и, получив решение этой системы при $V_6 = 0$

$U_1 = –1$, $U_2 = 0$, $U_3 = –18$, $U_4 = 0$,
$V_1 = 1/3$, $V_2 = 1/5$, $V_3 = 1$, $V_4 = 1/5$, $V_5 = 1/2$,

отыскиваем $\Delta_{ij} = U_i + L_{ij} V_j - C_{ij}$:

$$U_i+L_{ij}V_j = \left|\begin{array}{ccccc|c} 2/3 & \mathbf{1} & 19 & 2 & 9 & -1 \\ 2/3 & \mathbf{1} & \mathbf{5} & \mathbf{2} & 2 & \mathbf{0} \\ -8 & -16 & \mathbf{2} & -13 & -15 & -18 \\ \mathbf{5} & 4 & 10 & 4 & \mathbf{5} & \mathbf{0} \end{array}\right| \quad \Delta_{ij} = \left|\begin{array}{ccccc|c} -1/3 & \circ & \mathbf{15} & -3 & 7 & -1 \\ -28/3 & \circ & \circ & \circ & -2 & \circ \\ -18 & -18 & \circ & -18 & -16 & -18 \\ \circ & 0 & 0 & -1 & \circ & \circ \end{array}\right|$$

Обнаружив положительные значения Δ_{ij} , переходим к другому плану, где, например, компонента X_{13} отлична от нуля (точнее, входит в число базисных переменных). Согласно обнаруженной "цепочке", элементы которой помечены * , строим систему уравнений:

$$X^1 = \left|\begin{array}{ccccc|c} & 10^* & +^* & & & \\ & 15^* & 20^* & 10 & & 5^* \\ & & 15 & & & \\ 4 & & & & 10 & 16 \end{array}\right|$$

$$\begin{aligned} \theta_{13} + \theta_{12} &= 0 \\ 20\,\theta_{13} + 5\,\theta_{23} &= 0 \\ 10\,\theta_{12} + 5\,\theta_{22} &= 0 \\ \theta_{22} + \theta_{23} + \theta_{26} &= 0 . \end{aligned}$$

Решив эту систему при $\theta = 1$, имеем

$$\theta_{12} = -1 ,\quad \Theta_{22} = 2 ,\quad \theta_{23} = -4 ,\quad \theta_{26} = 2 .$$

$$X^1 = \left|\begin{array}{ccccc|c} & 5 & 5 & & & \\ & 25 & & 10 & & 15 \\ & & 15 & & & \\ 4 & & & & 10 & 16 \end{array}\right|$$

Отыскав согласно (13)

$$\gamma = min\left(\frac{10}{1}, \frac{20}{5} \right) = 5 ,$$

получаем умножением на θ_{ij} значения корректур для компонент плана :

$\Delta X_{13} = 5$, $\Delta X_{12} = -5$, $\Delta X_{22} = 10$, $\Delta X_{23} = -20$, $\Delta X_{26} = 10$; отсюда находим компоненты плана X^1.

Вновь строим систему уравнений:
$U_1 + 10\,V_2 = 1$, $U_2 + 10\,V_4 = 2$, $U_4 + 15\,V_1 = 5$, $U_2 + 5\,V_2 = 1$, $U_2 + V_6 = 0$, $U_4 + 10\,V_5 = 5$, $U_1 + 20\,V_3 = 4$, $U_3 + 20\,V_3 = 2$, $U_4 + V_6 = 0$, и, отыскав решение этой системы при $V_6 = 0$

$$U_1 = -1\ ,\quad U_2 = 0\ ,\quad U_3 = -3\ ,\quad U_4 = 0\ ,$$
$$V_1 = 1/3\ ,\quad V_2 = 1/5\ ,\quad V_3 = 1/4\ ,\quad V_4 = 1/5\ ,\quad V_5 = 1/2\ ,$$

имеем

$$U_i + L_{ij}V_j = \begin{array}{|ccccc|c|} \hline 2/3 & \mathbf{1} & \mathbf{4} & 2 & 9 & -1 \\ 2/3 & \mathbf{1} & 5 & \mathbf{2} & 2 & \mathbf{0} \\ 7 & -7/5 & \mathbf{2} & 2 & -0.5 & -3 \\ \mathbf{5} & 4 & 2.5 & 4 & \mathbf{5} & \mathbf{0} \\ \hline \end{array} \quad \Delta_{ij} = \begin{array}{|ccccc|c|} \hline -1/3 & \circ & \circ & -3 & \mathbf{7} & -1 \\ -28/3 & \circ & 0 & \circ & -2 & \circ \\ -3 & -17/5 & \circ & -3 & -2.5 & -3 \\ \circ & 0 & 0 & -1 & \circ & \circ \\ \hline \end{array}$$

Так как обнаруживаются $\Delta_{15} > 0$, необходимо перейти к другому плану, где компонента X_{15} отлична от нуля. Согласно показанной ниже "цепочке", строим очередную систему уравнений:

$$X^2 = \begin{array}{|ccccc|c|} \hline & 5^* & 5 & & +^* & \\ & 25^* & & 10 & & 15^* \\ & & 15 & & & \\ 4 & & & & 10^* & 16^* \\ \hline \end{array} \qquad \begin{array}{c} \theta_{15} + \theta_{12} = 0, \\ 20\,\theta_{15} + 10\,\theta_{45} = 0, \\ 10\,\theta_{12} + 5\,\theta_{22} = 0, \\ \theta_{22} + \theta_{26} = 0, \quad \theta_{45} + \theta_{46} = 0 \end{array}$$

Решив эту систему при $\theta_{15} = 1$, имеем

$$\theta_{12} = -1\ ,\quad \theta_{45} = -2\ ,\quad \theta_{22} = 2\ ,\quad \theta_{26} = -2\ ,\quad \theta_{46} = 2\ .$$

Отыскав $\gamma = min\left(\frac{5}{1}, \frac{10}{2}, \frac{10}{2}\right) = 5$, получаем умножением на θ_{ij} значения корректур для компонент плана

$$\Delta X_{15} = 5,\ \Delta X_{12} = -5,\ \Delta X_{22} = 10,\ \Delta X_{45} = -10,\ \Delta X_{26} = -10,\ \Delta X_{46} = 10;$$

откуда получаем план X^2 (здесь мы оставили в базисе $X_{12} = 0$, а не X_{45}, так как при одинаковой производительности затраты меньше, $R_{12} > R_{45}$) .

Вновь строим систему уравнений:

$$U_1 + 10\,V_2 = 1\ ,\quad U_2 + 5\,V_2 = 1\ ,\quad U_3 + 20\,V_3 = 2\ ,$$
$$U_1 + 20\,V_2 = 4\ ,\quad U_2 + 10V_4 = 2\ ,\quad U_4 + 15\,V_1 = 5\ ,$$
$$U_1 + 20\,V_5 = 2\ ,\quad U_3 + 1\,V_6 = 0\ ,\quad U_4 + V_6 = 0\ .$$

Получив решение этой системы при $V_6 = 0$:

$$U_1 = -1\ ,\quad U_2 = 0\ ,\quad U_3 = -3\ ,\quad U_4 = 0\ ,$$
$$V_1 = 1/3\ ,\quad V_2 = 1/5\ ,\quad V_3 = 1/4\ ,\quad V_4 = 1/5\ ,\quad V_5 = 3/20\ ,$$

имеем

$$U_i + L_{ij}V_j = \begin{array}{|ccccc|c|} \hline 2/3 & \mathbf{1} & \mathbf{4} & 2 & \mathbf{2} & -1 \\ 2/3 & \mathbf{1} & 5/4 & \mathbf{2} & 3/5 & \mathbf{0} \\ 7 & -7/5 & \mathbf{2} & 2 & -9/4 & -3 \\ \mathbf{5} & 4 & 5/2 & 4 & 3/2 & \mathbf{0} \\ \hline \end{array} \quad \Delta_{ij} = \begin{array}{|ccccc|c|} \hline -1/3 & \circ & \circ & -3 & \circ & -1 \\ -28/3 & \circ & -17/4 & \circ & -17/4 & \circ \\ -3 & -17/5 & \circ & -3 & -13/4 & -3 \\ \circ & \mathbf{0} & -5/2 & -1 & -7/2 & \circ \\ \hline \end{array}$$

Поскольку все $\Delta_{ij} \geq 0$, можно утверждать оптимальность плана X^2. Однако наличие $\Delta_{42} = 0$ для небазисной компоненты плана определяет возможность существования и других оптимальных планов.

Очевидно, что при сохранении идеологии метода решения распределительная задача, в отличие от транспортной, не гарантирована от противоречивости условий и тем более от получения нецелочисленного оптимального плана даже при целочисленных исходных показателях. Само собой, решение возникающих здесь систем алгебраических уравнений реализуется не столь примитивно, как в методе Данцига.

4.3. Задачи на транспортных сетях

Рассмотрим некоторые задач, решение которых базируется на понятии максимального потока в транспортной сети.

4.3.1. Задача о максимальном потоке

Рассмотрим транспортную сеть, где выделены пункты 0 (вход, источник) и n (выход, сток) и каждой дуге (отрезку), связывающей пункты i и j, сопоставлено число $d_{ij} > 0$, называемое *пропускной способностью* дуги. Эта величина характеризует максимальное допустимое количество вещества (воды, газа, самолетов, вагонов и др.), которое может проходить по дуге в единицу времени. Количество вещества, реально проходящего по дуге от i до j , называем потоком по дуге $(i\,,j)$ и обозначаем через X_{ij} .

Очевидно, что

$$0 \le X_{ij} \le d_{ij} \text{ для всех } i\,,j\,. \tag{1}$$

Если учесть, что все вещество, вошедшее в промежуточный пункт сети, должно полностью выйти из него, получаем

$$\sum_i X_{ij} = \sum_k X_{jk}\,,\ j \ne 0\,,n\,. \tag{2}$$

Из естественного требования равенства суммарного потока на входе и на выходе имеем

$$\sum_j X_{0\,j} = \sum_i X_{i\,n} = Z\,. \tag{3}$$

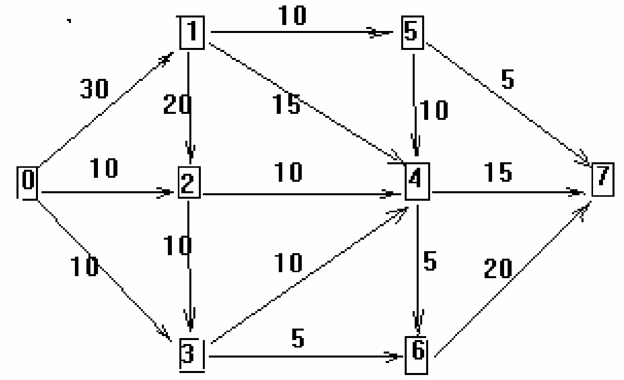
Величину Z называем *величиной потока* в сети и ставим задачу максимизации Z при условиях (1) – (3).

Решить задачу можно и симплексным методом, но едва ли эта возможность осуществима для реальных сетей.

Разобьем множество пунктов (вершин) сети на два непересекающихся подмножества U и V так, что вход $0 \in U$ и выход $n \in V$. Совокупность дуг, ведущих непосредственно из вершин множества U в вершины множества V, называют *разрезом сети*, а число

$$A(U, V) = \sum_{i \in U, j \in V} d_{ij}$$

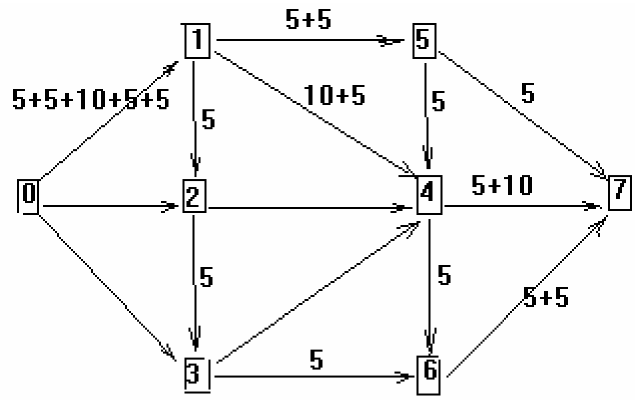
– пропускной способностью этого разреза. Так для ориентированной сети, приведенной на рисунке (числа на дугах-стрелках – пропускные способности), один из возможных разрезов определяется множествами $U = \{0, 1, 2, 3, 4\}$, $V = \{5, 6, 7\}$ и его пропускная способность равна $d_{15} + d_{36} + d_{46} + d_{47} = 10 + 5 + 5 + 15 = 35$.



Очевидно, что поток в сети не превышает величины пропускной способности любого ее разреза и равен пропускной способности минимального разреза.

В случае т. н. (0, n) – *плоских сетей*, т. е. сетей, которые можно изобразить на плоскости по одну сторону от линии, соединяющей вершины 0 и n, без пересечения дуг вне вершин (наша сеть относится к таковым), можно предложить простой наглядный алгоритм решения.

Берем самый «верхний» (по принципу левостороннего движения) путь от вершины 0 к вершине 7: [0 –1 – 5 – 7], отыскиваем минимальную пропускную способность составляющих его дуг, равную 5 , и уменьшаем пропускные способности дуг на эту величину.



Последующий поиск обнаруживает потоки по "верхним" путям : [0 –1 – 5 – 4 – 7], [0 – 1 – 4 – 7], [0 – 1 – 4 – 6 – 7] , [0 – 1 – 2 – 3 – 6 – 7] с пропускными способностями соответственно 5, 10, 5, 5. В итоге сеть оказалась разорванной и максимальный поток равен 30 .

Поскольку для больших сетей такой метод неприемлем, рассмотрим универсальный алгоритм поиска в матричной форме.

Строим матрицу D_0, куда заносим значения d_{ij} (для неориентированной дуги $d_{ij} = d_{ji}$); затем повторяем процесс поиска некоторого пути и коррекции потока на этом пути, используя процесс отмечаний.

Метим символом * нулевые строку и столбец матрицы (вход сети). В строке 0 отыскиваем $d_{0j} > 0$ и метим соответствующие столбцы индексами $\mu_j = 0$, $V_j = d_{0j}$ и переносим метки столбцов на строки. Затем берем i-ю отмеченную строку, ищем в ней непомеченный столбец с $d_{ij} > 0$, которому сопоставляем метки-индексы $\mu_j = i$, $V_j = \min(V_j, d_{ij})$. Метки столбцов переносим на строки и этот процесс продолжаем до тех пор, пока не будет отмечен n -й столбец.

$D_0 =$

	*	0/30	0/10	0/10	1/15	1/10	3/5	4/15	
	0	1	2	3	4	5	6	7	
0		30	10	10					*
1			20		15	10			0/30
2				10	10				0/10
3					10		5		0/10
4							5	15	1/15
5					10			5	1/10
6								20	3/5
7									4/15

$D_1 =$

	*	/15	0/10	0/10	2/10	1/10	3/5	5/5	
	0	1	2	3	4	5	6	7	
0		15	10	10					*
1	15		20		0	10			0/15
2				10	10				0/10
3					10		5		0/10
4		15					5	0	2/10
5					10			5	1/10
6								20	3/5
7					15				5/5

Затем "обратным ходом" по индексам выясняем путь, приведший к n-й вершине, и уменьшаем пропускные способности дуг пути (элементы матрицы) на V_n , увеличивая симметричные элементы на эту же величину. Такая процедура продолжается до тех пор, пока отмечание n -й вершины не станет невозможным.

Максимальный поток может быть найден вычитанием из исходной матрицы D_0 получаемой после вышеприведенной корректуры матрицы пропускных способностей $X = max(D_0 - D_k, 0)$.

Для рассмотренного примера строим матрицу D_0. Из строки 0 метим вершины 1, 2 и 3 (строки-столбцы) ндексами μ =0 и V, равными 30, 10 и 10. Из меченной строки 1 метим вершины 4 и 5 индексами μ = 1, $V_4 = min(30,15)=15$, $V_5 = min(30,10)=10$. Из строки 3 метим вершину 6 и, наконец, из строки 4 – вершину 7.

Обратным ходом по μ обнаруживаем путь: к вершине 7 от 4 , к 4 от 1 , к 1 от 0; корректируем элементы D_0 на величину V_7 =15.

Очередной шаг дает путь [0 – 1 – 5 – 7] с потоком 5.

Аналогичные действия приводят к матрице D_4, где дальнейшее отмечание невозможно. Отсюда получаем матрицу X_{max} максимального потока.

Этот алгоритм не накладывает никаких ограничений на транспортную сеть (любая ориентация, не обязательно плоская) и элементарно реализуется в программном виде.

4.3.2. Обобщенная задача о максимальном потоке

В отличие от описанной выше задачи здесь предполагается полная ориентированность сети, т. е. поток по дуге только в одном направлении, и ограничения на пропускную способность не только сверху, но и снизу.

Такая обобщенная задача, называемая иногда *задачей о потоке, совместимом с множеством значений*, состоит в максимизации

$$\sum_j X_{0j} = \sum_i X_{in} = Z$$

D_4=

	*	0/10	0/5	0/5	2/5	1/5			
	0	1	2	3	4	5	6		
0		10	5	5				7	*
1	20		20		0	5			0/10
2	5			10	5				0/10
3	5				10		0		0/5
4		15	5				0	0	2/5
5	5				10			0	1/5
6				5	5			10	
7					15	5	10		

X_{max} =

0	1	2	3	4	5	6	7
0	20	5	5				
1		0		15			
2			0	5			
3				0		5	
4						5	15
5				0			5
6							10
7							

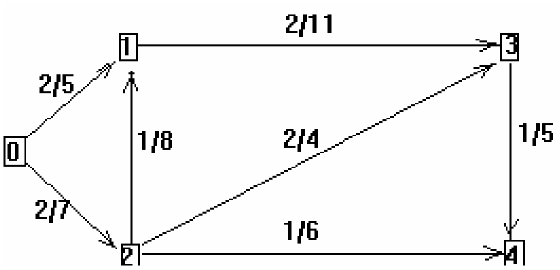
при условиях $\sum_i X_{ij} = \sum_k X_{jk}$, $j=1,..,n$; $0 \le b_{ij} \le X_{ij} \le d_{ij}$ для всех i, j .

Если в простой задаче о максимальном потоке всегда можно найти допустимый (например, нулевой) поток, то здесь такого потока может не существовать. Примером такой ситуации может служить задача, где при некотором j имеет место $\sum_i d_{ij} < \sum_k b_{jk}$.

Для поиска допустимого потока расширяем сеть добавлением вершин -1 (псевдовход) и $n+1$ (псевдовыход). Псевдовход соединяем со всеми вершинами, кроме 0 , дугами, пропускная способность которых равна сумме нижних границ по дугам, входящим в соответствующую вершину. Из всех вершин, кроме n, проводим в вершину $n+1$ дуги, пропускная способность которых равна сумме нижних границ пропускных способностей дуг, выходящих из соответствующей вершины. Кроме того, соединяем вершину n с вершиной 0 дугой с неограниченной пропускной способностью. Если нам удастся в такой расширенной сети найти максимальный поток, "насыщающий" псевдовыход, тем самым мы докажем существование допустимого потока.

В матричном виде алгоритм реализуется в два этапа.

На первом этапе строится матрица $D - B$, окаймляемая строками и столбцами -1 и $n+1$. Элементы с индексами $(-1, j)$ полагаем равными $\sum\limits_i b_{ij}$, элементы $(i, n+1)$ равными $\sum\limits_j b_{ij}$ и элемент $(n,0)$ – произвольно большому числу. К полученной матрице применяем алгоритм поиска максимального потока.



Если найденный максимальный поток не является насыщающим, задача неразрешима. В противном случае отбрасываем дополнительные строки и столбцы, исключаем связь $(n, 0)$; если $d_{0n} > 0$, то соответствующий элемент матрицы берем равным $d_{0n} - b_{0n}$ и 0 в противном случае. К полученной матрице вновь применяем алгоритм поиска максимального потока.

Пример. Рассмотрим сеть, приведенную выше, где числа на дугах - ограничения пропускных способностей снизу и сверху.

Строим исходные и расширенную матрицы.

$$
B = \begin{array}{c|ccccc|}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & & 2 & 2 & & \\
1 & & & & 2 & \\
2 & & 1 & & 2 & 1 \\
3 & & & & & 1 \\
4 & & & & & \\
\hline
\end{array}
\quad
D = \begin{array}{c|ccccc|}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & & 5 & 7 & & \\
1 & & & & 11 & \\
2 & & 8 & & 4 & 6 \\
3 & & & & & 5 \\
4 & & & & & \\
\hline
\end{array}
\quad
T_0 = \begin{array}{c|ccccccc|}
 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
-1 & & & 3 & 2 & 4 & 2 & \\
0 & & & 3 & 5 & & & 4 \\
1 & & & & & 9 & & 2 \\
2 & & & 7 & & 2 & 5 & 4 \\
3 & & & & & & 4 & 1 \\
4 & & \infty & & & & & \\
5 & & & & & & & \\
\hline
\end{array}
$$

Не рисуя ряд промежуточных матриц, найдем очевидные потоки с пропускной способностью r : $[\,-1 - 1 - 5\,]$, $r = 2$; $[\,-1 - 2 - 5\,]$, $r = 2$; $[\,-1 - 3 - 5\,]$, $r = 1$; $[\,-1 - 4 - 0 - 5\,]$, $r = 2$ и в результате обычной коррекции получаем матрицу T_4, из которой в процессе последующих отмечаний находим пути: $[-1-1-3-4-0-5]$, $r = 1$; $[-1-3-4-0-5]$, $r = 1$; $[-1-3-4-0-2-5]$, $r = 2$ и получаем матрицу T_7, показывающую исчерпание возможностей псевдовхода и псевдовыхода.

$$T_4 = \begin{array}{c|ccccccc|}
 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline
-1 & & & 1 & 0 & 3 & 0 & \\
0 & & & 3 & 5 & & 2 & 2 \\
1 & 2 & & & & 9 & & 0 \\
2 & 2 & & 7 & & 2 & 5 & 2 \\
3 & 1 & & & & & 4 & 0 \\
4 & 2 & \infty & & & & & \\
5 & & 2 & 2 & 2 & 1 & & \\ \hline
\end{array}
\qquad
T_7 = \begin{array}{c|ccccccc|}
 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline
-1 & & & 0 & 0 & 0 & 0 & \\
0 & & & 3 & 3 & & 6 & 0 \\
1 & 3 & & & & 8 & & 0 \\
2 & 2 & 2 & 7 & & 2 & 5 & 0 \\
3 & 4 & & 1 & & & 0 & 0 \\
4 & 2 & \infty & & & 4 & & \\
5 & & 4 & 2 & 4 & 1 & & \\ \hline
\end{array}$$

Проведя усечение T_7 к виду T_8 и обнаружив путь $[0-2-4]$ с $r = 3$, получаем матрицу T_9, в которой дальнейшее отмечание невозможно.

$$T_8 = \begin{array}{c|ccccc|}
 & 0 & 1 & 2 & 3 & 4 \\ \hline
0 & & 3 & 3 & & \\
1 & & & & 8 & \\
2 & 2 & 7 & & 2 & 5 \\
3 & & 1 & & & 0 \\
4 & & & & 4 & \\ \hline
\end{array}
\qquad
T_9 = \begin{array}{c|ccccc|}
 & 0 & 1 & 2 & 3 & 4 \\ \hline
0 & & 3 & 0 & & \\
1 & & & & 8 & \\
2 & 5 & 7 & & 2 & 2 \\
3 & & 1 & & & 0 \\
4 & & & 3 & 4 & \\ \hline
\end{array}
\qquad
X = \begin{array}{c|ccccc|}
 & 0 & 1 & 2 & 3 & 4 \\ \hline
0 & & 2 & 7 & & \\
1 & & & & 3 & \\
2 & & 1 & & 2 & 4 \\
3 & & & & & 5 \\
4 & & & & & \\ \hline
\end{array}$$

Отыскав матрицу $D - T_9$ и ограничиваясь только положительными значениями, получаем матрицу X максимального потока.

4.3.3. Венгерский метод для классической транспортной задачи

В отличие от ранее рассмотренного метода Данцига, где выбирался опорный план исходной задачи и проверялся на оптимальность посредством соотношений сопряженной задачи, здесь[9] берем какой-то план сопряженной задачи и, по соображениям второй теоремы двойственности, ищем решение исходной (максимальный поток в допустимой сети). Если таковое не удалось получить, то берем другой опорный план сопряженной задачи (другую допустимую сеть) и т. д.

Этот метод позволяет решать не только классическую транспортную задачу, но и транспортную задачу в сетевой постановке, где кроме «поставщиков» и «потребителей» присутствуют промежуточные

---

[9] Обсуждаемый метод называют венгерским, так как его идея высказана еще в 1931 году венгерским математиком Ф. Эгервари. Эта забытая работа была открыта в 1953 году американским математиком Г. Куном, который развил эту идею и назвал созданный им метод венгерским.

пункты и даже ограничения на пропускные способности отдельных дуг сети.

Рассмотрим идеологию и алгоритм метода на примере классической транспортной задачи минимизации

$$L(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}C_{ij}X_{ij} \tag{1}$$

при условиях

$$\sum_{j=1}^{n}X_{ij}=A_i\ ,\quad i=1\ ..\ m\ ; \tag{2}$$

$$\sum_{i=1}^{m}X_{ij}=B_j\quad ,\ \ j=1\ ..\ n\ \ ; \tag{3}$$

$$X_{ij}\geq\ 0,\quad i=1\ ..\ m\ ,\quad j=1\ ..\ n; \tag{4}$$

$$\sum_{i=1}^{m}A_i=\sum_{j=1}^{n}B_j\quad . \tag{5}$$

Как показано ранее, сопряженная задача сводится к максимизации

$$\widetilde{L}(\,U,V\,)\ =\sum_{i=1}^{m}A_i\,U_i+\sum_{j=1}^{n}B_jV_j \tag{6}$$

при условиях

$$U_i+V_j\ \leq\ C_{ij}\qquad \text{при всех}\ \ i\,,j\,. \tag{7}$$

Если построить матрицу

$$T=\|\,C_{ij}-U_i-V_j\,\|, \tag{8}$$

где

$$U_i=\min_j\ C_{ij}\,,\qquad i=1\ .\ .\ m, \tag{9}$$

$$V_j\ =\min_i\ [C_{ij}-U_i\,]\,,\ j=1\ .\ .\ n, \tag{10}$$

то очевидна неотрицательность элементов этой матрицы, т. е. найденные значения двойственных переменных удовлетворяют условиям (7).

Рассмотрим вспомогательную сеть, состоящую из дуг исходной сети, для которых $T_{ij}=C_{ij}-U_i-V_j$ равны нулю,

$$S=\{\,(\,i\,j\,)\,/\,C_{ij}-U_i-V_j=0\,\}\quad . \tag{11}$$

Если в ней найдем максимальный поток, удовлетворяющий (2)–(5), то по второй теореме двойственности он дает решение исходной задачи.

Для классической транспортной задачи воспользуемся компактной схемой поиска максимального потока с использованием матрицы $m\times n.$

Пусть найден некоторый поток в сети S, отвечающий условиям

$$\sum_{j=1}^{n} X_{ij} \le A_i \ (i = 1 \,..\, m) ; \ \sum_{i=1}^{m} X_{ij} \le B_j \ (j = 1 \,..\, n) ;$$

$$X_{ij} \ge 0 \quad (i = 1..m , \ j = 1..n).$$

Для строк i, в которых

$$\sum_{j=1}^{n} X_{ij} < A_i \ , \tag{12}$$

сопоставим метки

$$\mu_i = 0 \quad , \upsilon_i = A_i - \sum_{j=1}^{n} X_{ij} \quad . \tag{13}$$

Выбираем отмеченные строки (например, i-ю) и отмечаем неотмеченные столбцы j такие, что $(i\, j) \in S$, метками

$$\lambda_j = i \ , \quad \omega_j = \upsilon_i \ . \tag{14}$$

Затем берем отмеченные столбцы (например, j-й) и для неотмеченных строк i, в которых $X_{ij} > 0$, сопоставляем метки

$$\mu_i = j, \ \upsilon_i = min \ (\omega_j , X_{ij}) \ . \tag{15}$$

Повторяем процесс отмечания столбцов и строк до тех пор, пока не будет отмечен *ненасыщенный* столбец j^*, для которого

$$\sum_{i=1}^{m} X_{ij^*} < B_{j^*} \ . \tag{16}$$

Отыскиваем величину

$$\theta = min \ (\omega_{j^*} , B_{j^*} - \sum_{i=1}^{m} X_{ij^*}) \ , \tag{17}$$

определяющую величину потока в найденном пути; поочередно добавляем и вычитаем θ из значений X_{ij} в цепочке

$$(i_0\, j^*)(i_0\, j_1)(i_1\, j_1)(i_1\, j_2)\ (i_2\, j_2\ )\ \ldots (i_{k-1}\, j_k)(i_k\, j_k)$$

где

$$i_0 = \lambda_{j^*} , \ j_1 = m_{i_0} , i_1 = \lambda_{j_1} , \ j_2 = m_{i_1} , \ i_2 = \lambda_{j_2} , .., i_k = \lambda_{j_k} \ (m_{i_k} = 0)$$

и вновь продолжаем процесс отмечаний.

Если не удастся отметить ни одного из ненасыщенных столбцов, то перестраиваем сеть: отыскиваем величину

$$K = \min_{i \in I,\, j \notin J} T_{ij} \quad , \tag{18}$$

где I, J - множества отмеченных строк и столбцов, вычитаем K из отмеченных строк матрицы T и добавляем к отмеченным столбцам. Это действие эквивалентно корректуре двойственных переменных

$$U_i' = \begin{cases} U_i + K\,,\ i \in I \\ U_i \quad ,\ i \notin I \end{cases}, \qquad V_j' = \begin{cases} V_j - K\,,\ j \in J \\ V_j \quad ,\ j \notin J \end{cases}.$$

Такой процесс продолжается до тех пор, пока максимальный поток не окажется насыщающим для строк и столбцов.

Рассмотрим пример решения задачи при следующих данных.

$C =$

5	2	3	8	6
3	4	7	2	5
6	5	3	4	8
4	3	7	5	$B \setminus A$

Отыскав значения двойственных переменных (9)–(10) $U_1 = 2$, $U_2 = 2$, $U_3 = 3$, $V_1 = 1$,

$$V_2 = 0\,,\ V_3 = 0\,,\ V_4 = 0\,,$$

строим матрицу T и ищем в допустимой сети некоторый поток, например, по правилу минимальной стоимости:

$T_0 =$

2	0	1	6
0	2	5	0
2	2	0	1

$X_0 =$

	3			6
0			5	5
		7		8
4	3	7	5	

Отмечаем ненасыщенные строки

$$\mu_1 = 0\,,\ \upsilon_1 = 6 - 3 = 3\,;\quad \mu_3 = 0\,,\ \upsilon_3 = 8 - 7 = 1\,.$$

Из отмеченных строк по допустимым клеткам отмечаем неотмеченные столбцы $\lambda_2 = 1$, $\omega_2 = \upsilon_1 = 3$, $\lambda_3 = 3$, $\omega_3 = \upsilon_3 = 1$.

Попытка отметить неотмеченные строки по положительным значениям X_{ij} в отмеченных столбцах не увенчивается успехом. Следовательно, в выбранной сети найденный поток максимален и требуется перестройка сети.

Находим $\quad K = \min\limits_{\substack{i=1,3 \\ j=1,4}} T_{ij} = min\ (2\,,6\,,2\,,1) = 1\,;$

вычитаем K из строк 1 и 3 и добавляем к столбцам 2 и 3 матрицы T_0, сохраняя в новой сети тот же поток:

$T_1 =$

1	0	1	5
0	3	6	0
1	2	0	0

$X_0' =$

	3			6
0			5	5
		7	0	8
4	3	7	5	

Продолжая процесс отмечаний, имеем

$$\lambda_4 = 3\,,\ \omega_4 = \upsilon_3 = 1\,,\quad \mu_2 = 4\,,\ \upsilon_2 = max(\omega_4, X_{24}) = 1\,,\ \lambda_1 = 2,\ \omega_1 = \upsilon_2 = 1\,.$$

Поскольку отмечен ненасыщенный столбец 1, ищем величину коррекции потока $\theta = max(\omega_1\,, 4–0) = 1$ и выявляем “минимизирующую цепочку”:

$$j^* = 1,\ i_0 = \lambda_1 = 2,\ j_1 = \mu_2 = 4,\ i_1 = \lambda_4 = 3,\ j_2 = \mu_3 = 0,$$

которая определяет последовательность $[X_{21}\ \ X_{24}\ \ X_{34}]$.

Поочередное добавление–вычитание θ дает поток X_1.

Повторяя процесс отмечаний :

$$\mu_1 = 0,\ \upsilon_1 = 6–3 = 3,\ \lambda_2 = 1,\ \omega_2 = \upsilon_1 = 3,$$

обнаруживаем невозможность последующих отмечаний. Отыскиваем значение корректуры двойственных переменных

$$K = \min_{\substack{i=1 \\ j=1,3,4}} T_{ij} = \min\,(1, 1, 5) = 1\ ,$$

вычитаем K из строки 1 и добавляем к столбцу 2 матрицы T_1 , сохраняя в новой сети тот же поток :

$T_2 =$

0	0	0	4
0	4	6	0
1	3	0	0

$X_1' =$

0	3	0		6
1			4	5
		7	1	8
4	3	7	5	

Продолжая процесс отмечаний, имеем $\lambda_1 = 1,\ \omega_1 = \upsilon_1 = 3$.

Величина корректуры потока равна $\theta = \min\,(\omega_1\ , 6–3) = 3$ и минимизирующая цепочка, согласно

$$j^* = 1,\ i_0 = \lambda_1 = 1,\ j_1 = \mu_1 = 0,$$

состоит их единственного элемента X_{11}; отсюда получаем улучшенный поток X_2, который является насыщающим и дает решение поставленной задачи. Заметим, что при другой системе отмечаний

$\mu_1 = 0\ ,\ \upsilon_1 = 6 – 3 = 3\ ,$ $\lambda_4 = 3\ ,\ \omega_4 = \upsilon_3 = 3\ ,$

$\lambda_3 = 1\ ,\ \omega_3 = \upsilon_1 = 1\ ,$ $\mu_2 = 4\ ,\ \upsilon_2 = \min\,(\omega_4, X_{24}) = 3,$

$\mu_3 = 3\ ,\ \upsilon_3 = \min(\omega_3, X_{33}\ ) = 3\ ,$ $\lambda_1 = 2\ ,\ \omega_1 = \upsilon_2 = 3\ ,$

дающей $\theta = 3$ и цепочку $[X_{21}\ ,\ X_{24}\ ,\ X_{34}\ ,\ X_{33}\ ,\ X_{13}]$, получаем другой оптимальный поток X_2'.

$X_2 =$

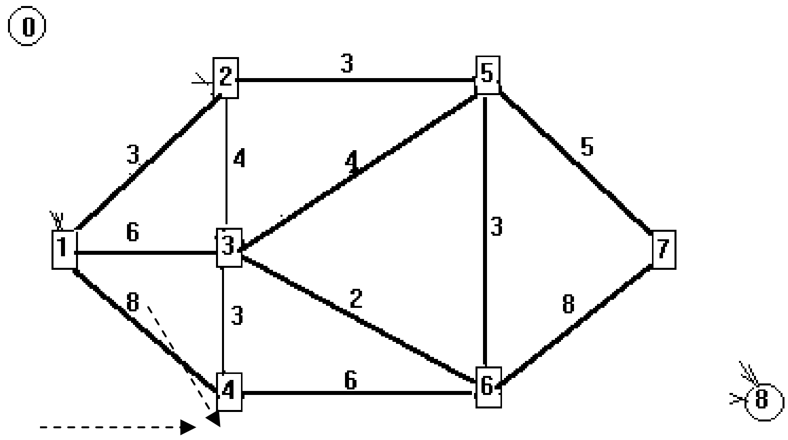
3	3	0		6
1			4	5
		7	1	8
4	3	7	5	

$X_2' =$

0	3	3		6
4			1	5
		4	4	8
4	3	7	5	

4.3.4. Венгерский метод для транспортной задачи в сетевой постановке

Рассмотрим более общую задачу на примере нижеприведенной сети, где числа на дугах определяют стоимость единичной перевозки, вершины 1 и 2 – пункты производства продукта в количествах 20 и 1 1 единиц, вершины 6 и 7 – пункты потребления в количествах 15 и 15 единиц.



	3			6

К тому же на маршрутах 2 – 5 и 3 – 6 пропускная способность не превышает соответственно 5 и 7 единиц продукта.

Воплощая основную идею венгерского метода, введем фиктивные вход 0 и выход 8. Соединим вход 0 с вершинами 1 и 2 дугами с пропускной способностью, равной объемам производства, и вершины 6 и 7 с выходом 8 – дугами с пропускной способностью, равной объемам потребления. Стоимости перевозок на этих дугах берем нулевыми.

В полученной сети ищем максимальный поток от входа к выходу, обеспечивающий минимум стоимости перевозок

$$L(X) = \sum_{(i\,j)} C_{ij}\, X_{ij} \qquad (1)$$

при условиях

$$\sum_j X_{0\,j} = A, \qquad (2) \qquad\qquad \sum_i X_{i\,j} = \sum_k X_{j\,k}\ ,\ j \neq 0\,,n \qquad (4)$$

$$\sum_i X_{i\,n} = A, \qquad (3) \qquad\qquad 0 \le X_{i\,j} \le d_{i\,j} \quad \text{для всех } i\,,j\,, \ (5)$$

и соблюдении баланса

$$A = \sum a_i = \sum b_j \quad . \qquad (6)$$

Можно переписать (4) в виде $\sum_i X_{i\,j} = V_j$ и $\sum_k X_{j\,k} = V_j$, где V_j – неопределенная величина, не превышающая A, и получить задачу, внешне похожую на классическую транспортную, но наличие ограничений на пропускные способности делает ее многократно более сложной.

Один из возможных алгоритмов состоит в следующем.

1. Отмечаем вход некоторым значком (множество отмеченных вершин в дальнейшем будем обозначать через M).

2. Отыскиваем неотмеченные вершины (j), в которые ведут ненасыщенные дуги с нулевыми стоимостями перевозок из вершин множества M, т. е. дуги с характеристиками $C_{Mj} = 0$, $X_{Mj} < d_{Mj}$. При отсутствии таковых переходим к пункту 3 алгоритма и при наличии – к пункту 4 .

3. Отыскиваем среди неотмеченных вершину, из которой идет дуга во множество M с нулевой стоимостью и ненулевой перевозкой, и переходим к пункту 4. При отсутствии таких – к пункту 7.

4. Выбранную вершину включаем во множество M и, если выход остался непомеченным, переходим к пункту 2 .

5. По меткам, сопоставленным отмеченным вершинам (откуда и сколько), отыскиваем путь от входа к выходу и его пропускную способность и корректируем суммарный поток и пропускные способности (по аналогии с поиском максимального потока в простейшем случае).

6. Если пропускные способности дуг, исходящих из вершины входа, не исчерпаны, переходим к пункту 1.

7. Выясняем наличие дуг, для которых $C_{Mj} > 0$ при $j \notin M$ или $C_{iM} < 0$ при $i \in M$. Если таковых нет, задача неразрешима.

8. Среди модулей найденных C_{Mj} и C_{iM} отыскиваем минимальное значение δ и корректируем матрицу стоимостей, вычитая δ из стоимостей на дугах, ведущих из множества M, и добавляя к стоимостям на дугах, ведущих во множество M. Возвращаемся к пункту 2.

Покажем реализацию алгоритма в матричной форме для вышеприведенной сети.

$$
C = \begin{array}{c|ccccccccc|c}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \\
\hline
0 & & 0 & 0 & & & & & & & * \\
1 & & & 3 & 6 & 8 & & & & & * \\
2 & & 3 & & 4 & & 1 & & & & * \\
3 & & 6 & 4 & & 3 & 5 & 2 & & & \\
4 & & 8 & & 3 & & & 6 & & & \\
5 & & & 1 & 5 & & & 3 & 5 & & \\
6 & & & & 2 & 6 & 3 & & 8 & 0 & \\
7 & & & & & & 5 & 8 & & 0 & \\
8 & & & & & & & & & & \\
\hline
 & & * & * & * & & & & & &
\end{array}
\qquad
D = \begin{array}{c|ccccccccc|}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
0 & & 20 & 10 & & & & & & \\
1 & & & \infty & \infty & \infty & & & & \\
2 & & \infty & & \infty & & 5 & & & \\
3 & & \infty & \infty & & \infty & \infty & 7 & & \\
4 & & \infty & & \infty & & & \infty & & \\
5 & & & \infty & \infty & & & \infty & \infty & \\
6 & & & & \infty & \infty & \infty & & \infty & 15 \\
7 & & & & & & \infty & \infty & & 15 \\
8 & & & & & & & & & \\
\hline
\end{array}
$$

Выбрав начальным нулевой поток X^0, задаем начальные метки:

$\mu_0 = -1$, $V_0 = 20 + 10 = 30$ (общий объем производства);

и с учетом $C_{01}=0, X_{01}<D_{01}$ и $C_{02}=0, X_{02}<D_{02}$:

$\mu_1 = 0$, $V_1 = min\,(V_0\,, D_{01}\,) = 20$;

$\mu_2 = 0$, $V_2 = min\,(V_0\,, D_{02}\,) = 10$.

$$
C_1 = \begin{array}{c|ccccccccc|c}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \\
\hline
0 & & 0 & 0 & & & & & & & \\
1 & & & 3 & 5 & 7 & & & & & * \\
2 & & 3 & & 3 & & 0 & & & & * \\
3 & & 7 & 5 & & 3 & 5 & 2 & & & \\
4 & & 9 & & 3 & & & 6 & & & \\
5 & & & 2 & 5 & & & 3 & 5 & & * \\
6 & & & & 2 & 6 & 3 & & 8 & 0 & \\
7 & & & & & & 5 & 8 & & 0 & \\
8 & & & & & & & & & & \\
\hline
 & & * & * & & & * & & & &
\end{array}
$$

Так как в отмеченных строках нет нулевых стоимостей $C_{Mj}, j \notin M$, дальнейшее отмечание невозможно. Найдя

$$\delta = \min_{\substack{i=0,1,2 \\ j \neq 0,1,2}} \left| C_{i\,j} \right| = min\,(6, 8, 4, 1) = 1,$$

вычитаем δ из отмеченных строк матрицы C, добавляя к отмеченным столбцам (получаем матрицу C_1) и, продолжая процесс отмечаний с учетом C_1 и D, можем отметить в дополнение к отмеченным вершину 5:

$\mu_5 = 2$, $V_5 = min\,(V_2\,, D_{25}) = 5$;

$$
C_2 = \begin{array}{c|ccccccccc|}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
0 & & 0 & 0 & & & & & & \\
1 & & & 3 & 2 & 4 & & & & \\
2 & & 3 & & 0 & & 0 & & & \\
3 & & 10 & 8 & & 3 & 8 & 2 & & \\
4 & & 12 & & 3 & & & 6 & & \\
5 & & & 2 & 2 & & & 0 & 2 & \\
6 & & & & 2 & 6 & 6 & & 8 & 0 \\
7 & & & & & & 8 & 8 & & 0 \\
8 & & & & & & & & & \\
\hline
\end{array}
$$

после чего опять-таки приходится расширять сеть путем коррекции матрицы C_1 на величину

$$\delta = \min_{\substack{i=0,1,2,5 \\ j\neq 0,1,2,5}} |c_{ij}| =$$

$$= min\,(5, 7, 3, 5, 3, 5) = 3,$$

получая матрицу C_2.

Продолжив отмечания, имеем:

$$\mu_3 = 2\,,\ V_3 = min\,(V_2, D_{23}) = 10;$$

$$\mu_6 = 5\,,\ V_6 = min\,(V_5, D_{56}) = 5\,; \qquad \mu_8 = 6\,,\ V_8 = min\,(V_6, D_{68}) = 5\,.$$

Обратным ходом по меткам выясняем путь $[0 \to 2 \to 5 \to 6 \to 8]$ с пропускной способностью $V_8 = 5$ и корректируем матрицу D, уменьшая ее элементы, соответствующие дугам пути в прямом направлении, и увеличивая симметричные элементы на $V_8 = 5$.

$X^1 =$

	0	1	2	3	4	5	6	7	8
0		0	5						
1			0	0	0				
2		0		0		5			
3		0	0		0	0	0		
4		0		0			0		
5			0	0			5	0	
6				0	0	0		0	5
7						0	0		0
8									

$D_1 =$

	0	1	2	3	4	5	6	7	8
0		20	5						
1			∞	∞	∞				
2	5	∞		∞		0			
3		∞	∞		∞	∞	7		
4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	10
7						∞	∞		15
8							5		

Повторяя процедуру отмечаний (на основе C_2, ,D_1 и X^1), имеем:

$\mu_0 = -1\,,\ V_0 = 20+5 = 25\,;$ $\qquad \mu_1 = 0\,,\ V_1 = min\,(V_0, D_{01}) = 20\,;$

$\mu_2 = 0\,,\ V_2 = min\,(V_0, D_{02}) = 5\,;$ $\mu_3 = 2\,,\ V_3 = min\,(V_2, D_{23}) = 5$

(вершину 5 отметить не удается, так как $D_{25}=0$). Соответственно $\delta = min(4,3,7,2) = 2$ (значение $C_{25}=0$ не учитываем, так как соответствующей дуги на этом этапе решения нет). После вычитания из отмеченных строк и добавления к отмеченным столбцам получаем матрицу C_3, и продолжая процесс отмечаний, имеем:

$C_3 =$

	0	1	2	3	4	5	6	7	8
0		0	0						
1			3	2	2				
2		3		0		$\boxed{0}$			
3		10	8		1	6	0		
4		14		5			6		
5			4	4			0	2	
6				4	6	6		8	0
7						8	8		0
8									

$\mu_6 = 3$, $V_6 = min(V_3, D_{36}) = 5$; $\mu_8 = 6$, $V_8 = min\,(V_6, D_{68}) = 5$, откуда имеем путь $[0 \to 2 \to 3 \to 6 \to 8]$ с пропускной способностью 5. Выполнив обычную коррекцию матрицы пропускных способностей и потока, получаем D_2 и X_2.

0 1 2 3 4 5 6 7 8   0 1 2 3 4 5 6 7 8

0		20	0						
1			∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	2		
$D_2=$ 4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	5
7						∞	∞		15
8							10		

0		0	10						
1			0	0	0				
2		0		5		5			
3		0	0		0	0	5		
$X^2=$4		0		0			0		
5			0	0			5	0	
6				0	0	0		0	10
7						0	0		0
8									

Повторяя процедуру отмечаний, имеем: $\mu_0 = -1$, $V_0 = 20$;

$\mu_1 = 0$, $V_1 = min\ (V_0\ , D_{01}\ ) = 20$; $\mu_3 = 1$, $V_3 = min\ (V_1\ ,D_{13}\ )= 20$;

$\mu_6 = 3$, $V_6 = min\ (\ V_3, D_{36}) = 2$; $\mu_8 = 6$, $V_8 = min\ (V_6\ ,D_{68}\ ) = 2$,

т. е. путь $[\ 0 \to 1 \to 3 \to 6 \to 8\ ]$ с величиной потока 2.

	0	1	2	3	4	5	6	7	8
0		18	0						
1	2		∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	0		
$D_3=$ 4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	3
7						∞	∞		15
8							12		

	0	1	2	3	4	5	6	7	8
0		2	10						
1			0	2	0				
2		0		5		5			
3		0	0		0	0	7		
$X^3=$4		0		0			0		
5			0	0			5	0	
6				0	0	0		0	12
7						0	0		0
8									

Повторяя процедуру отмечаний, имеем:

$\mu_0 = -1$, $V_0 = 18$;

$\mu_1 = 0$, $V_1 = min\ (V_0\ , D_{01}\ ) = 18$.

Находим $\delta = min(3, 2, 2) = 2$ и получаем матрицу C_4 , из которой

$\mu_3 = 1$, $V_3 = min\ (V_1\ , D_{13}\ ) = 18$;

$\mu_4 = 1$, $V_4 = min\ (V_1\ , D_{14}) = 18$.

(продолжение опять-таки невозможно).

	0	1	2	3	4	5	6	7	8
0		0	0						
1			1	0	0				
2		5		0		$\boxed{0}$			
3		12	8		1	6	$\boxed{0}$		
$C_4=$4		16		5			6		
5			4	4			0	2	
6				4	6	6		8	0
7						8	8		0
8									

Ищем $\delta = min(\ 1, 8, 6, 6) = 1$ и получаем матрицу C_5, из которой возникает возможность отметить вершину 2 (но не выход сети).

Опять-таки ищем $\delta = min\ (5, 5) = 5$ и перестраиваем матрицу стоимостей к виду C_6

	0	1	2	3	4	5	6	7	8
0		0	0						
1			0	0	0				

	0	1	2	3	4	5	6	7	8
0		0	0						
1			0	0	0				

2		6		1		$\boxed{0}$			
3		12	7		1	5	$\boxed{0}$		
C_5=4		16		5			5		
5			4	5			0	2	
6				5	7	6		8	0
7						8	8		0
8									

2		6		1		$\boxed{0}$			
3		12	7		1	0	$\boxed{0}$		
C_6=4		16		5			0		
5			9	10			0	2	
6				10	12	6		8	0
7						8	8		0
8									

Процедура отмечаний обнаруживает возможность перехода [0 →1 →4 →6 →8] с пропускной способностью 3 :

	0	1	2	3	4	5	6	7	8
0		15	0						
1	5		∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	0		
D_4= 4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	0
7						∞	∞		15
8							15		

	0	1	2	3	4	5	6	7	8
0		5	10						
1			0	5	0				
2		0		5		5			
3		0	0		0	3	7		
X^4=4		0		0			0		
5			0	0			8	0	
6				0	0	0		0	15
7						0	0		0
8									

Здесь обнаруживается возможность отмечания всех вершин, кроме 7 и 8. Находим $\delta = min\ (2, 8) = 2$ и перестраиваем матрицу стоимостей к виду C_7.

Теперь очевидна возможность отмечания всех вершин и обнаруживается путь [0→ 1→ 3→ 5 → 7→ 8] с пропускной способностью 15.

	0	1	2	3	4	5	6	7	8
0		0	0						
1			0	0	0				
2		6		1		$\boxed{0}$			
3		12	7		1	0	$\boxed{0}$		
C_7=4		16		5			0		
5			9	10			0	0	
6				10	12	6		6	$\boxed{0}$
7						10	10		0
8									

В итоге получаем D_5 и X^5 – возможности поставщиков исчерпаны и потребности удовлетворены.

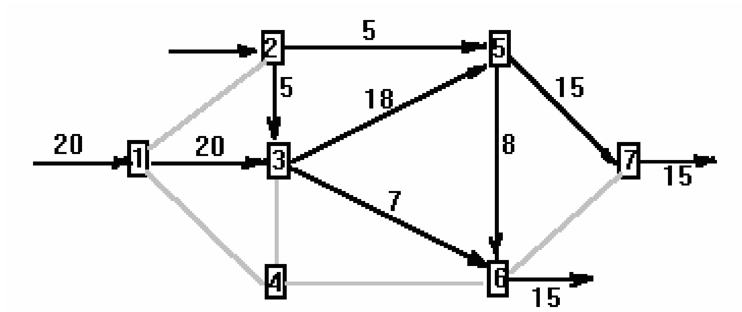
	0	1	2	3	4	5	6	7	8
0		0	0						
1	20		∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	0		
D_5= 4		∞		∞			∞		

	0	1	2	3	4	5	6	7	8
0		20	10						
1			0	20	0				
2		0		5		5			
3		0	0		0	18	7		
X^5=4		0		0			0		

5		∞	∞			∞	∞	
6			∞	∞	∞		∞	0
7					∞	∞		0
8						15	15	

5		0	0			8	15	
6			0	0	0		0	15
7					0	0		15
8								

Граф максимального потока представлен на рисунке.



Заметим, что матрицу X^5 можно получать и по окончании счета, вычитая D_5 из исходной матрицы D_0.

Нет сомнений, что мало-мальски приличную транспортную задачу вручную никто не будет решать и естественно прибегнуть к помощи компьютера (если у вас есть возможность стать обладателем соответствующего программного средства).

4.3.5. Транспортная задача по критерию времени

В рассматриваемой задаче критерием качества организации перевозок являются не суммарные затраты, а время реализации перевозок (подобные проблемы возникают при транспортировке скоропортящихся грузов, при переброске сил быстрого реагирования и т. д.).

Пусть имеется m поставщиков продукта в количествах A_i (i=1.. m) и n потребителей в количествах B_j (j = 1 .. n), причем соблюдается баланс предложения и спроса. Известно время $t_{i\,j}$ доставки груза от i-го поставщика к j-му потребителю, не зависящее от объема перевозки.

Требуется среди всех допустимых планов перевозок $X = \{ X_{ij} \}$ найти план, оптимальный по времени. Очевидно, что время, необходимое для реализации плана, совпадает с наибольшим временем отдельных перевозок и оптимальное время перевозок равно

$$T_{opt} = \min_{X} \max_{X_{i\,j}>0} t_{i\,j}$$

при условиях

$$\sum_{j=1}^{n} X_{i\,j} = A_i, \quad i = 1 \,..\, m$$

$$\sum_{i=1}^{m} X_{i\,j} = B_j, \quad j = 1 \,..\, n$$

$$X_{ij} \geq 0 \ , \quad i = 1 \,..\, m \ , \quad j = 1 \,..\, n$$

$$\sum_{i=1}^{m} A_i = \sum_{j=1}^{n} B_j \quad .$$

Алгоритм решения базируется на идеях венгерского метода для классической транспортной задачи и элементарном здравом смысле.

Предварительно выбирается минимальное значение t_{ij} и строится допустимая сеть по значениям t_{ij}, не превышающим выбранного.

В этой сети отыскивается максимальный поток. Если этот поток отвечает условиям задачи, то выбранное время оптимально. В противном случае выбирается очередное наименьшее время, сеть расширяется и в ней вновь ищется максимальный поток.

Очевидно, что для выбора начального времени разумнее отталкиваться не от минимального значения, а от максимального среди минимальных времен в строках и столбцах матрицы T.

Пример. Пусть задача определена данными, приведенными в таблице.

$T =$

1	13	17	18	18
2	18	10	10	10
16	1	4	12	12
11	9	13	7	*B\A*

Минимальные значения t_{ij} в строках равны 1, 2, 1 и в столбцах 1, 1, 4, 10.

Выбираем t^*=10, строим вспомогательную сеть по $t_{ij} \leq t^*$ и отыскиваем в ней начальное приближение для потока, например, методом северо-западного угла.

Так как найденный поток X^0 не является насыщающим, пытаемся его улучшить с использованием процесса отмечаний венгерского метода $\mu_1 = 0$, $\upsilon_1 = 18 - 11 = 7$, $\lambda_1 = 1$, $\omega_1 = \upsilon_1 = 7$.

$X^o =$

11				18
0		10	0	10
	9	3		12
11	9	13	7	*B\A*

Дальнейшее отмечание невозможно и приходится расширить сеть, взяв t^*=12 (появится возможность перевозки на маршруте 1 – 4 , поток X_o').

$X_o' =$

11				18
0		10	0	10
	9	3	0	12
11	9	13	7	*B\A*

Очевидно, что это расширение ничего нового не даст; берем t^*=13 (поток X_0'').

Отталкиваясь от ранее выбранного потока, пытаемся его улучшить. Продолжая процесс отмечаний, имеем

$$\lambda_2 = 1 \ , \ \omega_2 = \ \upsilon_1 \ = 7 \ ;$$

$$\mu_3 = 2 \ , \ \ \upsilon_3 \ = min(X_{3\,2}\, , \omega_2) = 7 \ ;$$

$$\lambda_2 \ = 3 \ , \ \omega_4 = \ \upsilon_3 \ = 7 \ .$$

$X_o'' =$

11	0			18
0		10	0	10
	9	3	0	12
11	9	13	7	*B\A*

Так как отмечен ненасыщенный столбец, отыскиваем минимизирующую цепочку [X_{34} X_{32} X_{12}] и корректируем ее на величину θ $= min$ $(\omega_4 , B_4) = 7$.

В итоге мы получаем насыщающий поток и можем утверждать, что минимальное время транспортировки составляет 13 единиц.

X_1 =

11	7			18
0		10	0	10
	2	3	7	12
11	9	13	7	*B\A*

4.3.6. Замечания

Исторически первые опыты в сфере исследования операций связаны с задачами транспортировки грузов, задолго до появления самого термина. Особо пристальное внимание к этой сфере было привлечено в США в годы Второй мировой войны в связи с организацией снабжения войск союзников на театре военных действий от Северной Африки и Европы до Тихого океана.

Не обошли вниманием эту тематику и советские математики, уже в началу 60-х разработавшие множество оригинальных методов решения сетевых задач. Так в опубликованной в 1962 г. монографии И. Я. Бирмана «Транспортная задача линейного программирования» рассматривались методы потенциалов и дифференциальных рент, задачи оптимальных перевозок угля по Сибири и Дальнему Востоку, распределения стройиндустрии Казахстана и другие.

Однако, несмотря на сотни многостраничных учебных пособий на тему транспортной задачи, лишь единицы могут претендовать на лаконичность и изящество доказательной базы, достигнутое С. Гассом [5].

Принципиальную роль в методологии решения сетевых задач сыграли фундаментальные работы Л. Форда и Д. Фалкерсона [13] с их алгоритмами, базирующимися на поиске максимального потока. Неплохой обзор этих методов можно найти в [14].

Фантастически обширную литературу собрала задача о коммивояжере (кратчайшим путем обойти все вершины транспортной сети, но только по одному разу, и вернуться в исходную вершину) – до сих пор предлагаются эвристические алгоритмы ее решения – от метода ветвей и границ до генетических алгоритмов.

Давно стали популярными азбучно простые, но требующие помощи компьютеров методы поиска кратчайших или самых длинных путей в транспортной сети (далее мы коснемся одного из этих методов, базирующегося на идеях Р. Беллмана [20]).

Заслуживает упоминания возникающая иногда при синтезе цепей задача поиска кратчайших связывающих сетей, реализуемая достаточно простым алгоритмом, описанным в статье Р. К. Прим [15].

Широкий круг комбинаторных задач, связанных с прогулками по вершинам и дугам (ребрам) сети (при учете ориентации используется термин «дуга» и в противном случае – термин «ребро»), решается с привлечением аппарата известной теории графов (среди множества пособий по теории графов начинающему исследователю рекомендуется обратиться к содержательной, немногословной, но доступной непрофессионалу замечательной книге К. Бержа [16]). Примерами подобных задач служат различные вариации задачи коммивояжера, поиска эйлеровых цепей и циклов (обход всех ребер по одному разу с возвратом или без такового), поиска гамильтоновых путей (обход вершин в ориентированной сети), проектирования транспортных маршрутов без пересечения дорог вне узловых пунктов, разработки структуры тайной организации и многие другие.

Однако существует множество сетевых задач, не укладывающихся в рамки упомянутых выше.

Так в приложениях возникают задачи о *многопродуктовых потоках* в сети с ограниченными пропускными способностями, где сложно установить сам факт возможности транспортировки или минимизировать суммарные затраты даже при линейных соотношениях между затратами и объемами перевозок.

5. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

5.1. Специфика нелинейных программ и методы их решения

Многообразие методов решения линейных программ имеет в своей основе идею упорядоченного перебора опорных планов (вершин множества планов) исходной или сопряженной задачи. Для нелинейных же программ простой метод решения, подобный симплексному, отсутствует по ряду причин.

Во-первых, множество планов может оказаться невыпуклым или иметь бесконечное количество "вер-

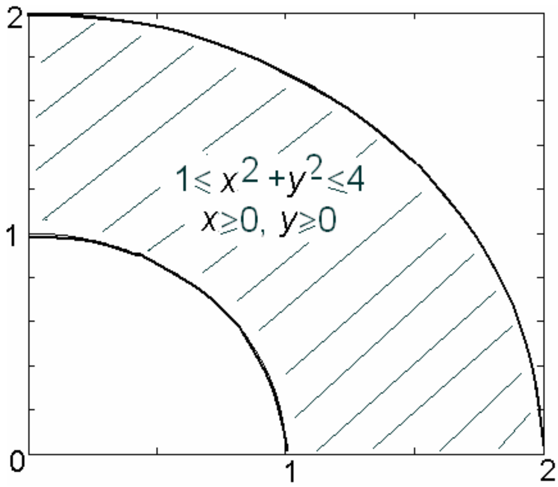


Рис. 1

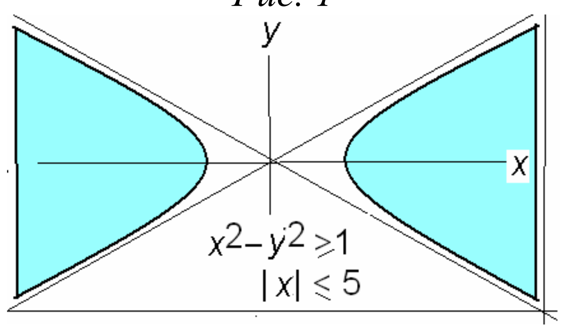


Рис. 2

шин" (рис. 1). Хуже того, оно может оказаться несвязным (рис. 2).

Во-вторых, искомые экстремумы могут достигаться как на границе, так и внутри множества планов.

В-третьих, в нелинейных программах может возникнуть проблема поиска **глобального** экстремума среди множества **локальных** (рис. 3)

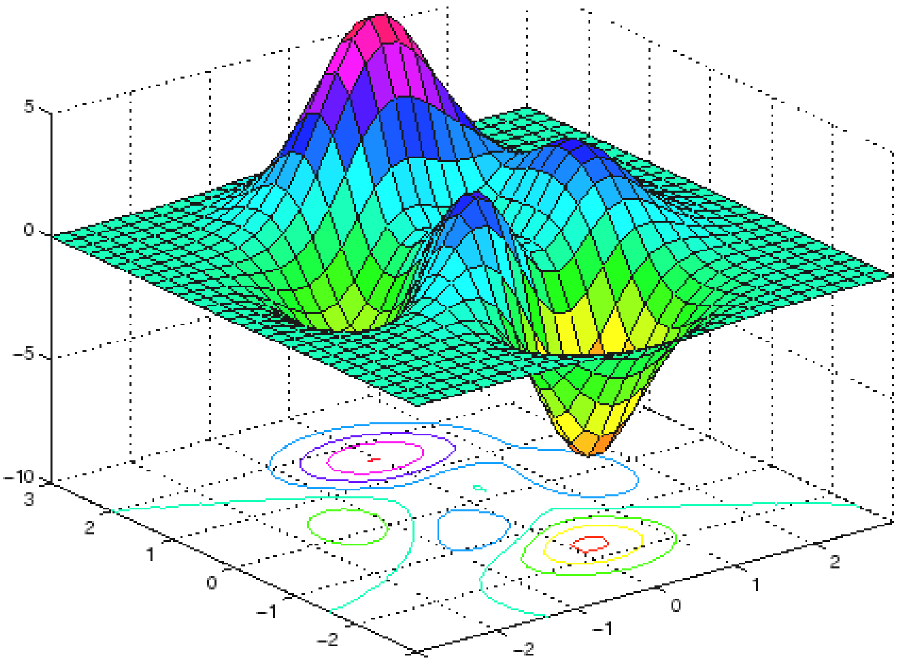


Рис. 3

Как мы показали ранее, использование аппарата производных или прямое табулирование целевой функции над множеством планов не решают проблему в случае более трех переменных. Поэтому каждая нелинейная программа требует индивидуального подхода, учитывающего ее вающего ее специфику.

Существующие методы нелинейного программирования можно подразделить на следующие основные классы.

1. *Градиентные методы*, в основе которых лежит свойство градиента функции в точке (вектора частных производных, вычисленного в точке) как указателя направления наибольшего роста функции в окрестности точки.

Так для удовлетворения желания найти точку минимума $F(X)$ при отсутствии ограничений одна из простейших разновидностей градиентных методов – *метод наискорейшего спуска* – предлагает выбрать некоторую точку (план) X_k и начальный шаг H_k, вычислить градиент функции в выбранной точке $grad\ F(X_k)$ и осуществить переход в направлении, обратном градиенту, с выбранным шагом. Если значение функции в новой точке меньше предыдущего, новая точка принимается за исходную и повторяется аналогичное действие. При попадании в точку с бóльшим значением шаг уменьшается (например, вдвое) и переход повторяется от предыдущей точки. Такие действия продолжаются до получения достаточно малого шага.

Например, если нужно решить систему уравнений

$$X^3 + Y^3 = 6, \qquad Y = e^{-X},$$

то можно заменить эту задачу задачей минимизации функции

$$F(X,Y) = [X^3 + Y^3 - 6]^2 + [Y - e^{-X}]^2$$

(если система имеет решение, то искомый минимум равен нулю).

Градиент этой функции определяется вектором

$$grad\, F(X,Y) = \{2\cdot[X^3 + Y^3 - 6]\cdot 3X^2 + 2\cdot[Y - e^{-X}]\, e^{-X},$$
$$2\cdot[X^3 + Y^3 - 6]\cdot 3Y^2 + 2\cdot[Y - e^{-X}]\}.$$

Выбираем начальную точку $M_0(2,0)$ и шаг h = 0.1. Здесь значение $F(M_0) \approx 4.0183$, градиент в этой точке $grad\, F(M_0)$ = [47.96, –0.27], а нормированный градиент (вектор единичной длины, составленный из компонент, деленных на корень из суммы их квадратов) $grad_н F(M_0)$ = [0.9999, –0.0056]. Смещаемся в точку $M_1 = M_0 - h\cdot grad_н F(M_0)$ = (1.90, 0.0056) и обнаруживаем, что $F(M_1) \approx 0.7601 < F(M_0)$. Из этой точки согласно $grad_н F(M_1)$ = [0.9999, –0.0161] смещаемся в точку M_2=(1.8000, 0.0022), $F(M_2) \approx 0.0548$ и т. д. до точки с координатами ~ (1.8165, 0.1626).

Существуют и более эффективные переходы по градиенту, связанные с выбором различного шага по разным координатам или с автоматическим определением шага (при каждом переходе решается задача поиска экстремума функции в заданном направлении). Гарантии же найти именно глобальный экстремум нет (при разных начальных данных для многоэкстремальных функций получаем разные решения).

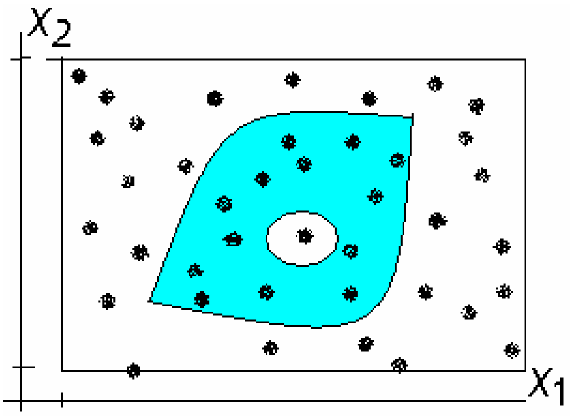
Градиентные методы для задач с ограничениями, где при смещениях по градиенту приходится сталкиваться с опасностью “выскочить” за пределы допустимого множества решений, существенно усложняются (модифицированный метод Ньютона, методы возможных направлений Зойтендейка, сопряженных градиентов, проектируемых градиентов Розена и др.).

Существует обширная литература по численному анализу, где значительное внимание уделяется градиентным и другим итерационным методам, но тем не менее решение нелинейных задач оптимизации при наличии ограничений почти всегда затруднительно.

2. *Методы Монте-Карло*. Здесь отыскивается n - мерный параллелепипед $\{a_i \le x_i \le b_i,\ i=1,..,n\}$, включающий в себя множество планов, и затем моделируются координаты N случайных точек с равномерным законом распределения в параллелепипеде (практически во всех программных средах предусмотрено наличие таких датчиков псевдослучайных чисел $\in(0,1)$). Для точек, попавших во множество

планов, вычисляются значения функции и запоминается точка текущего экстремума.

После этого, для достижения бо́льшей точности, берется параллелепипед меньших размеров с центром в найденной точке, и в нем вновь моделируются N случайных точек. Процесс такого *стохастического моделирования* заканчивается при малых размерах параллелепипеда.



Для отыскания приемлемого значения N можно воспользоваться законом больших чисел и прийти к оценке $N \sim 1/(4\varepsilon^2\delta)$, где ε – допустимая абсолютная погрешность, δ – вероятность ошибки. Если задаться погрешностью 0.001 и вероятностью ошибки 0.01, то имеем «фантастическое» значение N=25·10^6! Однако методы Монте-Карло имеют то преимущество над моделированием на детерминированной сетке, что точность получаемых оценок имеет порядок $1/\sqrt{N}$ и не зависит от размерности задачи (при $n > 3$ их эффективность несомненна). Естественно, эти методы никто не применяет при ручном счете, но они просты для программной реализации и часто используются при поиске начального приближения для градиентных методов.

3. *Методы динамического программирования*, сводящие многомерную задачу оптимизации к последовательности задач меньшей размерности. Их применение особенно успешно в случае *сепарабельных функций*, т.е. функций, представимых суммой функций одной переменной $F(X_1, X_2, .., X_n) = f_1(X_1) + f_2(X_2) + ... + f_n(X_n)$.

4. *Методы выпуклого программирования*, реализующие поиск минимума выпуклой (максимума вогнутой) функции на выпуклом множестве планов. Если множество планов – выпуклый многогранник, то эти методы допускают использование симплексного метода. В основе этих методов лежат т. н. двойственные оценки и теорема Куна-Таккера (понятие о двойственности в линейном программировании – частный случай).

5.2. Дробно-линейное программирование

Пусть стоит задача максимизации дробно-линейной функции

$$Q(X)=\frac{C_0+\sum\limits_{j=1}^{n} C_j X_j}{D_0+\sum\limits_{j=1}^{n} D_j X_j} \tag{1}$$

при линейных ограничениях

$$\sum_{j=1}^{n} A_{ij} X_j = B_i \qquad (i = 1 \,..\, m) \tag{2}$$

$$X_j \ge 0 \qquad (j = 1 \,..\, n) \tag{3}$$

Предположим, что знаменатель в (1) положителен при всех X, удовлетворяющих (2) – (3). Тогда при вводе обозначений

$$D_0 + \sum_{j=1}^{n} D_j X_j = \frac{1}{R} \; (R > 0) , \quad Z_j = R \cdot X_j , \; j = 1 \,..\, n \tag{4}$$

задача сведется к линейной программе максимизации

$$Q(Z,R) = C_0 R + \sum_{j=1}^{n} C_j Z_j \tag{5}$$

при условиях

$$D_0 R + \sum_{j=1}^{n} D_j Z_j = 1 \quad , \tag{6}$$

$$-B_i R + \sum_{j=1}^{n} A_{ij} Z_j = 0 \quad (i = 1 \,..\, m) , \tag{7}$$

$$R > 0 \quad , \quad Z_j \ge 0 \qquad (j = 1 \,..\, n) \quad . \tag{8}$$

Рассмотрим в качестве примера задачу максимизации дробно-линейной функции

$$\frac{-3+2X_1+4X_2-5X_3}{5+3X_1-X_2}$$

при условиях

$$\begin{array}{l} X_1 - X_2 \ge 0 \\ 5X_1+3X_2+10X_3 \le 15 \\ X_1 , X_2 , X_3 \ge 0 \end{array} .$$

С учетом (4) получаем задачу максимизации

$$-3R + 2Z_1 + 4Z_2 - 5Z_3$$

при условиях

$$\begin{array}{rl} 5R + 3Z_1 - Z_2 & = 1 \\ Z_1 - Z_2 & \ge 0 \\ -15R + 5Z_1 + 3Z_2 + 10Z_3 & \le 0 \\ R > 0 , \quad Z_1 , Z_2 , Z_3 & \ge 0 \end{array} .$$

C баз	Базис	План *Z*	–3	2	4	–5	0	0	–M	–M
			R	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7
–M	Z_6	1	5	3	–1	0	0	0	1	0
–M	Z_7	0	0	1	–1	0	–1	0	0	1
0	Z_5	0	–15	5	3	10	0	1	0	0
Δ*k*		–M	–5M	–3M	2M	5	M	0	0	0

C баз	Базис	План *Z*	– 3	2	4	– 5	0	0	–M
			R	Z_1	Z_2	Z_3	Z_4	Z_5	Z_7
–3	R	1/5	1	3/5	–1/5	0	0	0	0
–M	Z_7	0	0	1	–1	0	–1	0	1
0	Z_5	3	0	14	0	10	0	1	0
Δ*k*		–3/5	0	–M..	M+	5	M	0	0

C баз	Базис	План *Z*	–3	2	4	– 5	0	0
			R	Z_1	Z_2	Z_3	Z_4	Z_5
– 3	R	1/5	1	0	2/5	0	3/5	0
2	Z_1	0	0	1	–1	0	–1	0
0	Z_5	3	0	0	14	10	14	1
Δ*k*		–3/5	0	0	–7.2	5	-3.8	0

C баз	Базис	План *Z*	– 3	2	4	–5	0	0
			R	Z_1	Z_2	Z_3	Z_4	Z_5
– 3	R	4/35	1	0	0	–2/7	1/5	–1/35
2	Z_1	3/14	0	1	0	5/7	0	1/14
4	Z_2	3/14	0	0	1	5/7	1	1/14
Δ*k*		33/35	0	0	0	1/7	17/5	13/35

Отсюда $Z_1 = Z_2 = 3/14$, $Z_3 = 0$, $R = 4/35$ и соответственно

$$X_{opt} = (15/8,\ 15/8,\ 0)\ ,\quad max\ Q(X) = 33/35\ .$$

Следует заметить, что дробно-линейные программы представляют достаточно большой интерес для приложений; например, некорректная задача, преследующая взаимно противоположные цели типа максимизации прибыли при минимальных капиталовложениях, становится корректной, если поставить целью максимум прибыли на единицу затрат или минимум затрат на единицу прибыли.

5.3. Метод множителей Лагранжа

Приводимый здесь метод был предложен еще в 1797 г. Ж. Лагранжем для задачи поиска экстремумов функции $F(X)$ при условиях $f_i(X) = 0$ ($i = 1 .. m$).

Функция $\Phi(X,\lambda) = F(X) + \sum_{i=1}^{m}\lambda_i f_i(X)$ называется *функцией Лагранжа* и коэффициенты λ_i – *множителями Лагранжа.*

Можно доказать, что *необходимым условием экстремума исходной задачи является обращение в нуль всех частных производных функции Лагранжа.*

Например, при поиске экстремальных значений $F(X_1, X_2) = X_1 + X_2$ при единственном условии $X_1^2 + X_2^2 = 1$ строится функция Лагранжа

$$\Phi(X,\lambda) = X_1 + X_2 + \lambda\,(X_1^2 + X_2^2 - 1)$$

Строим систему уравнений

$$\frac{\partial\Phi}{\partial X_1} = 1 + 2\lambda X_1 = 0,\quad \frac{\partial\Phi}{\partial X_2} = 1 + 2\lambda X_2 = 0\ ,\quad \frac{\partial\Phi}{\partial\lambda} = X_1^2 + X_2^2 - 1 = 0,$$

решение которой дает $\lambda = \pm\, 1/\sqrt{2}$, $X_1 = X_2 = -\lambda$ и экстремальные значения целевой функции $-\sqrt{2}$ и $\sqrt{2}$.

При поиске параметров цилиндрической бочки, обеспечивающих максимум ее объема $V = \pi R^2 H$ при фиксированной площади ее оболочки $S = 2\pi R^2 + 2\pi RH$, берем функцию Лагранжа в виде $\Phi(R, H, \lambda) = \pi R^2 H + \lambda\cdot[2\pi R^2 + 2\pi RH - S]$ и приходим к системе уравнений

$$\frac{\partial\Phi}{\partial R} = 2\pi RH + \lambda\cdot 2\pi(2R + H) = 0;\quad \frac{\partial\Phi}{\partial H} = \pi R^2 + \lambda\cdot 2\pi R = 0;$$

$$\frac{\partial\Phi}{\partial\lambda} = 2\pi R^2 + 2\pi RH - S = 0,$$

решение которой дает $\lambda = -R/2$ и искомые оценки $H = 2R$, $R = \sqrt{S/6\pi}$.

Для определения типа найденного экстремума (*правило Сильвестра*) можно построить матрицу вторых производных $F(X)$, вычисленных в экстремальной точке, и определить знаки главных ее миноров. Если все они положительны, то найден минимум; если они чередуются, начиная с минуса, то найден максимум.

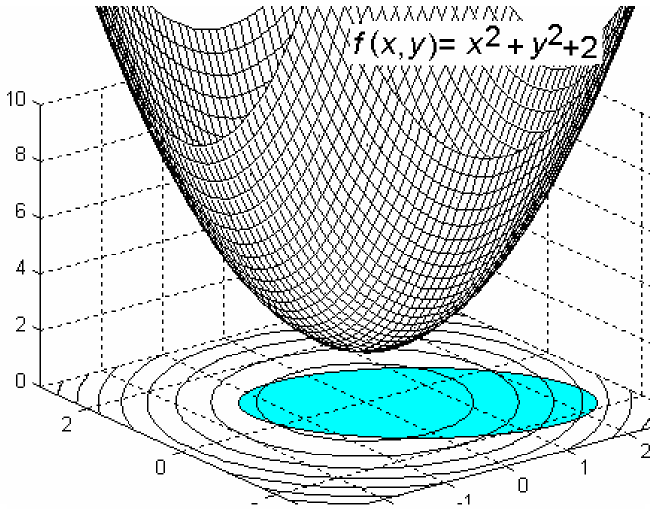
Простота метода множителей Лагранжа является кажущейся, поскольку он обычно приводит к нелинейной системе уравнений и не гарантирует тип отыскиваемого экстремума, кроме глобальных дает и множество локальных экстремумов. Тем не менее он полезен как база генерации идей и методов нелинейного программирования.

5.4. Теорема Куна-Таккера

Пусть стоит задача минимизации $F(X)$ при условиях :

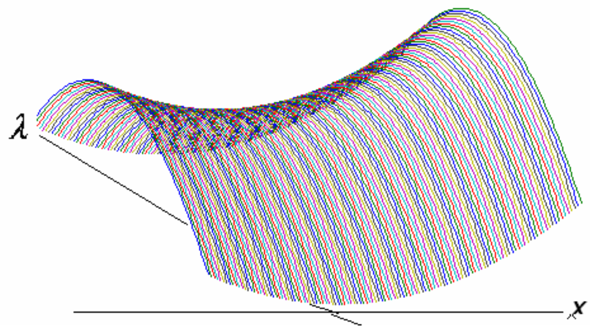
$f_i(X) \le 0$ ($i = 1\,..\,m$), (1)

$X \ge 0$, (2)



где X – n-мерный вектор, $F(X)$ и $f_i(X)$ – выпуклые функции, что обеспечивает выпуклость множества планов и единственность искомого минимума (мы считаем функцию выпуклой в некоторой точке, если главные миноры матрицы вторых производных положительны).

Введем функцию Лагранжа



$$\Phi(X,\lambda) = F(X) + \sum_{i=1}^{m} \lambda_i\, f_i(X)\,. \tag{3}$$

Теорема Куна – Таккера утверждает, что *вектор* $X^* \ge 0$ *является решением поставленной задачи тогда и только тогда, когда существует вектор* $\lambda^* \ge 0$ *такой, что* при всех $X \ge 0$, $\lambda \ge 0$

$$\Phi(X^*, \lambda) \le \Phi(X^*,\lambda^*) \le \Phi(X, \lambda^*)\ . \tag{4}$$

Так как функция Лагранжа в точке (X^*, λ^*) принимает минимум по X и максимум по λ, эта точка называется *седловой* и теорему называют *теоремой о седловой точке* или *теоремой о минимаксе*[10].

Достаточность условий этой теоремы доказывается сравнительно просто. Доказательство их необходимости предполагает выполнение *условий регулярности,* т. е. существования хотя бы одной допустимой точки X, где $f_i(X) < 0$ при всех i.

Если $F(X)$ и $f_i(X)$ дифференцируемы, то условия теоремы эквивалентны "локальным" условиям, утверждающим, что в точке (X^*,λ^*)

$$\frac{\partial \Phi}{\partial X} \ge 0\ ,\quad X^T \frac{\partial \Phi}{\partial X} = 0\ ,\quad X \ge 0\ , \tag{5}$$

$$\frac{\partial \Phi}{\partial \lambda} \le 0\ ,\quad \lambda^T \frac{\partial \Phi}{\partial \lambda} = 0\ ,\quad \lambda \ge 0\ . \tag{6}$$

---

[10] Работа над нелинейным программированием и теорией двойственности была начата Алом Такером и его студентами Х. Куном и Д. Джейлом в 1948 г., но, по утверждению Д. Данцига, создателем теоремы о двойственности является Джон фон Нейман, а Таккера и его коллег ценят как авторов строгого доказательства.

Остановимся на частных случаях задачи.

Если в поставленной задаче отсутствуют требования неотрицательности (2), то путем замены X разностью двух неотрицательных векторов можно показать, что (5) упрощается к виду $\frac{\partial \Phi}{\partial X} = 0$ (если нет требования неотрицательности для одной из компонент X, то обратится в нуль производная $\Phi(X)$ по соответствующей переменной).

Если $f_i(X) = 0$ при каком-то i, то заменой этого равенства системой неравенств $f_i(X) \le 0$, $-f_i(X) \le 0$ обнаруживаем, что в (6) соответствующая производная обращается в нуль и исчезает условие неотрицательности по λ_i; если $f_i(X) = 0$ при всех i, то (6) упрощается к виду $\frac{\partial \Phi}{\partial \lambda} = 0$.

Предельным случаем условий Куна – Таккера служат двойственные задачи линейного программирования. Так при минимизации $C^T X$ при условиях $AX \ge B$, $X \ge 0$ функция Лагранжа имеет вид

$$\Phi(X, \lambda) = C^T X + \lambda^T (AX - B)$$

и условия Куна – Таккера

$$\frac{\partial \Phi}{\partial X} = C - A^T \lambda \ge 0\ ,\ X^T \frac{\partial \Phi}{\partial X} = X^T (C - A^T \lambda) = 0\ ,\ X \ge 0$$

$$\frac{\partial \Phi}{\partial \lambda} = B - AX \le 0\ ,\ \lambda^T \frac{\partial \Phi}{\partial \lambda} = \lambda^T (B - AX) = 0,\quad \lambda \ge 0.$$

Отсюда с учетом $X^T C = C^T X$, $\lambda^T B = B^T \lambda$, $X^T A^T \lambda = \lambda^T AX$ получаем, что в седловой точке достигается минимум по X и максимум по λ, причем $C^T X = B^T \lambda$ и

$$AX \ge B\ , \quad A^T \lambda \le C\ ,$$
$$X \ge 0\ , \quad \lambda \ge 0\ .$$

5.5. Квадратичное программирование. Метод Вулфа – Фрэнка

Рассмотрим задачу минимизации квадратичной функции n переменных

$$F(X) = C^T X + X^T DX \tag{1}$$

при линейных ограничениях

$$AX \le B\ ,\ X \ge 0\ , \tag{2}$$

где A – матрица размерности m на n , C , X – n-мерные векторы, B– m-мерный вектор, D –положительно определенная n-мерная квадратная матрица[11] . Так целевая функция

$$F(X) = 7X_1 - 3X_2 + 22X_1^2 - 8X_1X_2 + X_2^2 - 10X_1X_3 + 12X_2X_3 + 40X_3^2$$

---

[11] Матрица называется положительно определенной, если положительны ее главные миноры

представится в виде (1), где

$$C = \begin{Vmatrix} 7 \\ -3 \\ 0 \end{Vmatrix}, \qquad D = \begin{Vmatrix} 22 & -4 & -5 \\ -4 & 1 & 6 \\ -5 & 6 & 40 \end{Vmatrix}$$

Положительная определенность матрицы D и линейность ограничений (выпуклое множество планов) позволяют использовать теорему Куна-Таккера.

Функция Лагранжа здесь имеет вид:

$$\Phi(X, \lambda) = C^T X + X^T DX + \lambda^T (AX - B) \tag{3}$$

и условия Куна-Таккера приводятся к форме:

$$\frac{\partial \Phi}{\partial X} = C + 2DX + A^T \lambda \geq 0, \quad X^T \frac{\partial \Phi}{\partial X} = 0, \quad X \geq 0; \tag{4}$$

$$\frac{\partial \Phi}{\partial \lambda} = AX - B \leq 0, \quad \lambda^T \frac{\partial \Phi}{\partial \lambda} = 0, \quad \lambda \geq 0. \tag{5}$$

Если ввести векторы ослабляющих переменных, то (4) и (5) примут вид:

$$\begin{aligned} &C + 2DX + A^T \lambda \quad - V = 0, \quad X^T V = 0, \; X \geq 0, \; V \geq 0; \\ &AX - B \qquad + Y \qquad = 0, \; -\lambda^T Y = 0, \; \lambda \geq 0, \; Y \geq 0. \end{aligned}$$

С учетом неотрицательности переменных можно поставить эквивалентную задачу *минимизации* (*до нуля*) функции

$$g(X, \lambda, Y, V) = X^T V + \lambda^T Y \tag{6}$$

при условиях

$$2DX + A^T \lambda \quad - V = -C \tag{7}$$

$$AX \qquad\quad + Y \qquad = B \tag{8}$$

$$X, \lambda, Y, V \geq 0. \tag{9}$$

Если обозначить

$$W^T = (X^T, \lambda^T, Y^T, V^T), \tag{10}$$

то (7) $-$(9) можно записать в виде:

$$RW = S, \quad W \geq 0, \tag{11}$$

где

$$R = \begin{Vmatrix} 2D & A^T & 0 & -E \\ A & 0 & E & 0 \end{Vmatrix}, \quad S = \begin{Vmatrix} -C \\ B \end{Vmatrix} \tag{12}$$

Таким образом метод П.Вулфа и М. Фрэнка сводит решение задачи к форме, допускающей применение симплексной процедуры.

Здесь отыскивается некоторый опорный план W^0. Если $g(W^0) = 0$, то этот план оптимален. В противном случае отыскивается градиент

$$grad\, g(W^0) = (V^T, Y^T, \lambda^T, X^T) \tag{13}$$

и его компоненты на один шаг симплексного преобразования (перехода к другому опорному плану) принимаются за коэффициенты «линейной формы». После выбора нового опорного плана выполняются вышеописанные рассуждения.

Пример. Минимизировать

$$F(X_1, X_2) = -4\,X_1\ -6\,X_2\ + X_1^{\,2}\ +3\,X_2^{\,2}$$

при условиях

$$2\,X_1 + \ X_2 \le \ 4\ ,$$

$$X_1\ , X_2 \ge \ 0.$$

Ставим задачу минимизации до нуля функции

$$g(X_1\ ,X_2\ , \lambda\ , Y, V_1\ ,V_2\ ) = X_1\,V_1 + X_2\,V_2\ \ +\lambda Y$$

при условиях (8), где

$$R = \begin{array}{|cc|c|c|cc|} \hline 2 & 0 & 2 & 0 & -1 & 0 \\ 0 & 6 & 1 & 0 & 0 & -1 \\ \hline 2 & 1 & 0 & 1 & 0 & 0 \\ \hline \end{array} \qquad S = \begin{array}{|c|} \hline 4 \\ 6 \\ \hline 4 \\ \hline \end{array}$$

Ищем начальный опорный план методом искусственного базиса (не обращая внимания на коэффициенты, меньшие М):

C баз	Базис	План W							M	M
			X_1	X_2	λ	Y	V_1	V_2	Z_1	Z_2
M	Z_1	4	2	0	2	0	–1	0	1	0
M	Z_2	6	0	6	1	0	0	–1	0	1
	Y	4	2	1	0	1	0	0	0	0
Δ_k		10M	2M	6M	3M	0	–M	-M	0	0

C баз	Базис	План W							M
			X_1	X_2	λ	Y	V_1	V_2	Z_1
	λ	2	1	0	1	0	–1/2	0	1/2
M	Z_2	4	–1	6	0	0	1/2	–1	–1/2
	Y	4	2	1	0	1	0	0	0
Δ_k		M	–M	6M	0	0	M/2	–M	0

C баз	Ба-зис	План W	X_1	X_2	λ	Y	V_1	V_2
	λ	6	0	6	1	0	0	-1
	V_1	8	-2	12	0	0	1	-2
	Y	4	2	1	0	1	0	0
	Δ_k							

Итак, найден начальный опорный план исходной задачи

$$X_1 = X_2 = V_2 = 0\ , V_1 = 8\ , Y = 4\ , \lambda = 6,$$

для которого $g(W) = 0{\times}8 + 0{\times}0 + 6{\times}8 = 48 > 0$.

Находим градиент $grad\ g(W) = \{8, 0, 4, 6, 0, 0\}$ и берем его компоненты в качестве коэффициентов «линейной формы» (нормали к поверхности, определяемой функцией $g(W)$, в выбранной точке).

C	Базис	План	8	0	4	6	0	0

баз		W	X_1	X_2	λ	Y	V_1	V_2
4	λ	6	0	6	1	0	0	–1
0	V_1	8	–2	12	0	0	1	– 2
6	Y	4	2	1	0	1	0	0
Δ_k		48	4	30	0	0	0	–4

C баз	Базис	План W	12	0	0	6	2	0
			X_1	X_2	λ	Y	V_1	V_2
0	λ	6	0	6	1	0	0	–1
2	V_1	12	0	13	0	1	1	– 2
12	X_1	2	1	1/2	0	1/2	0	0
Δ_k		48	0	32	0	2	0	– 2

C баз	Базис	План W	0	0	0	7/13	20/13	12/13
			X_1	X_2	λ	Y	V_1	V_2
0	λ	**7/13**	0	0	1	–6/13	–6/13	–1/13
0	X_2	**12/13**	0	1	0	1/13	1/13	–2/13
0	X_1	20/13	1	0	0	6/13	–1/26	1/13
Δ_k		**0**						

Получен оптимальный план $X = (20/13\ , 12/13)$ – точка, лежащая на границе множества планов.

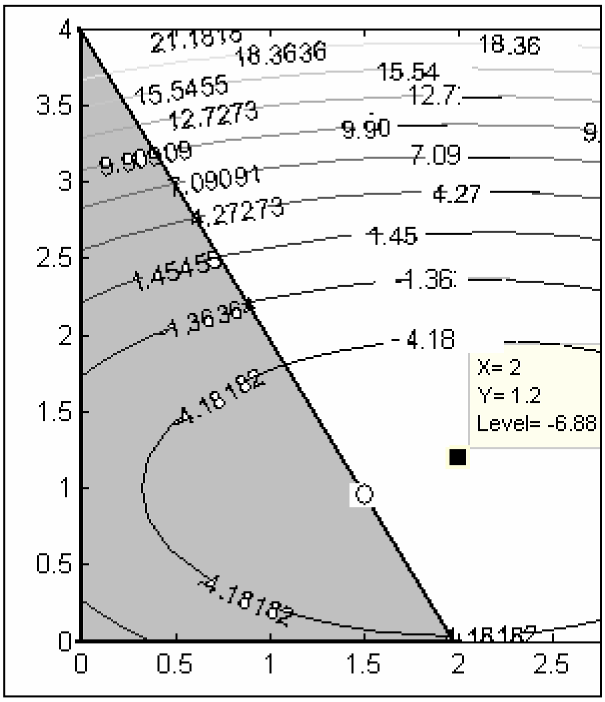
Если отвлечься от ограничений, то минимум рассматриваемой функции $F(X_1, X_2) = -\ 4\ X_1 - 6\ X_2\ + X_1^{\ 2}\ +3\ X_2^{\ 2}$ достигается в точке $X = (2,1)$ за пределами множества планов.

На приведенном ниже рисунке, полученном с помощью процедуры *contour* в среде MatLab, представлены линии уровня $F(X_1, X_2)=$ *const* и наглядно видно взаимное расположение точек минимума с учетом ограничений и без такового.

Заметим, что при ограничениях $AX = B$, $X \geq 0$ исчезнет переменная Y, снимутся условия неотрицательности на λ и потребуется минимизировать X^TV при условиях:

$$2DX + A^T\lambda - V = -C$$
$$AX = B$$
$$X, V \geq 0 .$$

При других отклонениях постановки задачи от выбранного здесь стандарта можно элементарными приемами прийти к нему (вместо максимизации $F(X)$ искать минимум $-F(X)$, условие $AX \geq B$ заменить на $B - AX \leq 0$ и т.п.).



5.6. Геометрическое программирование

Формулировка и идеология т. н. задачи геометрического программирования [23] принадлежат американским математикам Р. Дж. Даффину и Э. Л. Питерсону и специалисту по электротехнике К. Зенеру.

Прежде чем приступать к рассмотрению идеологии геометрического программирования (ГП), введем понятие позинома.

Функция

$$g(x_1, x_2, ..., x_n) = \sum_{k=1}^{m} c_k x_1^{\alpha_{1k}} x_2^{\alpha_{2k}} ... x_n^{\alpha_{nk}} , \qquad (1)$$

где $x_1>0$, .., $x_n>0$; $c_1>0$, .., $c_m>0$; a_{ik} – произвольные действительные числа, называ*ется позиномом* n-го порядка (каждое из слагаемых называют *одночленным позиномом*).

Задача геометрического программирования может быть сформулирована в форме:

найти **наименьшее значение** позинома $g_0(x_1, .., x_n)$ при условиях положительности значений переменных

$$x_1>0, .., x_n>0. \qquad (2)$$

и ограничениях (их называют *вынужденными*)

$$g_i(x_1, .., x_n) \leq 1 , \quad i=1, …, p \qquad (3)$$

где все функции $g_i(x_1, .., x_n)$ – позиномы.

Класс позиномов достаточно широк и ими описывается множество известных закономерностей в технике, экономике и других сферах. Так объем цилиндра $v=\pi r^2 h$ и сила тока $I=UR^{-1}$ представляются одночленными позиномами 2-го порядка; в принципе даже линейную функцию с положительными коэффициентами можно рассматривать как позином, поскольку например $x_3= x_1^0 \cdot x_2^0 \cdot x_3^1 \cdot x_4^0$ … В принципе любое линейное неравенство, например $2 \cdot x_1 - 3 \cdot x_2 + 4 \cdot x_3 - 1 \le 0$, заменой переменных $x_i=ln(z)$ и возведением экспоненты в степень преобразуется к одночленному позиному третьего порядка $0.3679 \cdot z_1^2 \cdot z_2^{-3} \cdot z_3^4 \le 1$.

Разумеется, на практике подобными преобразованиями можно заниматься лишь «из любви к искусству», ибо проще линейных функций в математике ничего нет.

Прежде чем переходить к проблематике решения поставленной задачи, вспомним известное из школьной математики утверждение: «среднее арифметическое положительных величин не меньше их среднего геометрического»

$$\tfrac{1}{m}(u_1 + u_2 + \ldots + u_m) \ge \sqrt[m]{u_1 \cdot u_2 \cdot \ldots \cdot u_m} \ .$$

Его обобщением может служить неравенство

$$\sum_{k=1}^{m} \delta_k u_k \ge \prod_{k=1}^{m} u_k^{\delta_k} \ , \tag{4}$$

где $u_k \ge 0$ и $\delta_k > 0$ – произвольные числа, $\sum\limits_{k=1}^{m} \delta_k = 1$.

Неравенство (4) можно преобразовать и к другой форме

$$\sum_{k=1}^{m} u_k \ge \prod_{k=1}^{m} \left(\frac{u_k}{\delta_k} \right)^{\delta_k} \tag{4a}$$

или в развернутом виде

$$u_1 + u_2 + \ldots + u_m \ge \left(\frac{u_1}{\delta_1} \right)^{\delta_1} \left(\frac{u_2}{\delta_2} \right)^{\delta_2} \ldots \left(\frac{u_m}{\delta_m} \right)^{\delta_m}$$

Обратившись к (1) и обозначив ее слагаемые

$$u_k = c_k x_1^{\alpha_{1k}} x_2^{\alpha_{2k}} \ldots x_n^{\alpha_{nk}} \ ,$$

получаем правую часть (4а) в виде т.н. *преддвойственной* функции

$$V(\delta, x) = \prod_{k=1}^{m} \left(\frac{c_k}{\delta_k} \right)^{\delta_k} \prod_{j=1}^{n} x_j^{\alpha_{j1}\delta_1 + \alpha_{j2}\delta_2 + \ldots + \alpha_{jm}\delta_m} \tag{5}$$

Если подобрать ненулевые (положительные) значения δ_k так, чтобы степени величин x_j обратились в нуль, то есть найти соответствующее решение системы $n + 1$ уравнений с m неизвестными (!!)

$$\alpha_{j1} \delta_1 + \alpha_{j2} \delta_2 + \ldots + \alpha_{jm} \delta_m = 0 \ , \ j=1,\ldots,n; \tag{6}$$

$$\delta_1 + \delta_2 + ... + \delta_m = 1 ,$$

преддвойственная функция превратится в независимую от x *двойственную функцию*

$$v(\delta) = \prod_{k=1}^{m} \left(\frac{c_k}{\delta_k} \right)^{\delta_k} \tag{7}$$

с заведомо положительным значением и обнаруживается неравенство $g(X) \geq v(\delta)$, из которого следует принципиальное существование искомого минимума функции $g(X)$. Условия (6) авторы [23] называют *условиями ортогональности и нормировки*.

Следует сразу обратить внимание на то, что решение системы (6) создает значительные принципиальные или технические трудности, связанные с соотношением между размерностью исходной задачи n и числом слагаемых m в минимизируемом позиноме.

Возьмем в качестве примера задачу [23] минимизации функции

$$g_0(X) = \frac{40}{x_1 x_2 x_3} + 40 x_2 x_3 + 20 x_1 x_3 + 10 x_1 x_2 .$$

Двойственная функция здесь выступает в форме

$$v(\delta) = \left(\frac{40}{\delta_1} \right)^{\delta_1} \cdot \left(\frac{40}{\delta_2} \right)^{\delta_2} \cdot \left(\frac{20}{\delta_3} \right)^{\delta_3} \cdot \left(\frac{10}{\delta_4} \right)^{\delta_4} \tag{7a}$$

и поиск δ_k осуществляется из системы уравнений:

(коэффициенты при x_1)	$-\delta_1$		$+\delta_3$	$+\delta_4$	$=0$
(коэффициенты при x_2)	$-\delta_1$	$+\delta_2$		$+\delta_4$	$=0$
(коэффициенты при x_3)	$-\delta_1$	$+\delta_2$	$+\delta_3$		$=0$
	δ_1	$+\delta_2$	$+\delta_3$	$+\delta_4$	$=1$

В данном случае решение системы однозначно (считайте, что вам повезло $m = n+1 = 4$!):

$$\delta_1 = 2/5 ,\ \delta_2 = 1/5 ,\ \delta_3 = 1/5 ,\ \delta_4 = 1/5.$$

Подставляя найденное решение в двойственную функцию (7а), получаем $v(\delta)=100$ и можем гарантировать выполнение условия $g_0(X) \geq 100$, то есть существование искомого минимума.

Как мы увидим из дальнейшего, минимум $g_0(X^*) = v(\delta)=100$ и справедливо равенство $u_k = \delta_k \cdot v(\delta)$, откуда возникает нелинейная (!) система

$$\frac{40}{x_1 x_2 x_3} = \frac{2}{5} \cdot 100 ; \qquad 40 x_2 x_3 = \frac{1}{5} \cdot 100 ;$$

$$20 x_1 x_3 = \frac{1}{5} \cdot 100 ; \qquad 10 x_1 x_2 = \frac{1}{5} \cdot 100 ,$$

прологарифмировав которую и решив возникающую систему m=4 линейно зависимых уравнений с n=3 неизвестными, получаем решение исходной задачи x_1=2, x_2=1, x_3=0.5.

Вернемся к общему случаю задачи минимизации позинома $g_0(x_1, .., x_n)$ при ограничениях x_j>0 (i=1…n) , $g_k(x_1, .., x_n)$≤1 (k= 1,.., p), где все функции $g_k(x_1, .., x_n)$ – позиномы .

Пусть m_k – количество членов в позиноме g_k $(x_1, .., x_n)$ и $\sum_{k=0}^{p} m_k = m$.

Из соображений компактности последующих формулировок, введем символику для следующих множеств натуральных чисел:

$$
\begin{aligned}
&J_0 = \{1,2,..., m_0\}, \\
&J_1 = \{m_0 + 1, m_0 + 2,..., m_0 + m_1\}, \\
&J_p = \left\{ \sum_{i=0}^{p-1} m_i + 1, \sum_{i=0}^{p-1} m_i + 2,..., \sum_{i=0}^{p-1} m_i + m_p \right\}.
\end{aligned}
\tag{8}
$$

Объединение всех множеств J_k дает последовательность натуральных чисел {1… m}, называемую в литературе *неприводимым* множеством.

Если перенумеровать все члены позиномов числами из соответствующих множеств, то каждый позином можно теперь записать в эквивалентном виде

$$g_k(x_1, x_2,..., x_n) = \sum_{i \in J_k} c_k\, x_1^{\alpha_{1i}} x_2^{\alpha_{2i}} ... x_n^{\alpha_{ni}} \tag{9}$$

Естественно, что все коэффициенты c_k>0 (k=1, 2,..., m), а показатели степени – произвольные действительные числа. Совокупность этих показателей может быть представлена т. н. *матрицей экспонент*, составленной из n строк (по числу переменных величин задачи) и m столбцов (по общему числу слагаемых в позиномах).

Двойственная программа состоит в поиске максимального значения функции

$$V(\delta) = \left[\prod_{k=1}^{m} \left(\frac{c_k}{\delta_k} \right)^{\delta_k} \right] \prod_{s=1}^{p} \lambda_s(\delta)^{\lambda_s(\delta)} \tag{\textbf{10}}$$

где
$$\lambda_s(\delta) = \sum_{j \in J(s)} \delta_j\,, \; s=1,2,\dots,p\,, \tag{11}$$

$$\delta_k \geq 0 \;\; (k=1,2,\dots,m), \tag{12}$$

$$\delta_1 + \delta_2 + ... + \delta_{m0} = 1\,, \tag{13}$$

$$\sum_{i=1}^{m} \alpha_{ij} \delta_i = 0\,, \quad j=1,2,..,n\,. \tag{14}$$

Следует заметить, что максимизируемая функция $V(\delta)$ является непрерывной в области значений δ, поскольку даже при $\delta = 0$ соблюдается предельное $\delta^{\pm\delta} = 1$.

Если система (1 – 2) ограничений прямой задачи совместна (не противоречива, имеет хотя бы одно решение) и к тому же существует хотя бы одна точка $X(x_1, .., x_n)$, где условия (2) выполняются как строгие неравенства $g_1(X) < 1, \ldots, g_p(X_n) < 1$, ее называют *сильно совместной*.

Не останавливаясь на обосновании приведенной формулировки двойственной задачи, приведем следующие фундаментальные теоремы.

Первая теорема двойственности.

Пусть ограничения прямой задачи **сильно совместны** и существует точка X^*, в которой достигается минимум целевой функции $g_0(X)$ (глобальный или локальный). Тогда :

(I) условия (12)–(14) двойственной задачи совместны и существует точка δ^*, в которой достигается максимум целевой функции $V(\delta)$ (глобальный или локальный), причем $g_0(X^*) = V(\delta^*)$;

(II) если X^* – оптимальное решение прямой задачи, существуют множители Лагранжа $\mu^* \geq 0$ такие, что для *функции Лагранжа*

$$L(X,\mu) = g_0(X) + \sum_{k=1}^{p} \mu_k \left[g_k(X) - 1\right], \tag{15}$$

имеет место

$$L(X^*,\mu) \leq g_0(X^*) = L(X^*,\mu^*) \leq L(X,\mu^*) \tag{16}$$

для всех $X>0$ и $\mu \geq 0$. Кроме того, оптимальное решение двойственной задачи имеет вид

$$\delta_i^* = \begin{cases} c_i \prod\limits_{s=1}^{n} x_s^{\alpha_s^{is}} \Big/ g_0(X), & i \in J(0) \\ \mu_k \cdot c_i \prod\limits_{s=1}^{n} x_s^{\alpha_s^{is}} \Big/ g_0(X), & i \in J(k),\ k = 1,..,p \end{cases} \tag{17}$$

(здесь $X=X^*$, $\mu=\mu^*$) и

$$\lambda_k(\delta^*) = \mu_k^* / g_0(X^*), \qquad k = 1,.., p\ ; \tag{18}$$

(III) для оптимального решения двойственной задачи δ^* любая точка минимума X^* прямой задачи удовлетворяет системе

$$c_i \prod_{s=1}^{n} x_s^{\alpha_s^{is}} = \begin{cases} \delta_i^* \cdot V(\delta^*), & i \in J(0) \\ \delta_i^* / \lambda_k(\delta^*), & i \in J(k) \end{cases} \tag{19}$$

где k принимает все положительные значения такие, что $\lambda_k(\delta^*)>0$.

Вторая теорема двойственности.

Если ограничения прямой задачи **совместны** и существует точка δ^* с положительными компонентами, удовлетворяющая ограничениям двойственной задачи, то существует точка X^*, удовлетворяющая ограничениям прямой задачи, в которой достигается минимум целевой функции $g_0(X)$ (глобальный или локальный).

Обе теоремы, равно как и формулировку пары двойственных задач, можно получить из условий теоремы Куна – Таккера, хотя и не с тем изяществом, как для задач линейного программирования.

В отличие от симплексной процедуры для линейной программы, при решении задачи ГП может возникнуть ряд отнюдь не тривиальных подзадач, в частности связанных с т. н. *коэффициентом трудности* $m-n-1$.

Так при минимизации $g_0(x, y) = y$ при ограничениях $x, y>0$

$$g_1(x,y) = \frac{x^4}{y^4} + \frac{y^{1/2}}{x} \le 1$$

Здесь число переменных $n = 2$, число вынужденных ограничений $p = 1$, $m_0 = 1$, $m_1 = 2$ и соответственно $m = m_0 + m_1 = 3$, $J_0 = \{1\}$, $J_1 = \{2,3\}$, степень трудности $t = m-n-1 = 0$, вектор коэффициентов $C = \{1,1,1\}$ и матрица степеней при неизвестных ($x_1 = x$, $x_2 = y$)

$$A = \begin{array}{c|c|cc|c} & J_0 & J_1 & & \\ \hline & 0 & 4 & -1 & x_1 \\ & 1 & -4 & 0{,}5 & x_2 \end{array}$$

Следующий этап – постановка двойственной задачи (10)–(14). Для нашего примера максимизируемая целевая функция имеет вид

$$V(\delta) = \left(\frac{1}{\delta_1}\right)^{\delta_1} \left(\frac{1}{\delta_2}\right)^{\delta_2} \left(\frac{1}{\delta_3}\right)^{\delta_3} (\delta_2 + \delta_3)^{\delta_2 + \delta_3}$$

и ее множество планов с учетом А определяется положительными (в пределе неотрицательными) решениями системы

$$\begin{array}{lll} & 4\,\delta_2 - \delta_3 & = 0 \\ \delta_1 - & 4\,\delta_2 + 0{,}5\delta_3 & = 0 \\ \delta_1 & & = 1 \end{array}$$

Поскольку степень трудности решаемой задачи равна нулю, решение этой системы можно попытаться найти традиционным методом Гаусса (последовательных исключений) или симплексным методом, гарантирующим неотрицательные решения.

Для нашего примера $\delta = [1,\ 0.5,\ 2]$ и $V(\delta) = 3.4939$.

Получив эту оценку, обращаемся к (19) для поиска решения исходной задачи.

Так в нашем примере возникает линейно зависимая система

$$x_2=\delta_1\cdot V(\delta)=3.4939\,,\qquad {x_1^4}\Big/{x_2^4}=\left(\frac{\delta_2}{\delta_2+\delta_3}\right)=\frac{1}{5}\,,\qquad {\sqrt{x_2}}\Big/{x_1}=\left(\frac{\delta_3}{\delta_2+\delta_3}\right)=\frac{4}{5}\,,$$

откуда легко найти $x_1= 2.34$, $x_2 = 3.49$. В общем же случае для линеаризации решаемой системы можно прибегнуть к логарифмированию. Если обозначить $z_j = \log_{10}(x_j)$, то получаем систему, матрица которой совпадает с транспонированной матрицей степеней

$$\begin{array}{rrl} & z_2 & = 0.5433 \\ 4z_1 & -\ 4z_2 & = -0.6990 \\ -z_1 & +0.5z_2 & = -\ 0.0969 \end{array}$$

В нашем случае решение этой (линейно зависимой) системы единственно $z_1 = \log_{10}(x_1) = -0.0198$, $z_2= \log_{10}(x_2) = -0.2334$, и однозначно получается приведенное выше решение поставленной задачи.

Если же степень трудности решаемой задачи ненулевая, то на пути решения встают значительные трудности. Так решая более простую, на первый взгляд, задачу минимизации [23]

$$g_0(X)=\frac{40}{x_1x_2}+40x_2+20x_1+10x_1x_2$$

при отсутствии вынужденных ограничений, имеем $n = 2$, $p = 0$, $m = 4$, C=[40,40,20,10], матрицу степеней

$$\begin{array}{c} J_0 \\ \mathrm{A}=\left[\begin{array}{ccc|c} -1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \end{array}\right]\begin{array}{l} x_1 \\ x_2 \end{array} \end{array}$$

и получаем задачу максимизации

$$V(\delta)=\left(\frac{40}{\delta_1}\right)^{\delta_1}\left(\frac{40}{\delta_2}\right)^{\delta_2}\left(\frac{20}{\delta_3}\right)^{\delta_3}\left(\frac{10}{\delta_4}\right)^{\delta_4}$$

на множестве неотрицательных решений системы уравнений (3 уравнения с 4 неизвестными)

$$\begin{array}{rrrrl} -\delta_1 & & +\ \delta_3 & +\ \delta_4 & = 0 \\ -\delta_1 & +\ \delta_2 & & +\ \delta_4 & = 0 \\ \delta_1 & +\ \delta_2 & +\ \delta_3 & +\ \delta_4 & = 1 \end{array}$$

и в итоге приходим к нетривиальной задаче максимизации нелинейной функции (гарантий ее вогнутости, т. е. единственности искомого максимума, нет) при линейных ограничениях .

Когда коэффициент трудности невелик (в нашем примере он равен 1), предлагается свести задачу к задаче максимизации меньшей размерности.

Здесь отыскивается некоторое базисное решение системы (например, с использованием симплексной процедуры с полным искусственным базисом до полного исключения искусственных переменных), т.е. система разрешается относительно $n+1$ переменных с сохранением неотрицательности правой части системы. Так вышеприведенная система сводится к виду:

$$\begin{array}{llll} & & \delta_3 & + 0.667\,\delta_4 = 0.333 \\ & \delta_2 & & + 0.667\,\delta_4 = 0.333 \\ \delta_1 & & & -0.333\,\delta_4 = 0.333 \end{array}$$

Полагая внебазисные $t=m-n-1$ переменных некоторым неотрицательным величинам r_i ($i = 1, .., t$), используя представления базисных через небазисные в $V(\delta)$ и приходим к максимизации функции t переменных в области, определяемой системой $n+1$ линейных неравенств.

В нашем примере

$$\begin{array}{lll} \delta_1 = 0.333 & +0.333\,r & >0 \\ \delta_2 = 0.333 & -0.667\,r & >0 \\ \delta_3 = 0.333 & -0.667\,r & >0 \end{array}$$

$$V(r) = \left(\frac{120}{1+r}\right)^{\frac{1}{3}(1+r)} \cdot \left(\frac{120}{1-2r}\right)^{\frac{1}{3}(1-2r)} \cdot \left(\frac{60}{1-2r}\right)^{\frac{1}{3}(1-2r)} \cdot \left(\frac{10}{r}\right)^{r}$$

и задача свелась к максимизации функции одной переменной в области $0 < r < 0.5$.

Без особых затруднений получив $r \sim 0.0953$, находим оптимальный вариант значений $\delta = (0.3651, 0.2698, 0.2698, 0.09536)$ и $V(\delta) = 106.5455$.

Дальнейшие действия идентичны ранее рассмотренному примеру, где решаем переопределенную, линейно зависимую систему (19):

$$\frac{40}{x_1 x_2} = \delta_1 V(\delta)\ , \qquad 40x_2 = \delta_2 V(\delta)\ ,$$

$$20x_1 = \delta_3 V(\delta), \quad 10x_1 x_2 = \delta_1 V(\delta),$$

из которой $x_1 = 1.4373$, $x_2 = 0.7186$.

Геометрическое программирование не нашло столь массовой известности как линейное или квадратическое, поскольку еще лет 10 назад вычислительные возможности компьютеров при значительных размерностях решаемых задач не соответствовали их трудоемкости.

6. ВВЕДЕНИЕ В ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

6.1. Многошаговые процессы принятия решений

В начале шахматной партии игрок делает ход, определяемый некоторым выбором из 20 возможных, например любимый ход гроссмейстера Бендера e2-e4. Следующий его ход определяется ответным ходом партнера и возникшей совокупностью возможных выборов и т. д. Как найти последовательность выборов, приводящую к достижению успеха? Перебор всех вариантов, которые могут возникнуть в процессе игры, невозможен даже для современного суперкомпьютера и прогнозы гибели шахматной игры как таковой преждевременны.

Во многих сферах человеческой деятельности принимать решения без учета последствий чревато тяжкими последствиями. Получение максимума удовольствия на каждом очередном этапе влечет гибель от жажды в пустыне, экологические катастрофы или уголовную ответственность за невыполнение договорных обязательств. Вся история человечества полна примерами безрассудных, сиюминутных решений с гибельными последствиями.

Настоящий «хозяин» лесохозяйственного предприятия, ставящий целью получение максимума прибыли в течение ближайшего столетия, едва ли откажется от затрат на восстановление лесов и будет руководствоваться на каждом этапе деятельности примитивным принципом «руби побольше».

Автоматическое управление ракетой-перехватчиком на очередном участке ее траектории не повторяет управляющих характеристик предыдущего этапа, хотя конечная цель остается неизменной.

Единственно разумно требовать от человека поэтапного (пошагового) принятия решений, не забывая о необходимости достижения некоторой привлекательной конечной цели.

Для одного человека такая цель может свестись к желанию «побыстрее урвать и смыться на Ривьеру», другой хочет, чтобы его потомки смогли нормально жить на родной земле.

Для одного человека цель может исчисляться в денежном эквиваленте, для другого – в эстетическом удовольствии или моральном удовлетворении.

Не возникает сомнений в разумности использования для управления сложными системами методов математического моделирования, которое подчас дает решения, неожиданные не только для начинающего, но и для опытного управляющего.

Уже ранее мы цитировали Козьму Пруткова о невозможности «объять необъятное», утверждая невозможность исчерпывающего математического моделирования действительности. Как правило, отсутствует глубокое знание законов развития природы и общества, отсутствует достоверная информация об изучаемой системе, вмешивается субъективная оценка последствий решения и др.

Естественно, что, создавая математическую модель управления сложной системой, разработчик не может ограничиться рассмотрением только «верхушки айсберга», но и не должен погрязнуть в деталях непринципиального характера. Последствия переупрощения очевидны , и математическая символика используется для бесполезного представления азбучных понятий. Учет же большого количества факторов приводит к известному «проклятию размерности», делающему задачу выбора оптимальной политики практически неразрешимой.

Для иллюстрации математических проблем, возникающих при исследовании многошаговых процессов принятия решений, рассмотрим следующую идеализированную задачу [19].

6.2. Многошаговый процесс распределения однородного ресурса

Пусть имеется X денежных единиц, часть которых Y используется для вложений в сферу A, а оставшаяся часть $X-Y$ в сферу B. В течение некоторого периода эти вложения дают доход, определяемый значениями некоторых функций $g(Y)$ и $h(X-Y)$ соответственно, который используется для удовлетворения каких-либо личных потребностей. По истечении периода вложения изымаются за $\alpha\cdot Y+\beta\cdot(X-Y)$ денежных единиц (как правило, $\alpha, \beta < 1$), которые в очередном периоде используются для вложений в те же сферы.

Требуется найти политику разделения денежного ресурса, которая обеспечивала бы получение максимального дохода за N периодов.

В случае одношагового процесса (при N=1) доход, зависящий от начального ресурса X и выбора доли $Y \in [0, X]$, равен

$$R_1(X,Y) = g(Y) + h(X-Y),$$

Столь же очевидно, что максимальный доход в одношаговом процессе при начальном ресурсе X равен

$$\max_{0\le Y\le X} [\, g(Y) + h(X-Y)\,]$$

(с такой ситуацией столкнемся, например, достигнув последнего периода, когда достаточно принимать решение лишь на шаг вперед).

В случае двух периодов ($N = 2$) суммарный доход зависит от начального ресурса X, выборов в первом и втором периодах Y и Y_1:

$$R_2(X, Y, Y_1) = [\, g(Y) + h(X-Y)\,] + [\, g(Y_1) + h(X_1-Y_1)\,],$$

где

$$X_1 = \alpha\cdot Y+\beta\cdot(X-Y) \quad ; \quad 0 \le Y \le X \quad ; \quad 0 \le Y_1 \le X_1 \,.$$

Аналогично при любом $N > 1$ доход за N периодов складывается из дохода в первом периоде и суммарного дохода в <u>последующих</u> $N-1$ периодах

$$R_N(X,Y,Y_1,Y_2,..) = g(Y) + h(X-Y) + \sum_{k=1}^{N-1}[\, g(Y_k) + h(X_k - Y_k)\,],$$

где

$$X_1 = \alpha\cdot Y + \beta\cdot(X-Y)\ ; \qquad X_k = \alpha\cdot Y_{k-1} + \beta\cdot(X_{k-1} - Y_{k-1}),\quad k=2\ldots N-1;$$

$$0 \le Y \le X\ ,\quad 0 \le Y_1 \le X_1\ ,\quad 0 \le Y_2 \le X_2\ ,\quad \ldots\ .$$

Следовательно, задача поиска максимального дохода в N-шаговом процессе при заданном начальном денежном ресурсе X сводится к максимизации функции N неизвестных Y, Y_1 ,Y_2 , . . при указанных ограничениях, т. е. к задаче математического программирования.

Если функции g и h линейны, то имеем дело с задачей линейного программирования и нет принципиальных преград для ее решения при конкретном значении X. Если же они нелинейны, то возникает ряд неприятных осложнений, связанных в первую очередь с проблемой размерности (*проклятие размерности*).

Как уже неоднократно указывалось ранее, решение одномерной задачи не представляет ни малейших затруднений. Можно при конкретном значении X найти производную максимизируемой функции по Y , приравнять ее нулю, решить полученное уравнение с целью найти т. н. «критические» точки и вычислить значения функции в критических точках, лежащих в диапазоне от 0 до X , и в двух граничных точках. Другой путь связан с примитивным табулированием на заданном количестве m точек интервала или выбором какой-нибудь более эффективной по времени численной процедуры.

При $N >> 1$ примитивное табулирование приводит к объему вычислений порядка m^N (m – число выборов по отдельной переменной), подчас нереальному для суперкомпьютеров.

А как быть в случае, когда величина исходного ресурса заранее точно неизвестна и лишь имеется информация о диапазоне возможных его значений? Ведь здесь даже при линейных функциях придется либо решать очень много линейных программ, либо решать достаточно сложную, нестандартную параметрическую линейную программу.

6.3. Принцип оптимальности и рекуррентные соотношения

Один из путей изучения многошаговых процессов связан с использованием интуитивного принципа оптимальности, сформулированного в 1957 году выдающимся американским математиком Р. Беллманом[12] в книге «Динамическое программирование» [19] и определяющего фундаментальное свойство оптимальной стратегии (политики, поведения).

Р. Беллман

«Оптимальное поведение обладает тем свойством, что, каковы бы ни были первоначальное состояние и решение в начальный момент, последующие решения должны составлять оптимальное поведение относительно состояния, полученного в результате начального решения».

Это *рекурсивное* определение можно интерпретировать самым примитивным высказыванием: если вы намерены добиться наилучшего эффекта вашей многолетней деятельности, то на любом ее этапе, независимо от того, в каких состояниях вы оказывались ранее и какие действия предпринимали, действуйте оптимально (с точки зрения конечной цели), насколько вам позволяет то состояние, в которое вы загнали себя предыдущими действиями.

Если на любом этапе своей целенаправленной деятельности руководствоваться стремлением к достижению некоторой высшей цели, то вся последовательность действий будет оптимальна.

Заметьте, что принцип оптимальности определяет особенности политики, направленной не на получение сиюминутной выгоды, а на достижение некоторой удаленной цели.

Обратимся к поставленной выше задаче поэтапного распределения ресурса, где оптимальность политики определяется достижением максимума суммарного дохода.

Обозначим через $\boldsymbol{F_k(Z)}$ ***– суммарный доход в k -шаговом процессе при начальном ресурсе Z и использовании оптимальной политики*** (максимальный суммарный доход за k шагов при начальном ресурсе Z).

Как мы уже отмечали ранее, доход в k –шаговом процессе можно представить суммой дохода на первом шаге и дохода на последующих

---

[12] Ричард Эрнст Беллман (1920–1984) – американский математик, один из ведущих специалистов в области прикладной математики и вычислительной техники, автор 619 статей и 39 книг. Получил многочисленные результаты, связанные с применением динамического программирования в разных областях математики (вариационное исчисление, автоматическое регулирование, теория аппроксимации, исследование операций и др.).

$k\!-\!1$ шагах (из всех возможных представлений нам нравится именно такое!)

$$R_k(X,Y,Y_1,Y_2,..) = g(Y) + h(X-Y) + \sum_{i=1}^{k-1}[\,g(Y_i) + h(X_i - Y_i)\,].$$

Если после первого шага вспомнить о принципе оптимальности и необходимости дальнейшей оптимальной политики, то доход в k-шаговом процессе сложится из <u>дохода на первом шаге</u> $g(Y) + h(X-Y)$ и <u>максимального дохода на последующих</u> $k\!-\!1$ <u>шагах</u>, которые начнутся уже при денежном ресурсе $\alpha\cdot Y+\beta\cdot(X-Y)$, т. е.

$$F_k(X) = max\left[h(X) + F_{k-1}(\beta X),\ g(X) + F_{k-1}(\alpha X)\right]$$

(здесь мы уже предполагаем оптимальность последующих выборов Y_1, Y_2, …).

Если подобрать величину Y (выбор на первом шаге при ресурсе X) так, чтобы эта сумма была максимальна, то мы получим оценку максимального дохода в $k-$шаговом процессе с начальным ресурсом X (!):

$$\boxed{F_k(X) = \max_{0 \le Y \le X}\left[g(Y) + h(X-Y) + F_{k-1}(\alpha Y + \beta(X-Y))\right]}, \qquad (1)$$

(такое соотношение справедливо при любом $k \geq 2$).

Если учесть, что максимальный доход в одношаговом процессе при начальном денежном ресурсе X равен

$$F_1(X) = \max_{0 \le Y \le X}\left[g(Y) + h(X-Y)\right], \qquad (2)$$

напрашивается последовательность решения поставленной задачи: найти функцию $F_1(X)$ и затем на основе рекуррентных соотношений (1) – функций $F_2(X), \ldots, F_N(X)$.

В дополнение к введенным ранее обозначениям $F_k(Z)$ будем обозначать через $Y_k(Z)$ *– оптимальный выбор на первом шаге k - шагового процесса с начальным ресурсом Z*.

6.4. Структура решения

Как уже сказано, решение задачи для N -шагового процесса разделения ресурса начинаем поиском функции $F_1(X)$ – максимального дохода в одношаговом процессе при начальном денежном ресурсе X и соответствующих значений Y, обеспечивающих максимум, т. е. функции $Y_1(X)$ – оптимального выбора для одношагового процесса с начальным ресурсом X. Тем самым мы узнаем, как следует действовать, если длительность предстоящего процесса составит только один шаг и начальный ресурс для него окажется равным X.

На следующем этапе решения, обладая информацией о функции F_1 при произвольном аргументе, выясняем, каков же окажется максимальный доход в случае двухшагового процесса при произвольном начальном ресурсе X

$$F_2(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X-Y) + F_1(\alpha Y + \beta(X-Y)) \right]$$

и $Y_2(X)$ – оптимальный выбор на первом шаге двухшагового процесса при начальном ресурсе X.

Затем аналогично можно получить оценки функций максимального дохода $[F_3(X), \ldots, F_N(X)]$ и поведений на первом шаге процесса соответствующей длительности $[Y_3(X), \ldots, Y_N(X)]$.

Появляется возможность отыскания оптимальной политики при конкретном начальном ресурсе $X = Z$.

Оптимальный выбор на первом шаге N-шагового процесса при этом ресурсе равен $\hat{y}_1 = Y_N(Z)$, в результате чего к началу второго шага начальный ресурс будет равен $Z_1 = \alpha\hat{y}_1 + \beta(Z - \hat{y}_1)$. Следовательно, оптимальный выбор на втором шаге совпадет с оптимальным выбором на первом шаге оставшегося $N–1$ -шагового процесса, т. е. $\hat{y}_2 = Y_{N-1}(Z_1)$, и к началу третьего шага ресурс станет равным $Z_2 = \alpha\hat{y}_2 + \beta(Z_1 - \hat{y}_2)$. Оптимальный выбор на третьем шаге совпадет с оптимальным выбором на первом шаге оставшегося $N–2$ –шагового процесса, т. е. $Y_{N-2}(Z_2)$ и т. д. Оптимальный выбор на последнем шаге будет определяться значением функции $Y_1(Z_{N-1})$ оптимального выбора для одношагового процесса.

6.5. Простейший случай: выпуклые и линейные функции

Пусть функции $g(X)$ и $h(X)$ являются выпуклыми (выпуклыми вниз, $g''(X) \ge 0$ и $h''(X) \ge 0$). При поиске

$$F_1(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X-Y) \right]$$

имеем дело с максимумом суммы выпуклых функций, которая является выпуклой функцией. Очевидно, что максимум выпуклой функции в интервале достигается на одном из концов интервала и, если допустить, что $g(0) = h(0) = 0$, то $F_1(X) = max\{h(X), g(X)\}$ и $Y_1(X)$ равно 0 или X соответственно.

Так как $F_1(X)$ выпукла, то при поиске

$$F_2(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X-Y) + F_1(\alpha Y + \beta(X-Y)) \right]$$

мы вновь сталкиваемся с поиском максимума суммы выпуклых функций

$$F_2(X)=max\left[h(X)+F_1(\beta X), g(X)+F_1(\alpha X)\right]$$

и оптимальный выбор на первом шаге двухшагового процесса с начальным ресурсом X, т. е. $Y_2(X)$, равен 0 или X.

По индукции можно показать, что при любом $k > 1$

$$F_k(X)= \max_{0 \le Y \le X} \left[h(X)+F_{k-1}(\beta X),\ g(X)+F_{k-1}(\alpha X)\right]$$

и $Y_k(X)$ равно 0 или X соответственно.

Пример. Пусть $g(X) = 9{\cdot}X$, $h(X) = 5{\cdot}X$, $\alpha = 0.4$, $\beta = 0.75$, $N = 4$. Так как функции линейны (частный случай выпуклости), то при $X > 0$

$$F_1(X) = max\ \{ 5\cdot X,\ 9{\cdot}X \} = 9{\cdot}X\ ; \qquad Y_1(X) = X\ .$$

Далее

$$F_2(X)=max\{ 5X+F_1(0.75X), 9X+F_1(0.4X)\}=$$
$$= max\ (5X + 9\cdot 0.75{\cdot}X,\ 9{\cdot}X + 9\cdot 0.4{\cdot}X) =$$
$$= max\ (11.75\cdot X, 12.6\cdot X) = 12.6{\cdot}X; \qquad Y_2(X) = X;$$

$$F_3(X)=max\{ 5X+F_2(0.75X), 9X+F_2(0.4X)\}=$$
$$= max\ (5X + 12.6\cdot 0.75{\cdot}X,\ 9X + 12.6\cdot 0.4{\cdot}X) =$$
$$= max\ (14.45\cdot X, 14.04\cdot X) = 14.45{\cdot}X; \qquad Y_3(X) = 0;$$

$$F_4(X)=max\{ 5X+F_3(0.75X), 9X+F_3(0.4X)\}=$$
$$= max\ (5X + 14.45\cdot 0.75{\cdot}X,\ 9X + 14.45\cdot 0.4\cdot X) =$$
$$= max\ (15.8375\cdot X, 14.78{\cdot}X) = 15.8375{\cdot}X; \qquad Y_4(X) = 0\ .$$

Обратите внимание на то, что мы нашли не значения функций $F_k(X)$, $Y_k(X)$, а сами функции, что позволит нам выяснить оптимальную политику при любом начальном ресурсе Z.

Величина дохода в заданном 4-шаговом процессе равна $F_4(Z)$=15.8375$\cdot Z$. Оптимальный выбор на первом шаге $\{\hat{y}_1\}$ совпадает с $Y_4(Z)$=0, т. е. весь денежный ресурс вкладывается в оборудование типа В. Оптимальный выбор на втором шаге $\{\hat{y}_2\}$ совпадает с оптимальным выбором на первом шаге оставшегося трехшагового процесса с изменившимся начальным ресурсом

$$Z_1 = 0.4\ \hat{y}_1 + 0.75\ (Z - \hat{y}_1) = 0.75\ Z\ ,$$

т.е. с $Y_3(Z_1) = 0$ (на втором шаге политика та же, что и на первом). Оптимальный выбор на третьем шаге $\{\hat{y}_3\}$ совпадает с оптимальным выбором на первом шаге оставшегося двухшагового процесса с начальным ресурсом

$$Z_2= 0.4\ \hat{y}_2 + 0.75\ (Z_1 - \hat{y}_2) = 0.5625\ Z\ ,$$

т.е. с $Y_2(Z_2)$=Z_2 (весь ресурс в сферу А). Наконец, на последнем (четвертом) шаге оптимальный выбор $\{\hat{y}_4\}$ совпадает с оптимальным выбором для оставшегося одношагового процесса с начальным ресурсом

$$Z_3 = 0.4\,\hat{y}_3 + 0.75\,(Z_2 - \hat{y}_3) = 0.225\,Z,$$

т.е. с $Y_1(Z_3) = Z_3$ (весь ресурс в сферу А).

Скептически настроенный читатель заметит, что рассмотренный пример можно было свести к элементарной задаче линейного программирования с 4 неизвестными:

максимизировать

$$L(Z) = 9\,(Y_1+Y_2+Y_3+Y_4) + 5\,(Z - Y_1 - Z_1 - Y_2 + Z_2 - Y_3 + Z_3 - Y_4)$$

при условиях

$$0 \le Y_1 \le Z\,; \quad 0 \le Y_2 \le Z_1\,; \quad 0 \le Y_3 \le Z_2\,; \quad 0 \le Y_4 \le Z_3\,;$$

$$Z_1 = 0.4Y_1 + 0.75(Z - Y_1); \quad Z_2 = 0.4Y_2 + 0.75(Z_1 - Y_2); \quad Z_3 = 0.4Y_3 + 0.75(Z_2 - Y_3).$$

Но решение этой задачи, например, симплексным методом придется выполнять для конкретного значения Z, тогда как решение с помощью рекуррентных соотношений дает готовую политику для любого Z.

6.6. Эффективность метода динамического программирования

Из приведенных выше замечаний и решений можно сделать некоторые выводы относительно метода динамического программирования, базирующегося на принципе оптимальности и реализуемого, в частности, в форме метода рекуррентных соотношений.

1. **Метод динамического программирования позволяет свести *N*-мерную задачу оптимизации к совокупности задач меньшей размерности** (очевидно, что легче решить 20 одномерных задач оптимизации, чем одну 20-мерную).

2. **Метод ориентирован на решение не конкретной задачи, а целого класса подобных задач**.

3. Как будет видно из дальнейшего рассмотрения численных реализаций, **появление дополнительных ограничений подчас облегчает решение задачи за счет уменьшения объема перебора вариантов**.

Идеология, использованная при рассмотрении задачи разделения ресурса, может быть обобщена на случай любого многошагового процесса принятия решений.

Если обозначить $F_k(x)$ – эффект некоторой деятельности за k периодов при оптимальной политике и начальном состоянии x, а через y – некоторое поведение на начальном этапе, выбираемое из множества допустимых поведений $Y(x)$, то

$$F_k(x) = \underset{y \le Y(x)}{extr}\, \mathrm{T}\left[y, F_{k-1}(x(y))\right]$$

где Т – некоторая функция от принимаемого решения и наилучшего эффекта от деятельности на последующих $k-1$ этапах, которая начнется уже от состояния $x(y)$.

Ниже мы рассмотрим несколько достаточно простых задач [19,20] для иллюстрации метода динамического программирования (построение рекуррентных соотношений и тем самым сведение многомерной задачи оптимизации к последовательности задач меньшей размерности; аналитическое решение рекуррентных соотношений и выявление структуры полученного решения).

6.7. Задача складирования однородного продукта

Пусть имеется склад вместимостью B с начальным запасом V некоторого продукта, цены на который подвержены сезонным изменениям, но мы тем не менее обладаем надежным прогнозом.

В начале i-го сезона часть сохраненного продукта Y_i можно продать по цене P_i и в конце сезона закупить X_i этого продукта по цене C_i. Образовавшийся на складе запас хранится до следующего сезона. Хотелось бы найти политику продажи–покупки, максимизирующую суммарный доход за N сезонов.

Если пренебречь затратами на хранение и сопутствующими факторами, то задачу можно свести к максимизации линейной функции

$$\sum_{i=1}^{N}[\,P_i\,Y_i - C_i\,X_i\,]$$

при условиях

$$0\le Y_1\le V,\qquad X_1\ge 0,\qquad V_1=V-Y_1+X_1\le B;$$
$$0\le Y_2\le V_1,\qquad X_2\ge 0,\qquad V_2=V_1-Y_2+X_2\le B;$$
$$0\le Y_3\le V_2,\qquad X_3\ge 0,\qquad V_3=V_2-Y_3+X_3\le B;\ \ldots.$$

т. е. к задаче линейного программирования с $2N$ неизвестными и $4N$ ограничениями, которую можно решить симплексным методом при фиксированном значении V.

Рассмотрим задачу с других позиций, для чего введем обозначения:

k – число предстоящих сезонов;

$F_k(V)$ – максимальный доход за k сезонов при начальном запасе V;

P_k ,C_k – цены продажи–покупки в первом из предстоящих k сезонов;

Y , X – объемы продажи-покупки в первом из этих сезонов;

$Y_k(V)$, $X_k(V)$ – оптимальные объемы продажи –покупки в этом сезоне (на первом шаге k – шагового процесса с начальным запасом V).

Для процесса длительностью в один этап (с таковым имеем дело, когда до конца процесса останется один шаг), если к его началу запас составит V единиц,

$$F_1(V) = max\ \{ P_1 Y - C_1 X \}\ , \qquad (1)$$

где область максимизации определяется условиями $0 \le Y \le V$; $X \ge 0$.

Очевидно, что максимум здесь достигается при $Y = V$ и $X = 0$ (естественно при завершении деятельности по купле-продаже продать весь запас и ничего не покупать), т. е.

$$F_1(V) = P_1 V\ ;\quad Y_1(V) = V\ ;\ X_1(V) = 0\ . \qquad (1a)$$

Обратимся к случаю k –шагового процесса при $k > 1$, повторив традиционные рассуждения.

Доход в k -шаговом процессе складывается из дохода на первом шаге $\{P_k Y - C_k X\}$ и дохода на оставшихся k–1 шагах.

Если мы желаем действовать оптимально, то обязаны руководствоваться принципом оптимальности: независимо от начального состояния (запаса V) и начального поведения (объемов Y, X продажи-покупки на первом шаге) дальнейшая политика должна быть оптимальной, исходя из возникающего состояния (запаса $V - Y + X$).

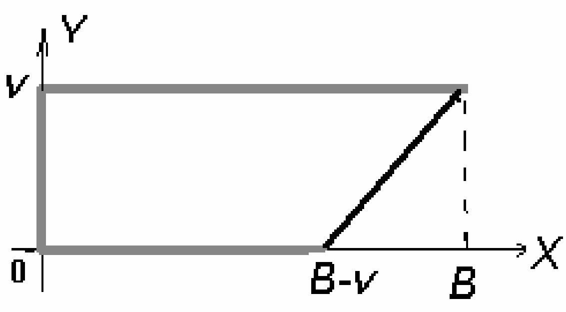
Тогда доход на оставшихся k –1 шагах будет равен $F_{k-1}(V - Y + X)$ и с учетом желания с первого шага действовать оптимально

$$F_k(V) = max\ \{ P_k Y - C_k X + F_{k-1}(V - Y + X) \}\ ,\ k = 2\ ..\ N\ , \qquad (2)$$

где область максимизации определяется условиями:

$$0 \le Y \le V\ ;\quad X \ge 0\ ;\quad V - Y + X \le B\ .$$

Легко видеть, что множество планов (область максимизации) при любом k – многоугольник с координатами вершин, линейно зависящими от V (см. рисунок).



Так как функция $F_1(V) = P_1 V$ *линейна* по V, то при поиске $F_2(V)$ мы сталкиваемся с максимизацией функции, *линейной* по X и Y, над указанным множеством планов, т. е. с задачей линейного программирования с экстремумами в вершинах множества. Можно показать, что аналогичное явление наблюдается при любом $k > 1$.

Соответственно (2) упрощается к виду:

$$F_k(V) = max\begin{cases} F_{k-1}(V) & ;\ Y_k(V)=0\ ;\ X_k(V)=0 \\ P_kV + F_{k-1}(0) & ;\ Y_k(V)=V\ ;\ X_k(V)=0 \\ -C_k(B-V)+F_{k-1}(B) & ;\ Y_k(V)=0;\ X_k(V)=B-V \\ P_kV - C_kB + F_{k-1}(B) & ;\ Y_k(V)=V;\ X_k(V)=B \end{cases}$$

Так исходная задача линейного программирования с $2N$ неизвестными сведена к N элементарным линейным программам с двумя неизвестными. Более того, здесь мы имеем решение при любых V и B.

Пример. Пусть $N = 5$ и

Сезон	1	2	3	4	5
Продажная цена (P)	7	6	4	4	8
Закупочная цена (C)	6	7	5	3	5

Решение задачи дает:

$$F_1(V) = 8\ V\ ; \qquad Y_1(V) = V;\ X_1(V)=0;$$

$$F_2(V) = max\begin{cases} F_1(V) = 8V \\ 4V + F_1(0) = 4V \\ -3(B-V)+F_1(B) = 3V+5B \\ 4V - 3B + F_1(B) = 4V + 5B \end{cases} = 4V+5B;\quad Y_2(V)=V;\ X_2(V)=B;$$

$$F_3(V) = max\begin{cases} F_2(V) = 4V+5B \\ 4V + F_2(0) = 4V+5B \\ -5(B-V)+F_2(B) = 5V+4B \\ 4V - 5B + F_2(B) = 4V + 4B \end{cases} = 4V+5B;\ Y_3(V)=0,V;\ X_3(V)=0;$$

$$F_4(V) = max\begin{cases} F_3(V) = 4V\ +5B \\ 6V + F_3(0) = 6V+5B \\ -7(B-V)+F_3(B) = 7V+2B \\ 6V - 7B + F_3(B) = 6V + 2B \end{cases} = 6V+5B;\ Y_4(V)=V;\ X_4(V)=0;$$

$$F_5(V) = \ max\begin{cases} F_4(V) = 6V\ +5B \\ 7V + F_4(0) = 7V+5B \\ -6(B-V)+F_4(B) = 6V+5B \\ 7V - 6B + F_4(B) = 7V + 5B \end{cases} = 7V+5B;\ Y_5(V)=V;\ X_5(V)=0,B.$$

Отсюда видим, что искомый максимум дохода в 5-шаговом процессе при начальном запасе V равен $F_5(V) = 7V+5B$ и оптимальная политика по шагам определяется последовательностью действий.

1. Объем продажи = $Y_5(V)=V$ – весь запас.
   Объем закупок = $X_5(V) = 0$ или B – ничего или полный склад.
2. Объем продажи = Y_4(0 или B)= 0 или B – весь запас, если он есть.

Объем закупок $= X_4(0$ или $B)=0$ – ничего не покупать.

3. Объем продажи = $Y_3(0)=0$ – можно все продать, но нечего.
   Объем закупок = $X_3(0)=0$ –ничего не покупать.
4. Объем продажи = $Y_2(0)=0$ – продать запас, но его нет.
   Объем закупок = $X_2(0)=B$ – закупить полный склад.
5. Объем продажи = $Y_1(B) = B$ – продать весь запас.
   Объем закупок = $X_1(V)$ – ничего не покупать.

Идеология рассмотренного решения не претерпит изменения, если появятся дополнительные ограничения на объемы продаж-покупок или дополнительные затраты, пропорциональные объемам

Читатель может сопоставить необходимые затраты энергии на решение этой задачи методом линейного программирования и использованным здесь методом.

6.8. Задача надежности многокомпонентных схем

При конструировании сложной аппаратуры одной из основных является задача надежности, которая решается иногда путем дублирования компонент [19, 20]. Например, при построении последовательной схемы из N ступеней надежность (вероятность безотказной работы) равна произведению вероятностей безотказной работы ее ступеней.

Если j -я ступень содержит $1+M_j$ компонент-дублеров и $\Phi_j(M_j)$ – вероятность ее безотказной работы, то надежность схемы $P(N)$ равна произведению значений $\Phi_j(M_j)$ при $j = 1, 2, .. , N$.

Очевидно, что большое количество дублеров ведет к росту стоимости, объема и появлению дополнительных ошибок. Если учесть только фактор стоимости, то возникает ограничение

$$\sum_{j=1}^{N} C_j\ M_j \le S\ ,$$

где S – заданная предельная стоимость дублирования схемы и C_j – стоимость одного дублера в j -й ступени. Не вызывает сомнения, что значения M_j неотрицательны и целочисленны.

Если вообразить N -шаговый процесс, на первом шаге которого выясняется число дублеров для N–ой ступени, на втором для N–1 -ой и т. д. и обозначить через $F_k(S)$ максимум надежности k -компонентной схемы, то из принципа оптимальности получим систему рекуррентных соотношений вида:

$$F_k(S) = \max_X \{\Phi_k(X)\cdot F_{k-1}(S - C_k X)\}\ ,\ k = 2, \ldots ,N\ ;$$

$$F_1(S) = \max\ \{\ \Phi_1\ (X)\ \}\ ,$$

где X – число дублеров k -ой ступени; X=0, 1,2, ...; C_kX не превышает S.

Требование целочисленности, даже при достаточно простых функциях $\Phi(X)$, не дает возможности получения аналитического решения приведенной системы и вынуждает прибегнуть к численному

6.9. Упражнения

Здесь мы предлагаем ряд задач, решение которых поможет читателю убедиться в глубине усвоения метода рекуррентных соотношений и, может быть, откроет некоторые «изюминки» (решения некоторых из этих задач можно найти в [20]).

1. Пусть имеется стадо из X голов скота и возможность отправки его части Y в конце года на рынок с получением дохода $\Phi(Y)$. Оставшаяся часть $X–Y$ за год увеличивается в α раз ($\alpha > 1$) . Требуется найти политику продажи, максимизирующую суммарный доход за N лет.

Покажите, что задача сводится к системе рекуррентных соотношений

$$F_k(X) = \max_{0 \le Y \le X} \{ \Phi(Y) + F_{k-1}(\alpha (X-Y)) \} , \; k = 2 , .. , N \; ;$$

$$F_1(X) = \max_{0 \le Y \le X} \{ \Phi(Y) \} \; .$$

Какой будет структура оптимальной политики в случае, когда $\Phi(X)$ – выпуклая возрастающая функция и $\Phi(0) = 0$?

Что изменится в приведенной системе при учете процентного соотношения между рождающимися особями разного пола?

Изменится ли структура оптимальной политики в случае невыпуклой функции $\Phi(X)$ (например вогнутой, т. е. при $\Phi''(X) <0$) ?

2. Предположим, что из имеющегося капитала X часть денег Y может использоваться на приобретение облигаций, а часть Z – на приобретение акций ($Y + Z \le X$) . Как должен использоваться капитал, чтобы за N лет он достиг максимума, если за год вложения в облигации и акции дают прирост в A и B раз соответственно?

Убедитесь, что решение определяется соотношениями

$$F_{k \ge 1}(X) = \max_{0 \le Y \le X} \{ F_{k-1}(\alpha Y + B(X-Y))) \} \; ; \quad F_0(X) = X \; .$$

найдите решение и согласуйте его со здравым смыслом.

3. Требуется разбить некоторое число C на N слагаемых так, чтобы их произведение было максимально.

Осознайте, что задачу можно свести к рекуррентным соотношениям вида $F_k(C) = \max_{0 \le X \le C} \{ X \cdot F_{k-1}(C-X) \} , \; k=1 .. N \; ; \quad F_0(C) = 1$.

Несмотря на очевидность искомого решения, получите формально оптимальную политику разбиения (при решении этой задачи люди, неискушенные в элементарной математике, приходят к выводам, едва ли согласующимся со здравым смыслом).

4. Медведь-арендатор обладает X кг моркови и Y кг морковных семян. Часть Z моркови он может обменять у Лисы на $R(Z)$ кг меда. В очередном сезоне оставшаяся морковь может быть использована для выращивания семян, а имеющиеся семена – для выращивания моркови.

Допустим, что наш Медведь работает в зоне устойчивого земледелия и урожайность является известной величиной; более того, Лиса отличается честностью в торговых операциях.

Постройте рекуррентные соотношения, позволяющие Медведю найти политику, максимизирующую суммарный обьем полученного меда за N сезонов. Какие изменения возникнут в соотношениях и оптимальной политике при учете дополнительных условий, связанных с ограниченностью посевных площадей, складских помещений и т. д.

5. Потребность в машинах данного типа является некоторой функцией от времени. Требуется определить порядок их приобретения, при котором эта потребность удовлетворяется с минимальными расходами, при следующих условиях:

– покупка одной машины обходится в P денежных единиц;

– эксплуатация машины в интервале времени стоит M денежных единиц;

– стоимость содержания и ремонта в течение интервала времени является функцией от имеющегося количества машин (не обязательно линейной).

Покажите, что задача сводится к системе соотношений

$$F_N(X) = \min_{X+Z\geq R} \left[P\cdot Z + M\cdot (X+Z) + F_{N-1}(X+Z) \right],$$

где X – число машин к моменту начала процесса; значения R представляют собой потребности в машинах на текущий момент.

6. Покажите, что поиск максимума $\sum_{i=1}^{N} \Phi(X_i)$ при условиях $\sum_{i=1}^{N} X_i =$ C, $X_i \geq 0$, $i=1 \dots N$; $\Phi(X)$ выпукла, сводится к соотношениям

$$F_N(C)= \max_{0\leq X\leq C} \left[\Phi (X) + F_{N-1} (C-X) \right], \ N \geq 2,$$

$$F_1(C)= \max_{0\leq X\leq C} \{ \Phi(X) \}.$$

Рассмотрите структуру решения задачи для случая вогнутой функции $\Phi(X) = X - b X^2$, $b > 0$, $0 \leq C \leq \frac{1}{2b}$.

7. Рассмотрим ситуацию последовательной обработки на двух машинах N различных деталей, если известно время A_i и B_i обработки i -й детали на соответствующих машинах [20]. Очевидно, что первая машина будет загружена полностью, но вторая может периодически оказываться в состоянии простоя. Как найти порядок обработки, минимизирующий время ее простоя и тем самым общее время обработки. Эта задача называется задачей Джонсона или задачей планирования производственной линии.

Обозначив через X_i – простой в ожидании i -й детали, имеем

$$X_1 = A_1;$$

$$X_1 + X_2 = max\,(A_1 + A_2 - B_1, A_1);$$

$$X_1 + X_2 + X_3 = max\,(A_1 + A_2 + A_3 - B_1 - B_2, A_1 + A_2 - B_1, A_1); \ldots$$

$$\sum_{i=1}^{N} X_i = \max_{1 \le k \le N} \{ \sum_{i=1}^{k} A_i - \sum_{i=1}^{k-1} B_i \}.$$

Читатель может обратиться к [20], где с помощью рекуррентных соотношений показано условие необходимости перестановки в паре деталей (i, j)

$$min\,(A_j, B_i) < min\,(A_i, B_j).$$

Соответственно среди всех значений A_i и B_i ищем наименьшее. Если найденное значение совпадает с некоторым A_i, то i -ю деталь ставим на обработку первой; если оно совпадает с некоторым B_i, то последней. Эту процедуру повторяем для всех остальных деталей.

Пусть, например, время обработки задано таблицей

i	1	2	3	4	5	6	7	8
A_i	4	4	30	6	2	9	13	9
B_i	5	1	4	30	3	13	9	9

Минимальное из значений равно 1 и соответствует B_2 – вторая деталь обрабатывается последней. Минимальное из значений (кроме второго столбца) соответствует A_5 – пятая деталь обрабатывается первой. Минимальное из значений в столбцах, кроме 2 и 5, равно A_1 и соответственно среди рассматриваемых сейчас деталей эта деталь обрабатывается первой и т. д. В итоге такого упорядочения получаем:

i	5	1	4	8	6	7	3	2
A_i	2	4	6	9	9	13	30	4
B_i	3	5	30	9	13	9	4	1

Время простоя второй машины при первичном порядке равно

$$max\,(4,\ 4+4-5,\ 4+4+30-5-1,\ 4+4+30+6-5-1-4,\ 4+4+30+6+2-5-1-4-30,\ 4+4+30+6+2+9-5-1-4-30-3,\ 4+4+30+6+2+9+13-5-1-4-30-3-13,\ 4+4+30+6+2+9+13-5-1-4-30-3-13) = max\,(4, 3, 32, 34, 6, 12, 12, 12) = 34.$$

Простой при оптимальной перестановке составит

max (2, 2+4–3, 2+4+6–3–5, 2+4+6+9–3–5–30, 2+4+6+9+9–3–5–30–9, 2+4+6+9+9+13–3–5–30–9–13, 2+4+6+9+9+13+30–3–5–30–9–13–9, 2+ 4 +6+9+9+13+30+4–3–5–30–9–13–9–4)=*max* (2,3,4, –17, –17, –17, 4, 4) = 4.

В процессе решения можно было заметить, что существует и другой оптимальный порядок обработки, связанный с неоднозначностью установки детали 8.

К сожалению, столь простого решения задачи Джонсона для случая последовательной обработки на $L > 2$ машинах нет.

7. ВЫЧИСЛИТЕЛЬНЫЙ АЛГОРИТМ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

7.1. Численное решение рекуррентных соотношений

Рассматривавшиеся ранее примеры допускали аналитическое решение за счет выпуклости или линейности исследуемых функций.

В случае функций более сложной природы решение рекуррентных соотношений приходится производить численно.

Обратимся к по-ставленной ранее задаче разделения ресурса:

$$F_k(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X-Y) + F_{k-1}(\alpha Y + \beta(X-Y)) \right], \; k = 2,.., N;$$

$$F_1(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X-Y) \right]$$

Очевидно, что знание значения $F_{k-1}(X)$ при одном конкретном значении X не дает возможности поиска $F_k(X)$ при том же аргументе. Поэтому приходится выбирать сетку M значений X в интервале от нуля до некоторого предельного значения $X=Z$, соответствующего максимально возможному начальному ресурсу, с каким-то шагом и отыскиваем $F_1(X)$ и $Y_1(X)$ в узлах этой сетки, решая с этой целью M одномерных задач оптимизации.

Приступив к поиску значений $F_2(X)$ и $Y_2(X)$ в тех же узлах, мы сталкиваемся с тем, что при различных Y значения $\alpha Y+\beta(X-Y)$ не совпадают с узлами сетки и приходится прибегнуть к приближенной оценке значений $F_1(\alpha Y+\beta(X-Y))$ путем интерполяции.

Напомним, что если задана таблица значений $R(T)$ на равномерной сетке с шагом h, то для нахождения $R(t)$ при $T_i \le t \le T_{i+1}$ можно прибегнуть к линейной интерполяции

$$R(t) = R(T_i) + (t - T_i)\,\frac{R(T_{i+1}) - R(T_i)}{h}$$

(можно прибегнуть и к табличной интерполяции более высоких порядков по формулам Ньютона, Стирлинга или др.). Аналогично отыскиваются $F_3(X)$, $F_4(X)$ и т. д.

В результате получается таблица:

X	$F_1(X)$	$Y_1(X)$	$F_2(X)$	$Y_2(X)$	$F_3(X)$	$Y_3(X)$	...	$F_N(X)$	$Y_N(X)$
0	$F_1(0)$	$Y_1(0)$	$F_2(0)$	$Y_2(0)$	$F_3 0)$	$Y_3(0)$	...	$F_N(0)$	$Y_N(0)$
h	$F_1(h)$	$Y_1(h)$	$F_2(h)$	$Y_2(h)$	$F_3(h)$	$Y_3(h)$	...	$F_N(h)$	$Y_N(h)$
$2h$	$F_1(2h)$	$Y_1(2h)$	$F_2(2h)$	$Y_2(2h)$	$F_3(2h)$	$Y_3(2h)$	...	$F_N(2h)$	$Y_N(2h)$
...	...	...	...	...	...	...	...	...	...
Z	$F_1(Z)$	$Y_{11}(Z)$	$F_2(Z)$	$Y_2(Z)$	$F_3(Z)$	$Y_3(Z)$	...	$F_N(Z)$	$Y_N(Z)$

Для поиска решения при $X = C$ в пределах сетки достаточно провести "обратный ход":

– оптимальный выбор на первом шаге N-шагового процесса $\hat{y}_1 = Y_N(C)$,

– оптимальный выбор на втором шаге $\hat{y}_2 = Y_{N-1}(C_1)$, $C_1 = \alpha\,\hat{y}_1 + \beta\,(C - \hat{y}_1)$,

– оптимальный выбор на третьем шаге $\hat{y}_3 = Y_{N-2}(C_2)$, $C_2 = \alpha\hat{y}_2 + \beta(C_1 - \hat{y}_2)$ ...

Пример. Пусть при $X = 10$ требуется найти решение системы

$$F_k(X) = \min_{0 \le Y \le X} \left\{ \frac{4}{Y} + \frac{9}{X-Y} + F_{k-1}(X - Y) \right\}, \; k = 2, 3, \ldots$$

$$F_1(X) = \min_{0 \le Y \le X} \left\{ \frac{4}{Y} + \frac{9}{X-Y} \right\}$$

Выберем сетку значений X от 1 до 10 с шагом 1 (можно взять и меньший шаг). Тогда с учетом того, что минимум функции для $F_1(X)$ достигается при Y, удовлетворяющих условию $9Y^2 - 4(X-Y)^2 = 0$, т.е. при $Y = 0.4\,X$, отыскиваем $F_1(X) = 25/X$ (то же самое можно получить перебором значений Y в пределах от 0 до X).

Затем отыскиваем $F_2(X)$ и $F_3(X)$ перебором допустимых значений Y или использованием аппарата производных (здесь функции выпуклые, что дает единственность искомого минимума, и возникающие уравнения для производных легко решаются аналитически) и получаем таблицу:

X	$F_1(X)$	$Y_1(X)$	$F_2(X)$	$Y_2(X)$	$F_3(X)$	$Y_3(X)$
1	25.0	0.4	61.3	0.26	107.8	0.19
2	12.5	0.8	30.6	0.51	53.9	0.38
3	8.3	1.2	20.4	0.77	35.9	0.58
4	6.3	1.6	15.3	1.02	26.9	0.77
5	5.0	2.0	12.3	1.28	21.6	0.96
6	4.2	2.4	10.2	1.54	17.9	1.15
7	3.6	2.8	8.8	1.79	15.4	1.34
8	3.1	3.2	7.7	2.06	13.5	1.54
9	2.8	3.6	6.8	2.30	12.0	1.73
10	2.5	4.0	6.1	2.56	10.8	1.92

Отсюда $F_3(10) = 10.8$; выбор на первом шаге $\bar{y}_1 = Y_3(10) = 1.92$; выбор на втором шаге $\bar{y}_2 = Y_2(10–1.92) = Y_2(8.08) = 2.06+(2.30–2.06)\cdot 0.08 = 2.08$; выбор на третьем шаге $\bar{y}_3 = Y_1(8.08–2.08) = Y_1(6) = 2.4$.

Единственная опасность, которая возникает на пути численного решения, связана с возникающей иногда «проблемой расширения сетки».

Примером могла бы служить задача разделения ресурса при $\alpha, \beta > 1$. Пусть при решении соотношения типа

$$F_k(X) = \max_{0 \le Y \le X} \{\ldots + F_{k-1}(3 \cdot (X - Y))\}$$

выбрана сетка значений X от 0 до 10 . При поиске максимума очередной функции потребуется знание значений предыдущей функции на сетке от 0 до 30 и для того, чтобы найти значение $F_5(10)$, потребуется задать $F_1(X)$ в диапазоне от 0 до 810.

7.2. Классические примеры постановки и численного решения

7.2.1. Задача о загрузке корабля

Рассмотрим известную «задачу о загрузке корабля», которую в других эквивалентных постановках называют «задачей о рюкзаке» или «задачей линейного раскроя».

Пусть имеется корабль грузоподъемности W, который может быть загружен предметами N типов. Если обозначить через W_k и C_k вес и ценность предмета k -го типа и через X_k количество таких предметов, то можно поставить задачу загрузки корабля грузом максимальной ценности в виде:

максимизировать $\sum_{k=1}^{N} C_k X_k$ при условиях

$$\sum_{k=1}^{N} W_k X_k \le W, \quad X_k \ge 0 \text{ при всех } k.$$

Если потребовать неделимости предметов, т. е. целочисленности X_k, то возникшую задачу целочисленного линейного программирования можно решать методом ветвей и границ или методом Гомори, получая решение для конкретного значения W . Появление дополнительных условий приведет к усложнению процесса решения.

Поскольку нам безразлично, в каком порядке грузить наши предметы, то вообразим искусственный n -шаговый процесс, на первом шаге которого загружается X предметов n -го типа (на втором $n–1$ -го, затем $n-2$ -го и на последнем – первого типа). Тогда, обозначив через $F_n(W)$ суммарную ценность загрузки в n –шаговом процессе при заданной грузоподъемности W и при использовании оптимальной политики, на основе принципа оптимальности сводим задачу к системе рекуррентных соотношений

$$F_n(W) = max\ [\ C_n X + F_{n-1}(W - W_n X)\ ]\ ,\quad n = 2, 3, .., N\ ;$$
$$F_1(W) = max\ [\ C_1 X\ ]\ ,$$

где область максимизации определяется целыми значениями X в диапазоне от нуля до целой части отношения W / W_n .

Пусть $X_n(W)$ – оптимальное количество предметов, загружаемое на первом шаге n -шагового процесса при начальной грузоподъемности W .

Пример. Решим поставленную задачу при N=3, C_1=8, C_2=7, C_3 =4, W=10, W_1=4, W_2=3, W_3=2.

Возьмем сетку значений W от 0 до 10. Очевидно, что $F_1(W)$ равно C_1, умноженному на максимально возможное X, равное целой части отношения W к W_1.

При поиске $F_2(W)$ приходится перебирать значения X в диапазоне от 0 до целой части отношения W к W_2 и выбирать среди соответствующих значений максимизируемой функции наибольшее. Например, при $W = 10$ перебираем X, равные 0, 1, 2, 3 и

$$F_2\,(10) = max\ [\ 0+F_1(10),\ 7+F_1(7),\ 14+F_1(4),\ 21+F_1(1)\ ] =$$
$$= max\ [\ 16\ ,\ 15\ ,\ 22\ ,\ 21\ ] = 22\ ;\quad X_2\,(10) = 2.$$

Аналогично отыскиваем значения $F_3\,(W)$ и $X_3\,(W)$.

X	$F_1(W)$	$X_1(W)$	$F_2(W)$	$X_2(W)$	$F_3(W)$	$X_3(W)$
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	4	1
3	0	0	7	1	7	0
4	8	1	8	0	8	0,2
5	8	1	8	0	11	1
6	8	1	14	2	14	0
7	8	1	15	1	15	0,2
8	16	2	16	0	18	1
9	16	2	21	3	21	0
10	16	2	22	2	22	0,2

Полученная таблица позволяет найти оптимальную политику загрузки при любом W, не превышающем 10.

Так при W =8 оптимальное количество предметов третьего типа (загружаемых на первом шаге 3–шагового процесса) равно $X_3(8)=1$, оптимальное количество предметов второго типа (загружаемых на первом шаге оставшегося двухшагового процесса) равно $X_2(8–2·1) = X_2$ $X_2(6)= 1$, оптимальное количество предметов первого типа (загружаемых на последнем шаге трехшагового процесса) $X_1(6–3·2) = X_1(0) = 0$.

При W=10 возникают два варианта оптимальной загрузки;

1) $X_3 (10) = 0$; $X_2(10–2·0) = X_2(10) = 2$; $X_1(10–3·2) = X_1(4)=1$;
2) $X_3 (10) = 2$; $X_2 (10–2·2) = X_2(6) = 2$; $X_1(6–3·2) = X_1(0) = 0$.

Приведенный пример еще раз показывает, что решение методом динамического программирования обеспечивает получение политики при любом значении W, тогда как другие подходы к решению дают лишь частное решение. Более того, появление дополнительных ограничений (число предметов такого-то типа не менее или не более указанного значения) упрощает решение за счет уменьшения объема перебора (решение методами целочисленного программирования стало бы более трудоемким).

7.2.2. Задача планирования развития отрасли

Пусть в N регионах действуют или могут быть созданы предприятия по выпуску некоторой продукции и требуется довести выпуск до B единиц.

Введем обозначения:

D_j – существующая мощность j –го предприятия,

E_j – его максимальная возможная мощность,

$M_j = E_j – D_j$ – максимально возможный прирост мощности,

$Ф_j(X)$ –годовые затраты на приращение X мощности,

$Б = B – \sum D_j$ – недобор мощности.

Соответственно можно поставить задачу минимизации

$$F(X) = \sum_{j=1}^{N} Ф_j (X_j)$$

при условиях $\sum_{j=1}^{N} X_j = Б$, $0 \le X_j \le M_j$ $(j = 1 \,..\, N)$.

Заметим, что задача неразрешима, если $\sum M_j < Б$.

Сформулируем задачу в виде задачи N-шагового процесса принятия решений, на первом шаге которого принимается суждение о N-м предприятии, на втором о N–1 -ом и т.д.

Обозначим через $F_k(t)$ минимальные затраты на создание дополнительной мощности t на k предприятиях. Очевидно, что здесь достаточно ограничиться диапазоном значений t от 0 до $min(Б, \Sigma M_j)$.

При $k = 1$ имеем $F_1(t) = \Phi_1(t)$ и при $k > 1$

$$F_k(t) = min [\Phi_k(X) + F_{k-1}(t - X)],$$

где область минимизации определяется условиями $0 \le X \le min (M_k, t)$.

Пусть планируемый выпуск B=100, число предприятий N=4, существующие мощности D=(0, 20, 0, 0), предельные значения мощности E=(20, 60, 50, 30).

Тогда требуемый прирост мощности предприятий $Б$=80 и возможности прироста по предприятиям M=(20, 40, 50, 30).

Пусть затраты на приращение мощности X определяются таблицей:

X	0	10	20	30	40	50
$\Phi_1(X)$	0	10	13			
$\Phi_2(X)$	0	6	11	16	20	
$\Phi_3(X)$	0	9	18	26	32	38
$\Phi_4(X)$	0	8	16	19		

При k=1 достаточно взять сетку значений t от 0 до 20 и $F_1(t)$= $\Phi_1(t)$. При k=2 берем сетку значений t от 0 до 20 + 40 и

$$F_2(t) = min [\Phi_2(X) + F_1(t - X)] \quad \text{при } 0 \le X \le min(t ,40) :$$

t	X=0	X=10	X=20	X=30	X=40	$F_2 (t)$	$X_2(t)$
0	0+ 0					0	0
10	0+10	6+ 0				6	10
20	0+13	6+10	11+ 0			11	20
30		6+13	11+10	16+ 0		16	30
40			11+13	16+10	20 + 0	20	40
50				16+13	20+10	29	30
60					20+13	33	40

При $k = 3$ берем сетку значений t от 0 до $80 < 20 + 40 + 50$ и

$$F_3 (t) = min [\Phi_3 (X) + F_2(t–X)] \quad \text{при } 0 \le X \le min(t ,50) :$$

t	X=0	X=10	X=20	X=30	X=40	X=50	$F_3 (t)$	$X_3(t)$
0	0+ 0						0	0
10	0+ 6	9+ 0					6	0
20	0+11	9+6	18+0				11	0
30	0+16	9+11	18+6	26+0			16	0
40	0+20	9+16	18+11	26+6	32+0		20	0
50	0+29	9+20	18+16	26+11	32+6	38+0	29	0,10
60	0+33	9+29	18+20	26+16	32+11	38+ 6	33	0
70		9+30	18+29	26+20	32+16	38+11	42	10
80			18+30	26+29	32+20	38+16	51	20

При $k = 4$ берем сетку значений t от 0 до $80 < 20 + 40 + 50 +30$ и $F_4(t) = min\ [\ \Phi_4(X) + F_3(t–X)\ ]$ при $0 \le X \le min(t, 30)$:

t	X=0	X=10	X=20	X=30	$F_4(t)$	$X_4(t)$
0	0+ 0				0	0
10	0+ 6	8+ 0			6	0
20	0+11	8+ 6	16+ 0		11	0
30	0+16	8+11	16+ 6	19+ 0	16	0
40	0+20	8+16	16+11	19+ 6	20	0
50	0+29	8+20	16+16	19+11	28	10
60	0+33	8+29	16+20	19+16	33	0
70	0+42	8+33	16+29	19+20	39	10
80	0+51	8+42	16+33	19+29	48	30

Отсюда мы получаем, что для прироста мощности на 80 единиц требуются затраты $F_4(80)=48$ путем прироста по предприятиям:

предприятие 4 – на $X_4(80) = 30$,
предприятие 3 – на $X_3(50) = 0$ или 10,
предприятие 2 – на $X_2(50) = 30$ или $X_2(40) = 40$,
предприятие 1 – на $X_1(20) = 20$ или $X_1(0) =0$ единиц.

7.2.3. Календарное планирование трудовых ресурсов

Ставится задача регулирования численности рабочих во времени при известной минимальной потребности путем найма и увольнения с учетом того, что имеют место убытки в виде некоторых накладных расходов (оплата услуг биржи труда, расходы на обучение, оформление документов и т. п.) и необходимость оплаты простоев при лишней численности.

Введем обозначения:

k – число недель до конца исследуемого периода,

B_k –минимальная потребность в рабочих на первой из этих недель,

Y – наличие рабочих на этой неделе,

$C_1(Z_k)$ – убытки при превышении численности ($Z_k > 0$),

$C_2(S_k)$ – расходы по найму ($S_k > 0$) .

Если обозначить через $F_k(Y_0)$ минимальные расходы в течение k недель при начальной численности Y_0, то при $k > 1$

$$F_k(Y_0) = min\ [C_2(Y–Y_0) + C_1(Y–B_k) + F_{k–1}(Y)\ ] .$$

Значения Y, обеспечивающие здесь минимум, дают оптимальную численность на первой из k недель (на первом шаге k-шагового процесса) при начальной численности Y_0. Обозначаем их через $Y_k(Y_0)$.

Пример. Пусть потребности в рабочих в течение 5 недель равны соответственно 5 , 7 , 8 , 4 и 6 . Функции затрат равны

$$C1(Z)=\begin{cases}3\cdot Z\ ,\ Z>0\\ 0\ ,\ Z\le 0\end{cases}\ ,\qquad C2(S)=\begin{cases}4+2\cdot S\ ,\ S>0\\ 0\ ,\ S\le 0\end{cases}\ .$$

Откажемся от единой сетки значений Y_0. Так как перед последней неделей численность составляла не менее 4 человек, то отыскиваем

$$F_1(Y_0)=\min_{Y=6,7,\ldots}\ [\ C_2(Y-Y_0)+\ C_1(Y-6)\ ]$$

при возможной предыдущей численности $Y_0 = 4$, 5 , 6 , 7 и т.д.

Так как искомая численность соответствует последнему шагу процесса, то использовать численность свыше 6 бессмысленно и минимум достигается при $Y = 6$; при предыдущей численности 4 принять 2 рабочих, при численности 5 принять одного, при численности 6 приема не производить и при большей численности провести соответствующие увольнения (затраты на увольнение мы при постановке задачи не предусмотрели).

Y_0	$F_1(Y_0)$	$Y_1(Y_0)$
4	8	6
5	6	6
6	0	6
>6	0	6

Так как перед четвертой неделей численность составляла не менее 8 человек, то

$$F_2(Y_0)=\min_{Y=4,5,\ldots}\ [\ C_2(Y-Y_0)+C_1(Y-4)+\ F_1(Y)\ ]$$

при возможной предыдущей численности $Y_0 = 8$, 9 , 10 и т. д.

Y_0	Y=4	Y=5	Y=6	Y=7	Y>7	$F_2(Y_0)$	$Y_2(Y_0)$
8	0+0+8	0+3+6	0+6+0	0+9+0	. . .	6	6
>8	0+0+8	0+3+6	0+6+0	0+9+0	. . .	6	6

Так как перед третьей неделей численность составляла не менее 7 человек, то

$$F_3(Y_0)=\min_{Y=8,9,\ldots}\ [\ C_2(Y-Y_0)+C_1(Y-8)+\ F_2(Y)\ ]$$

при возможной предыдущей численности $Y_0 = 7$, 8 , 9 и т. д.

Y_0	Y=8	Y=9	Y=10	Y=11	Y>11	$F_3(Y_0)$	$Y_3(Y_0)$
7	6+0+6	8+3+6	10+6+6	12+9+6	. . .	12	8
8	0+0+6	6+3+6	8+6+6	10+9+6	. . .	6	8
9	0+0 +6	0+3+6	6+6+6	8+9+6	. . .	6	8
>9	0+0 +6	0+3+6	0+6+6	6+9+6	. . .	6	8

Так как перед второй неделей численность составляла не менее 5 человек, то

$$F_4(Y_0) = \min_{Y=7,8,\dots} [\, C_2(Y-Y_0) + C_1(Y-7) + F_3(Y) \,]$$

при возможной предыдущей численности $Y_0 = 5, 6, 7$ и т. д.

Y_0	Y=7	Y=8	Y=9	Y=10	Y>10	$F_4(Y_0)$	$Y_4(Y_0)$
5	8+0+12	10+3+6	12+6+6	14+9+6	. . .	19	8
6	6+0+12	8+3+6	10+6+6	12+9+6	. . .	17	8
7	0+0+12	6+3+6	8+6+6	10+9+6	. . .	12	7
8	0+0+12	0+3+6	6+6+6	8+9+6	. . .	9	8
>8	0+0+12	0+3+6	0+6+6	6+9+6	. . .	9	8

Наконец,

$$F_5(Y_0) = \min_{Y=5,6,\dots} [\, C_2(Y-Y_0) + C_1(Y-5) + F_4(Y) \,]$$

при возможной предыдущей численности $Y_0 = 0$, 1 , 2 и т.д.

Y_0	Y=5	Y=6	Y=7	Y=8	Y>8	$F_5(Y_0)$	$Y_5(Y_0)$
0	14+0+19	16+3+17	18+6+12	20+9+9	. . .	33	5
1	12+0+19	14+3+17	16+6+12	18+9+9	. . .	31	5
2	10+0+19	12+3+17	14+6+12	16+9+9	. . .	29	5
3	8+0+19	10+3+17	12+6+12	14+9+9	. . .	27	5
4	6+0+19	8+3+17	10+6+12	12+9+9	. . .	25	5
5	0+0+19	6+3+17	8+6+12	10+9+9	. . .	19	5
6	0+0+19	0+3+17	6+6+12	8+9+9	. . .	19	5
7	0+0+19	0+3+17	0+6+12	6+9+9	8+12+9	18	7
8	0+0+19	0+3+17	0+6+12	0+9+9	6+12+9	18	7,8
9	0+0+19	0+3+17	0+6+12	0+9+9	0+12+9	18	7,8
10	0+0+19	0+3+17	0+6+12	0+9+9	0+12+9	18	7,8

Если начальная численность не превышала 6, то на первой неделе численность должна быть доведена до 5.

На второй неделе численность доводится до 8, на третьей оставляется той же, на четвертой доводится до 6 и на последней остается неизменной.

В результате такой политики при начальной численности 0 затраты составят

$$[\,(4+2\cdot5)+0\,] + [\,(4+2\cdot3)+3\cdot1\,] + [\,0+0\,] + [\,0+3\cdot2\,] + [\,0+0\,] = 33.$$

Если начальная численность составляла 7 человек, то численность на первой и второй неделях – 7, на третьей – 8 и на последних двух – 6. При начальной численности, большей 7 , на первые две недели устанавливается численность 7 или 8, на третьей –8 и на последних – 6 .

7.2.4. Задачи о замене оборудования

Пусть имеется машина возраста t, которая может в течение сезона обеспечить прибыль в размере $R(t)$, причем в каждом сезоне имеется возможность ее замены новой машиной с затратами $C(t)$.

Найти политику замены, которая обеспечивает максимальную прибыль за N сезонов.

Естественно, что функция прибыли является убывающей по t из-за возрастания эксплуатационных расходов и к какому-то моменту времени оказывается более выгодным произвести модернизацию (замену) оборудования, чем продолжать эксплуатировать старое.

Если обозначить $F_N(t)$ максимальную прибыль за N сезонов от использования машины с начальным возрастом t, то получаем рекуррентные соотношения:

$$F_N(t) = max\begin{cases} -C(t)+R(0)+F_{N-1}(1) \\ R(t)+F_{N-1}(t+1) \end{cases}, \ N>1 \ ; \quad F_1(t) = max\begin{cases} -C(t)\ +R(0) \\ R(t) \end{cases}$$

(здесь первый выбор соответствует варианту замены, второй – отказу от замены).

Решение полученной системы осложняется наличием факта расширения сетки: знание значений $F_{N-1}(t)$ при $t \le T$ не дает возможности найти $F_N(T)$.

Поэтому придется воспользоваться здравым смыслом и принять значения $F_N(t)= 0$ при любом N и t, больших некоторого достаточно большого T.

Такой подход позволяет воспользоваться обычным вычислительным алгоритмом на сетке значений t от 0 до T.

В реальности идет непрерывное совершенствование технологических параметров выпускаемого оборудования и приходится учитывать не только срок службы, но и момент изготовления машины.

Поэтому более естественным будет рассмотрение этой задачи в условиях, где прибыль и затраты на замену являются функциями от времени:

$$F_N(t)=max\begin{cases} -C(t)+R(0)+F_{N-1}(1) \\ R(t)+F_{N-1}(t+1) \end{cases}, N>1; \ F_1(t)= max\begin{cases} -C_1(t)+R_1(0) \\ R_1(t) \end{cases}$$

(здесь функции $R_N(t)$, $C_N(t)$ определяют прибыль и затраты в первом из N сезонов).

Можно рассмотреть и другие постановки данной задачи. Например, добавить возможность покупки использованной машины возраста L и появление третьего выбора: $-S_N(t)+R_N(L)+F_{N-1}(L+1)$.

7.2.5. Задачи на узкие места. Двухотраслевой экономический комплекс

Как правило, никакая отрасль экономики не существует без взаимосвязи с другими отраслями: сельское хозяйство выступает потребителем продукции машиностроения и энергетики, равно как машиностроение и энергетика прямо или косвенно пользуется сельхозпродукцией. Обеспечение эффективного функционирования комплекса взаимосвязанных отраслей экономики является исключительно важной проблемой. Следует откровенно сказать, что в реальной жизни нет оптимальных решений, объявляемое оптимальным – часто плод спекуляций и реже лишь субъективное мнение, опирающееся на познанные количественные оценки взаимосвязей, учет случайных воздействий и принятые критерии оценки эффективности.

Тем не менее, иногда обнаруживаются комплексы, которые мало зависят от других и характеризуются известными жесткими взаимосвязями. Для них и возникают т. н. задачи на «узкие места» – задачи поиска производственных факторов, определяющих самые жесткие условия существования комплексов.

Рассматриваемая здесь задача является простейшим примером задач на «узкие места», достаточно внимательно рассмотренных в работах Р. Беллмана и С. Дрейфуса [19, 20]. Здесь же читатель может убедиться, что описанные ранее вычислительные процедуры метода рекуррентных соотношений могут послужить базой решения и такого рода задач.

Рассмотрим простейший вариант для комплекса автомобильной и сталелитейной промышленности в предположении, что в каждый момент известны запасы сырья, необходимого для производства, и производственные мощности.

В дискретные моменты времени приходится принимать решение об использовании имеющейся стали («узкое место» в производственном процессе):

– использовать ее (полностью или частично) в производстве дополнительной стали при имеющихся мощностях;

– использовать ее для увеличения существующей мощности производства стали;

– использовать ее для производства автомобилей.

Пусть в момент t состояние системы определяется величинами:

$X_s(t)$ – количество стали на складах;

$X_m(t)$ – мощность сталелитейных заводов;

(мощность автостроения для простоты считается неограниченной).

Обозначим количество используемой стали:

$Z_s(t)$ – для производства дополнительной стали;

$Z_m(t)$ –для прироста мощности сталелитейных заводов;

$Z_a(t)$ –для производства автомобилей.

В этих обозначениях можно представить условия:

$Z_a(t) \le A_1 \cdot X_s(t)$ – на автомобили расходуется не свыше определенного процента запасов стали ($A_1 < 1$) ;

$Z_s(t) \le X_m(t)$ – производство стали ограничено имеющимися мощностями;

$X_s(t+1)$ = остаток от предыдущего периода + $A_2 \cdot Z_s(t)$ – количество произведенной стали в интервале времени пропорционально расходу стали на ее воспроизводство;

$X_m(t+1)=X_m(t) + A_3 \cdot Z_m(t)$ – прирост мощности пропорционален расходу стали.

Примем выпуск автомобилей пропорциональным соответствующему расходу стали и обозначим через $F_N(C_1, C_2)$ максимальный возможный расход стали для производства автомобилей за N временных интервалов при начальном запасе стали $X_s(0) = C_1$ и начальной мощности сталелитейного производства $X_m(0)=C_2$.

Тогда

$$F_1(C_1, C_2) = A_1 \cdot C_1 ,$$

$$F_N(C_1, C_2) = max\,\{Z_a + F_{N-1}(C_1 - Z_a - Z_s - Z_m + A_2 \cdot Z_{s-1},\ C_2 + A_3 \cdot Z_m)\}$$

где область максимизации определяется условиями $Z_s, Z_m, Z_a \ge 0$ и

$$Z_s + Z_m + Z_a \le C_1;$$

$$Z_a \le A_1 C_1;$$

$$Z_s \le C_2 .$$

Нетрудно обнаружить здесь необходимость решения последовательности трехмерных задач линейного программирования с сохранением двумерных таблиц $F_k(C_1, C_2)$ и трех оценок оптимальных выборов.

7.2.6. Задача о трудной переправе

Рассмотрим задачу, которая по своей постановке кажется далекой от каких-либо приложений, но в свое время увлекательную для тысяч школьников [20] :

Группа, состоящая из 3 людоедов и 3 миссионеров, пытается переправиться через реку. В имеющейся лодке помещаются не более двух человек. Если число людоедов по какую-то сторону реки или в лодке превышает число миссионеров, то людоеды предаются кровожадным инстинктам и пожирают миссионеров. Каким должен быть план перевозок, чтобы после переправы вся группа осталась в живых ?

Очевидно, что обычный перебор вариантов является достаточно трудоемким и возрастает с ростом числа людей в группе.

Возьмем более общую ситуацию: по одну сторону реки находится M_1 миссионеров и N_1 людоедов, по другую – соответственно M_2 и N_2 и в лодке размещается не более K человек.

Пусть правило, предотвращающее пожирание миссионеров, определяется ограничением $R(m, n) > 0$, где m, n – количество миссионеров и людоедов в образовавшейся подгруппе (для нашей задачи это ограничение сводится к требованию: $m > n$ или $m = 0$ (друг друга людоеды не едят).

Так как общее число миссионеров и людоедов известно, то состояние изучаемой системы в любой момент определяется числом людей на одном берегу. Соответственно, введем функцию $F_L(M_1,N_1)$ – максимальное число людей на втором берегу после L переправ, если мы начинаем с M_1 , M_2 миссионеров и N_1 , N_2 людоедов на первом и втором берегах.

Если обозначить через X_1 , Y_1 и X_2, Y_2 число миссионеров и людоедов в лодке при переправе туда и обратно, то по принципу оптимальности

$$F_L(M_1, N_1) = max\, F_{L-1}(M_1 - X_1 + X_2, N_1 - Y_1 + Y_2) \,,\; L > 1 \,,$$

где область максимизации определяется условиями

$$0 \le X_1 \le M_1 \;;\; 0 \le X_2 \le M_2 + X_1 \;;\; 0 \le Y_1 \le N_1 \;;\; 0 \le Y_2 \le N_2 + Y_1 \;;$$
$$X_1 + Y_1 \le K \;;\; X_2 + Y_2 \le K;\; R(X_1, Y_1) \ge 0 \;;\; R(X_2, Y_2) \ge 0 \;;$$
$$R(M_1 - X_1, N_1 - Y_1) \ge 0 \;;\; R(M_1 - X_1 + X_2, N_1 - Y_1 + Y_2) \ge 0 \;;$$
$$R(M_2 + X_1, N_2 + Y_1) \ge 0 \;;\; R(M_2 + X_1 - X_2, N_2 + Y_1 - Y_2) \ge 0 \,.$$

В случае одношагового процесса имеем

$$F_1(M_1, N_1) = max\, [\, M_2 + X_1 + N_2 + Y_1 \,] \quad .$$

Построенными рекуррентными соотношениями нужно пользоваться до тех пор, пока не обнаружится значение L, при котором

$$F_L(M_1, N_1) = M_1 + N_1 + M_2 + N_2.$$

Применим полученные соотношения к нашей первичной задаче.

Перебирая допустимые сочетания значений M_1 , N_1 и запоминая лишь часть оптимальных выборов, имеем таблицу значений $F_L(M_1, N_1)$.

Отыскиваем оценки F_2 для состояний, где F_1 отлична от 6 (когда все оказались на втором берегу, незачем ездить с берега на берег).

M_1,N_1	$F_2(M_1,N_1)$	X_1	Y_1	X_2	Y_2
(0,3)	6	0	2	0	1
(2,2)	4	0	2	1	0
(3,0)	6	2	0	0	1
(3,1)	3	1	1	0	1
(3,2)	2	0	1	0	1
(3,3)	2	0	2	1	0

Аналогично оценки F_3 для состояний, в которых F_2 отлична от 6

M_1, N_1	$F_3(M_1,N_1)$	X_1	Y_1	X_2	Y_2
(2,2)	6	0	2	1	0
(3,1)	4	0	1	0	2
(3,2)	3	1	1	1	0
(3,3)	2	1	1	1	1

Точно так же находим F_4 и F_5

M_1, N_1	$F_4(M_1,N_1)$	X_1	Y_1	X_2	Y_2
(3,1)	6	1	0	0	1
(3,2)	4	0	2	0	1
(3,3)	3	0	2	0	1

M_1, N_1	$F_5(M_1,N_1)$	X_1	Y_1	X_2	Y_2
(3,2)	6	0	2	0	1
(3,3)	4	0	2	0	1

и, наконец, значение

$$F_6(3,3) = max[F_5(3,2), F_5(3,3)] = 6 \text{ при } X_1=0,\ Y_1=2,\ X_2=0,\ Y_2=1\ .$$

M_1,N_1	$F_1(M_1,N_1)$	X_1	Y_1	X_2	Y_2
(0,0)	6	0	0	0	0
(0,1)	6	0	1	0	0
(0,2)	6	0	2	0	0
(0,3)	5	0	2	0	0
(1,1)	6	1	1	0	0
(2,2)	3	1	1	0	1
(3,0)	4	2	0	0	1
(3,1)	3	2	0	0	1
(3,2)	2	1	1	0	1
(3,3)	1	1	1	1	0

Таким образом мы обнаруживаем возможность реализовать нашу задачу за 6 шагов с помощью политики по шагам:

	Берег1	Лодка →	Лодка ←	Берег2		Берег1	Лодка →	Лодка ←	Берег2
1.	3+3	0+2	0+1	0+1	4.	2+2	0+2	1+0	0+3
2.	3+2	0+2	0+1	0+2	5.	3+0	2+0	0+1	2+2
3.	3+1	1+0	0+1	1+1	6.	1+1	1+1	0+0	3+3

Существуют и другие подобные задачи (на переливание жидкостей между двумя сосудами, на взвешивание и т. п.), которые могут быть исследованы аналогичной методикой.

8. БЕСКОНЕЧНОШАГОВЫЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ

8.1. Бесконечношаговая аппроксимация и функциональные уравнения

Обратимся к рекуррентным соотношениям, построенным ранее для задачи разделения денежного ресурса:

$$F_{k+1}(X) = \max_{0 \le Y \le X} \{ g(Y) + h(X-Y) + F_k [\alpha Y + \beta (X-Y)] \}, \quad k = 2 \,..\, N-1;$$

$$F_1(X) = \max_{0 \le Y \le X} \{ g(Y) + h(X-Y) \}.$$

Как показано выше, при численном их решении объем хранимой информации имеет порядок $2\,MN$, где M – число узлов таблиц. Если даже не сохранять предшествующие таблицы $F_k(X)$ (таблицы поведений $Y_k(X)$ нуждаются в сохранении для последующего нахождения оптимальной политики по шагам), то объем хранимой информации сведется к $M\,(N+1)$. Само собой, что при больших N значительны как емкость сохраняемой информации, так и затраты времени вычислений.

Если $\alpha, \beta < 1$, то значения $\alpha \cdot Y + \beta (X-Y) < X$ с ростом N стремятся к нулю, т. е. при больших N значения $F_N(X)$ и $F_{N+1}(X)$ оказываются достаточно близкими. Отсюда напрашивается мысль о замене процесса большой длительности бесконечношаговым и переходе от рекуррентных соотношений к *функциональному уравнению* (в отличие от обычных уравнений, решением функциональных являются функции, а не скалярные величины) :

$$F(X) = \max_{0 \le Y \le X} \{ g(Y) + h(X-Y) + F(\alpha\, Y + \beta (X-Y)) \},$$

где $F(X)$ – максимальный доход в бесконечношаговом процессе с начальным ресурсом X (оптимальный выбор на первом шаге такого процесса будем обозначать через $Y(X)$).

Оставляя открытым вопрос о правомерности такого предельного перехода, остановимся на некоторых математических и вычислительных аспектах.

Проблемы решения функциональных уравнений связаны с рассмотрением условий существования $F(X)$ и единственности этого решения, а также условий непрерывности $F(X)$ и $Y(X)$.

Очевидно, что можно предложить такие функции $g(x)$ и $h(x)$ и параметры α, β, при которых решение не существует. Так для уравнения

$$F(X) = \max_{0 \le Y \le X} \{ Y + F(\alpha\,(X-Y)) \}$$

при $\alpha > 1$ решения нет (искомая функция стремится к бесконечности).

В [19] , например, показано, что для уравнения

$$F(X) = \max_{0 \le Y \le X} \left(e^{-10/Y} + e^{-15/(X-Y)} + F(0.8Y + 0.9(X-Y) \right)$$

обнаружатся существование, единственность и непрерывность решения $F(X)$, но разрывность функции поведения $Y(X)$.

Поэтому переход от рекуррентных соотношений к функциональным уравнениям должен сопровождаться доказательством существования и единственности решения, что на практике почти нереально. Соответственно, обычно пытаются решить уравнение и убедиться в наличии или отсутствии решения (вопрос о единственности решения остается открытым).

8.2. Методы решения функциональных уравнений

Для решения функциональных уравнений применяют два основных метода: *приближение в пространстве функций* и *приближение в пространстве поведений*.

Возьмем функциональное уравнение общего вида

$$F(X) = \max_{y} R[\, y, F(X)\,] .$$

Приближение в пространстве функций начинается выбором некоторой функции $F_0(X)$ с последующим использованием итерационного процесса (процесса последовательных приближений)

$$F_n(X) = \max_{y} R[\, y, F_{n-1}(X)\,] \;, \quad n = 1, 2, ...$$

Этот процесс продолжается, пока при некотором n не обнаружится близость $F_n(X)$ и $F_{n-1}(X)$ при всех X: функция $F_n(X)$ принимается за $F(X)$.

Приближение в пространстве поведений начинается с выбора начального поведения $y_0(X)$ и поиска функции $F_0(X)$, являющейся решением уравнения $F(X) = R[y_0(X), F_0(X)]$. Отыскивая $\max\limits_{y} R[\, y, F_0(X)\,]$,

получаем улучшенное поведение $y_1(X)$, для которого из уравнения $F(X) = R[y_1(X), F\ (X)]$ находим функцию $F_1(X)$. Продолжаем аналогичные действия до тех пор, пока очередные приближения поведений не окажутся достаточно близкими. Очевидно, что выбор этого метода наиболее разумен при небольшом числе возможных выборов.

Оба метода требуют сохранять лишь очередное и предыдущее значения приближений для искомых функций и поведений, т. е. емкость хранимых таблиц имеет порядок не более $3M$.

Ниже мы ограничимся лишь примерами задач прикладного характера.

8.3. Задача о кратчайшем пути в транспортной сети

Пусть задана транспортная сеть, в которой отмечены вершина с индексом 0 (вход) и вершина N (выход), и заданы длины дуг L_{ij}, связывающих вершины. Требуется найти путь кратчайшей длины от входа к выходу.

Обозначим через F_i длину кратчайшего пути от i -ой вершины до N -ой.

Тогда F_N=0 и для остальных вершин по принципу оптимальности

$$F_i = \min_{(i\ j)} [\, L_{ij} + F_j \,]\ ,\ i = 0,1, . , N\!-\!1$$

(из какой бы вершины i мы не исходили и в какую бы вершину j мы не перешли, дальнейший путь должен быть кратчайшим).

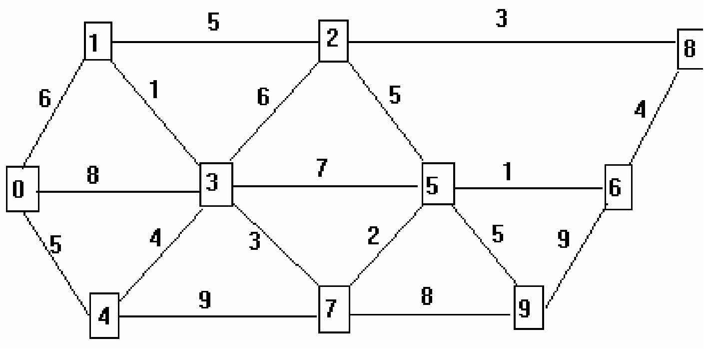
Для решения полученной системы можно воспользоваться приближением в пространстве функций, приняв за начальную функцию, равную 0 при i=N и бесконечности (большому числу) при остальных i.

Осуществляем итерационный процесс

$$F_i^{(k)} = \min_{(i\ j)} [\, L_{i\,j} + F_j^{(k-1)}]\ , i \neq N\, ,\ k = 1, 2, \ldots ; \qquad F_N^{(k)} = 0\ \ , k = 1, 2, \ldots$$

до совпадения k -го и (k–1) -го приближений. Если к тому же запоминать для каждого i индекс последующей вершины j, обеспечивающей минимум, то можно будет найти искомый кратчайший путь.

Пример.



L	0	1	2	3	4	5	6	7	8	9
0		6		8	5					
1	6		5	1						
2		5		6		5			3	
3	8	1	6		4	7		3		
4	5			4				9		
5			5	7			1	2		5
6						1			4	9
7				3	9	2				8
8			3				4			
9						5	9	8		

Задав начальные приближения для F_i равными ∞ (кроме F_N) и перебирая вершины в порядке роста индексов, получаем очередные приближения; например,

$$F_6^{(1)} = \min(1 + F_5^{(1)}, 4 + F_8^{(0)}, 9 + F_9^{(0)}) = \min(1 + 5, 4 + \infty, 9 + 0) = 6\, ,\ j = 5\, ; \ldots$$

$$F_2^{(2)} = \min(3 + F_1^{(2)}, 6 + F_3^{(1)}, 5 + F_5^{(1)}, 3 + F_8^{(1)}) = \min(\infty, \infty, 5 + 5, 3 + 10) = 10\, , \ldots$$

i	F^0	F^1	j^1	F^2	j^2	F^3	j^3	F^4	j^4	F^5	j^5
0	∞	∞	4	∞	4	18	3	17	3	17	3
1	∞	∞	3	∞	3	11	3	11	3	11	3
2	∞	∞	8	10	5	10	5	10	5	10	5
3	∞	∞	1	10	7	10	7	10	7	10	7
4	∞	∞	3	14	3	14	3	14	3	14	3
5	∞	5	9	5	9	5	9	5	9	5	9
6	∞	6	5	6	5	6	5	6	5	6	5
7	∞	7	5	7	5	7	5	7	5	7	5
8	∞	10	6	10	6	10	6	10	6	10	6
9	0	0	-	0	-	0	-	0	-	0	-

Из полученной таблицы легко по значениям j^5 выяснить кратчайший путь [0 – 1 – 3 – 7 – 5 – 9] с длиной $F_0 = 17$. Аналогично можно найти кратчайшие пути от любой вершины до 9-й.

Ускорения сходимости процесса итераций можно добиться, если при поиске k-го приближения ссылаться не на (k–1) -е приближение, а на последнюю из полученных оценок:

$$F_i^{(k)} = \min_{(i\,j)}[\,L_{i\,j} + F_j^{(s)}\,]\ , i \neq N\ ,\ s = \begin{cases} k \\ k-1 \end{cases},\ k = 1, 2, \ldots$$

8.4. Задача о критическом пути в сетевом графике

В сетевом планировании (см. ниже) возникает задача поиска пути максимальной длины (критического пути) от входа к выходу в ориентированной транспортной сети, не содержащей контуров (т. е. возможности возврата к пройденным вершинам).

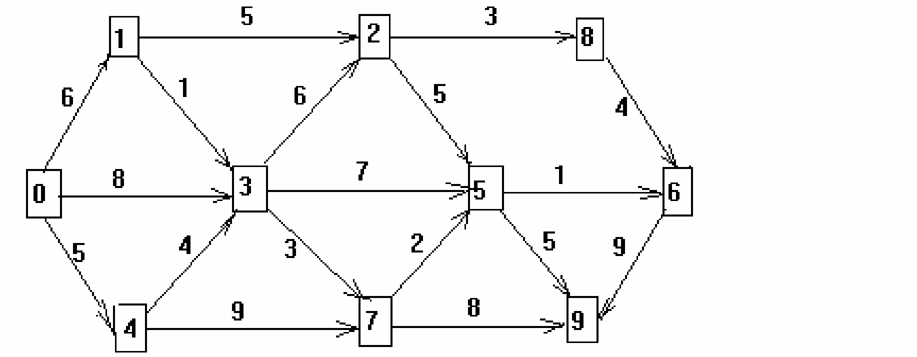
Если обозначить через F_i длину критического пути от i -ой вершины до N–ой, задача сводится к системе функциональных уравнений

$$F_i = \max_{(i\,j)} [\,L_{i\,j} + F_j\,],\ i = 0,1, .. ,N–1; \quad F_N=0\ .$$

Решение этой системы можно, как и в задаче о кратчайшем пути, вести приближением в пространстве функций (приняв начальные приближения для всех вершин, кроме N-ой, равными нулю). Если предварительно проранжировать вершины сети, то процесс итераций можно осуществить за один шаг.

Суть ранжировки поясним на примере нижеприведенной сети.

Отнесем к рангу 0 вершину входа 0. К рангу 1 отнесем вершины, в которые ведут дуги только из вершины ранга 0, т. е. вершины 1 и 4. К рангу 2 относим вершины, в которые ведут дуги только из вершин меньшего ранга, т.е. вершину 3. К рангу 3 отнесутся вершины 2 и 7 , к рангу 4 – вершины 5 и 8, к рангу 5 – вершина 6, к рангу 6 – вершина 9.



Перебирая вершины в порядке убывания ранга, получаем значения

$F_9=0$;

$F_6=9+F_9=9$, $j_6=9$;

$F_5=max[5+F_9, 1+F_6]=10$, $j_5=6$;

$F_8=4+F_6=13$, $j_8=6$;

$F_2=max[5+F_5, 3+F_8]=16$, $j_2=8$;

$F_7=max[2+F_5, 8+F_9]=12$, $j_7=5$;

$F_3=max[6+F_2, 7+F_5, 3+F_7]=22$, $j_3=2$;

$F_1=max[5+F_2, 1+F_3]=23$, $j_1=3$;

$F_4=max[4+F_3, 9+F_7]=26$, $j_4=3$;

$F_0=max[6+F_1, 8+F_3, 5+F_4]=31$, $j_0=4$

и критический путь [0 – 4 – 3 – 2 – 8 – 6 – 9] с длиной 31.

8.5. Выбор критерия оптимальности для бесконечношаговых процессов

Если отвлечься от некоторых элементарных процессов принятия решений, то всякий реальный процесс является частью некоторого бесконечного процесса и всегда можно ставить вопрос о его предыстории и о том, что ожидает в будущем. Если предыстория обычно не представляет интереса для последующего планирования (мы оставляем в стороне пользу опыта), то учет в модели достаточно далекого будущего является необходимым, ибо в противном случае мы за счет оптимизации на фиксированный интервал времени можем создать положение, которое приведет наших потомков к критической ситуации. Человечество уже накопило богатый опыт такого планирования по принципу сиюминутной выгоды.

Можно ли при построении оптимизационной модели учесть далекое будущее? Имеет ли смысл наряду с конечношаговыми процессами изучать бесконечношаговые?

Рассмотрение бесконечношаговых процессов базируется на *гипотезе стационарности*, т. е. постоянства или известной закономерности изменений характеристик среды или, по крайней мере, постоянства вероятностных оценок этих характеристик.

Замена рекуррентных соотношений функциональными уравнениями обеспечивает поиск универсальной политики для любого шага процесса, не говоря об оправданности такого подхода с вычислительной точки зрения.

На практике стационарность редко проявляется слишком долго и многие характеристики среды изменяются под влиянием неучтенных воздействий. Соответственно, найденная универсальная политика перестает быть оптимальной через некоторое время, но правомерна на первых шагах процесса.

На этом и строится т. н. "*скользящее планирование*", где оптимальная политика, выработанная для бесконечношагового процесса, используется на конечном числе шагов и периодически подвергается коррекции (пересчету) с учетом возникающих изменений параметров среды.

При конечном плановом периоде не возникает проблем с выбором критерия эффективности и мы спокойно говорим о максимальном доходе, минимальных издержках и т. п. В бесконечношаговом процессе эти критерии терпят крах, т. к. при любой реальной политике эти величины стремятся к бесконечности и выбор оптимальной политики невозможен. Попытка замены такого процесса N-шаговым при большом N не решает проблему, т. к. оптимальная политика для N - шагового процесса может быть неоптимальна для (N+1) и (N–1) - шагового.

Рассмотрим для примера таблицу значений эффектов по шагам некоторого процесса при разных политиках Q :

Q	шаг1	шаг2	шаг3	шаг4	шаг5	шаг6	. . .
1	3	2	1	3	2	1	. . .
2	3	1	3	1	3	1	. . .
3	1	6	–1	1	6	-1	. . .
4	2	2	2	2	2	2	. . .
5	1	3	1	3	1	3	. . .
6	2	2	1	2	2	1	. . .
7	4	0	0	4	0	0	. . .

Здесь с очевидностью обнаруживается, что политика 6 при любой длительности процесса дает худший эффект в сравнении, например, с политикой 1. Отдать предпочтение какой-либо из других политик затруднительно.

Обычно для бесконечношаговых процессов используются три критерия эффективности:

– *средний эффект* за отрезок времени (СЭ),

– *интегральный дисконтированный эффект* (ИДЭ),

– *эквивалентный средний эффект* (ЭСЭ).

Критерий СЭ интуитивно наиболее прост и определяется отношением суммы эффектов за несколько шагов к числу шагов. Для приведенного примера можно найти оценки СЭ по первым 6 шагам и попытаться сделать грубый прогноз на большую длительность.

Q	шаг1	шаг2	шаг3	шаг4	шаг5	шаг6	. . .	прогноз
1	3	5/2	2	9/4	11/5	2	. . .	2 +
2	3	2	7/3	2	11/5	2	. . .	2 +
3	1	7/2	2	7/4	13/5	2	. . .	2 ±
4	2	2	2	2	2	2	. . .	2
5	1	2	5/3	2	9/5	2	. . .	2 –
7	4	2	4/3	2	8/5	4/3	. . .	< 2

Если мы не ошиблись в прогнозе, с позиций СЭ политики 1 и 2 предпочтительнее других, а политика 7 хуже других (ниже мы уточним эти выводы).

Критерий ИДЭ состоит в вычислении интегрального показателя эффекта, дисконтированного (приведенного) к начальному моменту времени. Если эффект по шагам процесса равен [R_1 ,R_2 , ... , R_n ,...], то

$$\text{ИДЭ} = R_1 + \alpha R_2 + \alpha^2 R_3 + ... + \alpha^{n-1} R_n + ...$$

где $0 \le \alpha < 1$ называют *коэффициентом дисконтирования* (*приведения, скидки*).

Если максимальное из значений R_i (i = 1, 2, ...) меньше некоторого R, то

$$\text{ИДЭ} < R\ (1 + \alpha + \alpha^2 + \alpha^3 + ...) = R / (1 - \alpha)\ ,$$

т. е. при $\alpha < 1$ величина ИДЭ всегда ограничена и тем самым дает возможность сравнения политик.

В основе этого критерия лежит утверждение (гипотеза) об уменьшении полезности денег (денежная сумма сегодня полезнее той же денежной суммы завтра). В самом деле, если фирма обладает денежной единицей и имеет возможность вложений под P % годовых, то через период в k лет она будет обладать $S=(1+P/100)^k$ денежными единицами, т. е. обладание $1/S$ денежной единицы сегодня эквивалентно обладанию полной денежной единицей через период. В такой ситуации коэффициент дисконтирования равен

$$\alpha = \left| 1 \big/ 1 + \frac{p}{100} \right|^k$$

и убывает с ростом процентной ставки.

Сравним вышеприведенные политики с позиций ИДЭ. Имеем

$$\text{ИДЭ}(1) = 3 + 2\alpha + \alpha^2 + 3\alpha^3 + 2\alpha^4 + \alpha^5 + ... =$$
$$= (3 + 2\alpha + \alpha^2)\,(1 + \alpha + \alpha^2 + \alpha^3 + ...) = (3 + 2\alpha + \alpha^2) / (1 - \alpha^3).$$

Аналогично получаем

$$ИДЭ(2) = (3 + \alpha) /(1-\alpha^2) ,$$
$$ИДЭ(3) = (1+ 6\alpha-\alpha^2) / (1-\alpha^3) ,$$
$$ИДЭ(4) = 2 / (1-\alpha) ,$$
$$ИДЭ(5) = (1 +3\alpha) / (1-\alpha^2),$$
$$ИДЭ(6) = 4 / (1-\alpha^3) .$$

Сопоставляя полученные значения, обнаруживаем

ИДЭ(1) ≥ ИДЭ(2) при всех α (при α=0 равенство),
ИДЭ(1) ≥ ИДЭ(3) при всех α (при α=1 равенство),
ИДЭ(1) ≥ ИДЭ(4) при всех α (при α=1 равенство),
ИДЭ(1) ≥ ИДЭ(5) при всех α (при α=1 равенство),
ИДЭ(1) > ИДЭ(7) при $\alpha > 0.414$.

Очевидно, что политика 1 имеет преимущество над политиками 2, 3, 4, 5 и превосходит политику 7 при α > 0.414, что соответствует P<141 % .

При больших P политика 7 имеет преимущество за счет большей начальной денежной суммы.

Критерий ЭСЭ связывает подходы с позиций ИДЭ и СЭ [22]:

$$ЭСЭ(\alpha) = (1 - \alpha) \cdot ИДЭ(\alpha);$$
$$СЭ = \lim_{\alpha\to 1} ЭСЭ(\alpha) .$$

Для рассмотренного примера имеем

$$ЭСЭ(1) = (3+2\alpha+\alpha^2)/(1+\alpha+\alpha^2), \quad СЭ(1) =2 ;$$
$$ЭСЭ(2) = (3+\alpha)/(1+\alpha), \quad СЭ(2) = 2 ;$$
$$ЭСЭ(3) = (1+6\alpha-\alpha^2)/(1+\alpha+\alpha^2), \quad СЭ(3) = 2 ;$$
$$ЭСЭ(4) = 2, \quad СЭ(4) = 2 ;$$
$$ЭСЭ(5) = (1+3\alpha)/(1+\alpha), \quad СЭ(5) = 2 ;$$
$$ЭСЭ(7) = 4/(1+\alpha+\alpha^2), \quad СЭ(7) = 4/3 .$$

Появившаяся возможность точной оценки СЭ обнаруживает, что с позиций СЭ политика 4 хуже остальных политик.

8.6. Простейшая задача управления запасами: конечношаговый процесс

Рассмотрим задачу разработки программы выпуска некоторого изделия на плановый период из N отрезков времени в предположении наличия точного прогноза спроса на эту продукцию.

Пусть продукция, изготавливаемая в течение отрезка t, может быть использована для покрытия спроса на этом отрезке и имеется возможность хранения остатка до начала очередного отрезка времени.

Требуется разработать программу, при которой общая сумма затрат на производство и хранение минимальна при условии полного и своевременного удовлетворения спроса [25] .

Обозначим через k число временных интервалов до конца процесса, через X_k – объем производства в первом из предстоящих k периодов (при $N = 5$ индекс $k = 5$ соответствует первому шагу, $k = 4$ – второму и т. п.), через Y_k – уровень запаса на конец этого периода, через $C_k(X_k,Y_k)$ – функцию затрат на производство и хранение в этом периоде. Аналогично обозначим через S_k – спрос на продукцию, R_k – максимальный объем производства и T_k – максимальную емкость склада.

Если обозначить через $F_k(Y)$ минимальные затраты в k–шаговом процессе с начальным запасом Y, а через $X_k(Y)$ – оптимальный объем производства на первом шаге такого процесса, то возникает система рекуррентных соотношений

$$F_k(Y) = min\ [\ C_k(X_k,\ Y+X_k–S_k\ ) + F_{k–1}(Y+X_k –S_k\ )\ ]\ ,\ k = 2\ ...\ N\ ,$$

где область минимизации определяется условиями

$$0 \le X_k \le R_k\ ,\ 0 \le Y+X_k–S_k \le T_k\ .$$

В случае одношагового процесса напрашивается очевидная политика: производить лишь минимум продукции, необходимый для покрытия спроса, т. е.

$$F_1(Y) = C_1(S_1 –Y,\ 0)\ ,\quad X_1(Y) = S_1 –Y \quad \text{при } Y \le S_1$$

$$F_1(Y) = C_1(0,\ Y–S_1)\ ,\quad X_1(Y) = 0 \qquad \text{при } Y \ge S_1 .$$

Рассмотрим пример 6-шагового процесса с характеристиками производственной мощности R_k =5, емкости склада T_k =4 и спроса S_k =3, неизменными на всех шагах процесса. Возьмем неизменную функцию затрат $C_k(X,Y) = C(X) + H(Y)$, где $H(Y) = H{\cdot}Y$ – стоимость хранения Y единиц продукции, пропорциональная объему хранения. Если принять производственные затраты также пропорциональны объему, можно убедиться, что минимизация затрат обеспечится требованием производить лишь требуемое для покрытия спроса на каждом шаге и не тратить деньги на создание запасов. Поэтому возьмем ее нелинейной в виде:

$$C(X) = \begin{cases} A + B \cdot X\ , & X > 0 \\ \quad 0 & , \ X = 0 \end{cases}$$

(если в текущем интервале времени производится продукция, то, кроме непосредственных затрат, пропорциональных объему, присутствуют косвенные затраты на содержание контролеров, слесарей-ремонтников и др.). В этом случае в зависимости от стоимостных показателей H, A, B может возникнуть резон усиленной работы некоторое время с тем, чтобы затем какое-то время жить созданными запасами.

Пусть A=13 , B=2 , H=1. Примем диапазон целочисленных значений Y от 0 до 4 и с учетом

$$\text{при } Y < 3 \quad F_1(Y) = 13 + 2\cdot(3-Y)\,, \quad X_1(Y) = 3-Y;$$
$$\text{при } Y \geq 3 \quad F_1(Y) = 1\cdot(Y-3)\,, \quad X_1(Y) = 0\,;$$

имеем

Y	0	1	2	3	4
$F_1(Y)$	19	17	15	0	1
$X_1(Y)$	3	2	1	0	0

При $k>1$ $F_k(Y) = min\ [\begin{Bmatrix} 13+2\cdot X\ , & X>0 \\ 0\ , & X=0 \end{Bmatrix} + 1\cdot(Y+X-3) + F_{k-1}(Y+X-3)]$,

где область минимизации определена требованиями

$$0 \leq X \leq 5\ ,\ 0 \leq Y+X-3 \leq 4\ .$$

При $k = 2$ имеем расчетную таблицу значений

Y	X=0	X=1	X=2	X=3	X=4	X=5	$F_2(Y)$	$X_2(Y)$
0				19+0+19	21+1+17	23+2+15	38	3
1			17+0+19	19+1+17	21+2+15	23+3+0	26	5
2		15+0+19	17+1+17	19+2+15	21+3+0	23+4+1	24	4
3	0+0+19	15+1+17	17+2+15	19+3+0	21+4+1		19	0
4	0+1+17	15+2+15	17+3+0	19+4+1			18	0

При $k = 3$

Y	X=0	X=1	X=2	X=3	X=4	X=5	$F_3(Y)$	$X_3(Y)$
0				19+0+38	21+1+26	23+2+24	48	4
1			17+0+38	19+1+26	21+2+24	23+3+19	45	5
2		15+0+38	17+1+26	19+2+24	21+3+19	23+4+18	43	4
3	0+0+38	15+1+26	17+2+24	19+3+19	21+4+18		38	0
4	0+1+26	15+2+24	17+3+19	19+4+18			27	0

и т. д., получая в итоге сводную таблицу $F_k(Y)$, $X_k(Y)$.

Y	$F_1(Y)$	$X_1(Y)$	$F_2(Y)$	$X_2(Y)$	$F_3(Y)$	$X_3(Y)$	$F_4(Y)$	$X_4(Y)$	$F_5(Y)$	$X_5(Y)$	$F_6(Y)$	$X_6(Y)$
0	19	3	38	3	48	4	67	3,4	79	5	96	4
1	17	2	26	5	45	5	64	5	74	5	93	5
2	15	1	24	4	43	4	54	5	72	4	91	4
3	0	0	19	0	38	0	48	0	67	0	79	0
4	1	0	18	0	27	0	46	0	65	0	75	0

Если уровень запаса на начало процесса равен 0, то оптимальные объемы производства по периодам от начала процесса равны

$$X_1 = X_6\,(0) = 4\ , \quad X_2 = X_5(0+4-3) = 5\ , \quad X_3 = X_4\,(1+5-3) = 0\ ,$$
$$X_4 = X_3(3+0-3) = 4\ , \quad X_5 = X_2\,(0+4-3) = 5\ , \quad X_6 = X_1(1+5-3) = 0\ .$$

Здесь мы наблюдаем двукратное повторение последовательности объемов производства – производственный цикл [4 – 5 – 0] и средние затраты 96/6=16. Сохранится ли наблюдаемое явление для большей длительности? Вычислительный эксперимент дает положительный ответ при N=9 и 12, но для N=15 уже обнаруживается цикл [5–5–0–5–0] и средние затраты составят 15.8 <16. А для длительности 52 или 365 ? Конечно, при наличии программного средства это нетрудно выяснить, но может быть разумнее перейти от большой (!?) длительности к бесконечной и не тратить зря время и деньги?

8.7. Простейшая задача управления запасами: бесконечношаговый процесс

При неизменных во времени спросе и других характеристиках рассматриваемый процесс проявляет свойства стационарности – стремления некоторых итоговых его характеристик к постоянству .

Если обозначить через $F(Y)$ величину интегрального дисконтированного эффекта (ИДЭ) при начальном запасе Y и использовании оптимальной политики, то возникает функциональное уравнение

$$F(Y) = min\ [\ C(X, Y+X-S) + \alpha\ F(Y+X-S)\ ]\ ,$$

где α – коэффициент дисконтирования и область минимизации определяется условиями $0 \le X \le R$, $0 \le Y + X - S \le T$.

Поставленную задачу при дискретных спросе и объеме производства, по соображениям компактности записи, можно интерпретировать как систему переходов между состояниями (уровнями запаса).

Если обозначить Y через i, $Y+X-S$ через j, $C(X, Y+X-S)$ через C_{ij}, $F(Y)$ через $F(i)$, то C_{ij} можно понимать как затраты на переход из состояния i в состояние j и функциональное уравнение переписать в виде:

$$F(i) = \min_{j}\ [\ C_{ij} + F(j)\ ]\ , \qquad i = 0, 1, \ldots, T\ .$$

Обозначим через g средние затраты (СЭ) и через F_i – составляющие затрат, определяемые начальным состоянием (запасом). Тогда

$$F(i) = F_i + g\ /\ (1-\alpha)$$

и при $\alpha = 1$ приведенные выше функциональные уравнения примут вид:

$$F_i + g = \min_{j}\ [\ C_{ij} + F_j\ ]\ , \qquad i = 0, 1, \ldots, T.$$

Решение этих уравнений осуществляем приближением в поведениях.

Обратимся к примеру, рассмотренному в 8.6, и соответствующую систему управления запасами будем интерпретировать как систему с пятью состояниями (i = 0, 1, 2, 3, 4).

Для удобства последующей работы заблаговременно рассчитаем таблицу значений C_{ij} ($j = i+X-3$), где учитываются затраты на производство, равные $13+2\cdot X$ при $X > 0$ и нулю при $X = 0$, и затраты на хранение остатков, равные $1\cdot j$ (X не превышает 5).

$C_{ij} =$

$i \setminus j$	0	1	2	3	4
0	19+0	21+1	23+2		
1	17+0	19+1	21+2	23+3	
2	15+0	17+1	19+2	21+3	23+4
3	0+0	15+1	17+2	19+3	21+4
4		0+1	15+2	17+3	19+4

Берем за начальное поведение вполне разумную политику производства, предлагающую минимальный объем выпуска, лишь достаточный для покрытия текущего спроса.

Тогда для i=0 берем X=3, для i=1 – X=2, для i=2 – X=1, для i=3 и i=4 – X=0: начальное поведение определяется системой переходов

$$(0 \to 0), (1 \to 0), (2 \to 0), (3 \to 0), (4 \to 1).$$

Для этого поведения возникает система 5 уравнений с 6 неизвестными, разрешимая с точностью до константы (приняв некоторое F_i, например F_0, равным какой-то константе, другие F_i определяем с точностью до константы, тогда как значение g не зависит от этого выбора):

$$\begin{array}{ll} F_0 + g = 19 + F_0 & F_0 = 0, \quad g = 19, \\ F_1 + g = 17 + F_0 & F_1 = -2, \\ F_2 + g = 15 + F_0 & F_2 = -4, \\ F_3 + g = 0 + F_0 & F_3 = -19, \\ F_4 + g = 1 + F_1 & F_4 = -20. \end{array}$$

Обнаружив, что выбранная политика обеспечивает средние затраты, равные 19, попытаемся найти улучшенное поведение:

$i = 0$: $min[C_{0j} + F_j] = min$ [19+0, 22–2, 25–4, – , –] при j=0;
$i = 1$: $min[C_{1j} + F_j] = min$ [17+0, 20–2, 23–4, 26–19, –] при j=3;
$i = 2$: $min[C_{2j} + F_j] = min$ [15+0, 18–2, 21–4, 24–19, 27–20] при j=3;
$i = 3$: $min[C_{3j} + F_j] = min$ [0+0, 16–2, 19–4, 22–19, 25–20] при j=0;
$i = 4$: $min[C_{4j} + F_j] = min$ [– , 1–2, 17–4, 20–19, 23–20] при j=1.

Найденное улучшенное поведение определяет систему переходов:

$$(0 \to 0), (1 \to 3), (2 \to 3), (3 \to 0), (4 \to 1).$$

Строим и решаем соответствующую систему уравнений

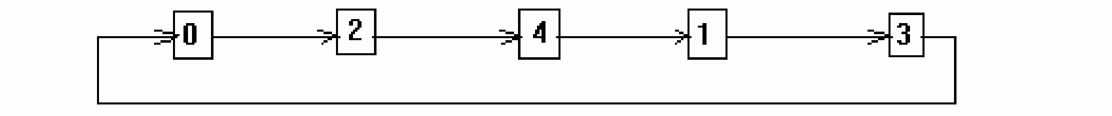
$$\begin{array}{ll} F_0 + g = 19 + F_0 & F_0 = 0, \quad g = 19, \\ F_1 + g = 26 + F_3 & F_1 = -12, \\ F_2 + g = 24 + F_3 & F_2 = -14, \\ F_3 + g = 0 + F_0 & F_3 = -19, \\ F_4 + g = 1 + F_1 & F_4 = -30. \end{array}$$

Видим, что и эта выбранная политика дает средние затраты, равные 19.

Аналогичные попытки улучшения дают поведения:

$$(0 \to 1), (1 \to 3), (2 \to 3), (3 \to 4), (4 \to 1), g = 17.3;$$
$$(0 \to 2), (1 \to 2), (2 \to 4), (3 \to 0), (4 \to 1), g = 17;$$
$$(0 \to 2), (1 \to 3), (2 \to 3), (3 \to 0), (4 \to 1), g = 16.3;$$
$$(0 \to 1), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1), g = 16;$$
$$(0 \to 2), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1), g = 15.8;$$
$$(0 \to 2), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1).$$

Поскольку два очередных приближения в поведениях совпали, можно утверждать, что оптимальная политика переходов между состояниями имеет вид:



и оптимальный производственный цикл определяется последовательностью объемов производства [5 – 5 – 0 – 5– 0]. В частности, при скользящем планировании найденную политику можно принять за отправную.

8.8. Бесконечношаговый процесс замены оборудования

Обратимся к задаче, рассмотренной в предшествующей главе для условий стационарного режима:

$$F_n(t) = max \left[\begin{array}{l} -C(t) + R(t) + F_{n-1}(1) \\ \quad R(t) + F_{n-1}(t+1) \end{array} \right.$$

Если рассмотреть бесконечношаговый процесс и обозначить $F(t)$ максимальный дисконтированный эффект при начальном возрасте t, то

$$F(t) = max \left[\begin{array}{l} -C(t) + R(0) + \alpha\, F(1) \\ \quad R(t) + \alpha\, F(t+1) \end{array} \right. .$$

Предполагая оптимальным возрастом замены $t = T$, имеем

$F(0)=R(0) + \alpha F(1)$; $F(1)=R(1)+\alpha F(2)$; . . $F(T)= -C(T)+R(0)+\alpha F(1)$,

откуда последовательными подстановками получаем

$$F(0) = R(0) + \frac{\alpha}{1-\alpha^T} \{ R(1) + \alpha R(2) + \alpha^2 R(3) + \ldots + \alpha^{T-1} [R(0) - C(T)] \}$$

Достаточно последовательно вычислять значения функции $F(0)$ при $T = 0, 1, 2,...$ до тех пор, пока сохраняется возрастание этих значений. Предельное значение T соответствует оптимальному возрасту замены машины на новую.

Можно рассмотреть и другие подходы к организации замены оборудования [20].

9. СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ

9.1. Специфика выбора критерия оптимальности

До сих пор мы рассматривали оптимизационные модели, где предполагалась полная **детерминированность** исходных данных, т. е. значения спроса, дохода, затрат и т. п. определялись однозначно. Однако часто значения параметров задачи являются функциями от многих факторов, неучтенных в математической модели, и с позиций решающего задачу выступают в роли случайных величин, для которых известны лишь параметры соответствующего распределения вероятностей или другие вероятностные оценки.

Естественно, что в таких условиях мы не можем говорить о точных значениях прибыли, затрат и т. п., а лишь об ожидаемых значениях этих величин или о вероятностях того, что указанные величины принимают значения в некотором заданном диапазоне.

Более того, если в случае детерминированного процесса обнаруживается одна или несколько оптимальных стратегий, представляющих вполне определенную последовательность управляющих воздействий (такие стратегии называют чистыми), то для стохастического (вероятностного) процесса оптимальная стратегия часто представляет совокупность таких воздействий, смешанных в некоторых пропорциях (соответственно такие стратегии и называют смешанными).

Не претендуя на исчерпывающее изложение этой необъятной тематики, рассмотрим несколько примеров многошаговых стохастических процессов принятия решений.

9.2. Управление запасами в условиях неопределенности

Рассмотрим уже упоминавшуюся ранее задачу управления запасами для ситуации, когда величина спроса является случайной. Соответственно, запланированный в начале очередного периода объем производства может оказаться заниженным по сравнению с возникающим спросом и для покрытия неудовлетворенного спроса приходится нести дополнительные расходы (штрафы, срочные заказы и т. п.).

Предположим, что известны функции (для простоты считаем их независимыми от номера сезона):

$K(X)$ – затраты на запланированный в очередном сезоне объем производства X;

$R(V)$ – дополнительные затраты на компенсацию неудовлетворенного спроса в объеме V;

$P(S)$ – плотность распределения случайной величины спроса S.

Напомним, что в случае непрерывной случайной величины S, математическое ожидание некоторой функции $F(S)$ равно значению интеграла

$$M[F(S)] = \int\limits_{(S)} F(S)P(S)dS,$$

где область интегрирования – диапазон возможных значений S (гипотетически от 0 до ∞) .

Обозначим через $F_k(Y)$ математическое ожидание издержек в k -шаговом процессе при начальном запасе Y и при использовании оптимальной политики.

Очевидно, что затраты на первом шаге процесса при начальном запасе Y равны сумме затрат на производство X единиц продукции, т. е. значению $K(X)$, и ожидаемых затрат на покрытие неудовлетворенного спроса. Так как при $S \le Y + X$ спрос удовлетворяется полностью, то последние равны

$$\int\limits_{Y+X}^{\infty} R(S - Y - X)P(S)dS.$$

Последующий $k-1$ -шаговый процесс начнется с начальным запасом $Y+X-S$ при S в диапазоне от 0 до $Y+X$ и нулевым начальным запасом при S, превысившим $Y+X$.

Тогда в соответствии с принципом оптимальности задача минимизации затрат в процессе конечной длительности сводится к рекуррентным соотношениям вида [19, 20]:

$$F_k(Y) = \min_X \{ K(X) + \int\limits_{Y+X}^{\infty} R(S - Y - X)P(S)dS +$$
$$+ \int\limits_0^{Y+X} F_{k-1}(Y + X - S)P(S)dS + F_{k-1}(0) \int\limits_{Y+X}^{\infty} P(S)dS \}, \ k>1;$$

$$F_1(Y) = \min_X \{ K(X) + \int\limits_{Y+X}^{\infty} R(S - Y - X)P(S)dS \}$$

Нетрудно построить и другие модификации рассматриваемой задачи. Так, если производимая в текущем сезоне продукция становится готовой для покрытия спроса лишь в следующем сезоне, рекуррентные соотношения преобразуются к виду

$$F_k(Y) = \min_X \{ K(X) + \int\limits_{Y}^{\infty} R(S - Y)P(S)dS +$$

$$+\int_0^Y F_{k-1}(Y+X-S)P(S)dS+F_{k-1}(0)\int_Y^\infty P(S)dS\ \},\ k>1\ ;$$

$$F_1(Y)=\min_X\{K(X)+\int_Y^\infty R(S-Y)P(S)dS\}.$$

В случае же аналогичной задержки на два сезона приходится ввести в рассмотрение функцию двух переменных $F_k(Y, Z)$ как минимум ожидаемых затрат в k-шаговом процессе с начальным запасом Y и объемом производства на предыдущем шаге Z:

$$F_k(Y,Z)=\min_X\{K(X)+\int_Y^\infty R(S-Y)P(S)dS+$$

$$+\int_0^Y F_{k-1}(Y-S+Z,X)P(S)dS+F_{k-1}(Z,X)\int_Y^\infty P(S)dS\ \},\ k>1.$$

Очевидно, что аналитическое решение для рекуррентных соотношений и функциональных уравнений получить трудно даже в случае простейших исходных функций. Численное решение, нереальное в эпоху Р. Беллмана, на современной компьютерной базе вполне осуществимо при разумной длительности процесса и не слишком высокой требовательности к точности оценок.

Нетрудно обобщить полученные рекуррентные соотношения и на случай бесконечношагового процесса, если прибегнуть к критерию интегрального дисконтированного эффекта (с возникающими здесь результатами можно ознакомиться в [19]).

Рассмотрим ситуацию, когда спрос определяется дискретной случайной величиной.

Если обратиться к простейшей задаче управления запасами, которая описывалась выше рекуррентными соотношениями

$$F_k(Y)=\min\ \{\ C_k(X)+H_k(Y+X-S_k)+F_{k-1}(Y+X-S_k)\ \}\ ,$$

то в ситуации, когда S_k принимает значения S_{k1} , S_{k2} ,..., S_{km} с вероятностями P_{k1}, P_{k2},..., P_{km} придется определить $F_k(Y)$ как минимальные ожидаемые затраты производства и хранения в k - шаговом процессе с начальным запасом Y.

Учитывая, что для дискретной случайной величины математическое ожидание равно сумме произведений всех ее возможных значений на соответствующие вероятности, получаем рекуррентные соотношения:

$$F_k(Y) = min \ \{ \ C_k(X) + \sum_{i=1}^{m}[\ H_k \ .(Y+X-S_{ki}) + F_{k-1}(Y+X-S_{ki}) \ ] \ P_{ki} \},$$

Область минимизации в условиях обязательного покрытия любого возможного спроса должна учитывать, что спрос должен удовлетворяться в любом случае $Y+X \geq \max\limits_{ki} \{S_{ki}\}$ и что остаток $Y+X-\min\limits_{ki} \{S_{ki}\}$ не превышает емкости склада.

Пример. Пусть максимальный объем производства равен 5, емкость склада равна 4, для всех k H_k=1, $C_k(X)$=13+2X при X>0 и равна 0 при X=0; спрос может быть равен 2 или 4 с равными вероятностями.

Здесь область минимизации определяется условиями $4 \leq Y+X \leq 6$ и $0 \leq X \leq 5$. Отыскивая

$F_1(Y) = min\{C(X)+0.5\cdot (Y+X-2)+0.5\cdot (Y+X-4)\}=min\{C(X)+Y+X-3\}$,

получаем

Y	0	1	2	3	4
$F_1(Y)$	22	20	18	16	1
$X_1(Y)$	4	3	2	1	0

Аналогично находим

$F_2(Y) = min \ \{ \ C(X) +(Y+X-3) + 0.5\cdot F_1(Y+X-2) +0.5\cdot F_1(Y+X-4) \ \}$:

Y	X=0	X=1	X=2	X=3	X=4	X=5	$F_2(Y)$	$X_2(Y)$
0					22+20	25+18	42	4
1				20+20	23+18	26+9.5	35.5	5
2			18+20	21+18	24+9.5		33.5	4
3		16+20	19+18	22+9.5			31.5	3
4	1+20	17+18	20+9.5				21	0

$F_3(Y) = min \ \{ \ C(X) + (Y+X-3) \ + 0.5\cdot F_2(Y+X-2) +0.5\cdot F_2(Y+X-4) \ \}$:

Y	X=0	X=1	X=2	X=3	X=4	X=5	$F_3(Y)$	$X_3(Y)$
0					22+37.75	25+33.5	58.5	5
1				20+37.75	23+33.5	26+27.25	53.25	5
2			18+37.75	21+33.5	24+27.25		51.25	4
3		16+37.75	19+33.5	22+27.25			49.25	3
4	1+37.75	17+33.5	20+27.25				38.75	0

Если в случае детерминированного процесса со спросом 3 для трехшагового процесса с нулевым начальным запасом мы получали оптимальную политику производства по шагам (4, 5, 0), здесь объемы производства несколько выше и равны (5, 4, 1).

Если обратиться к рассмотрению соответствующего бесконечношагового процесса, то функциональное уравнение с учетом дисконтирования имеет вид:

$$F(Y) = min\ \{\ C(X) + \sum_{i=1}^{m}[\ H.(Y+X-S_i\ ) + \alpha F(Y+X-S_i\ )\ ]\ P_i\ \}\ ,$$

Полагая $F(Y) = F_Y + G\ /\ (1-\alpha)$, где G – средние ожидаемые затраты при оптимальной политике и F_Y – составляющие затрат, определяемые начальным ресурсом, при $\alpha = 1$ получаем систему функциональных уравнений

$$F_Y + G = min\ \{\ C(X) + H.(Y+X- \sum_{i=1}^{m} P_i\ S_i\ ) + \sum_{i=1}^{m} P_i\ \ F_k\ \},$$

где k соответствует $Y+X- S_i$.

Для рассмотренного примера приходится решать систему функциональных уравнений (например, приближением в поведениях)

$$F_0 + G = min\ \{\ C(X) + H.(X-3) + 0.5\ (\ F_{X-2} + F_{X-4}\ )\}$$
$$F_1 + G = min\ \{\ C(X) + H.(X-2) + 0.5\ (\ F_{X-1} + F_{X-3}\ )\}$$
$$F_2 + G = min\ \{\ C(X) + H.(X-1) + 0.5\ (\ F_X + F_{X-2}\ )\}$$
$$F_3 + G = min\ \{\ C(X) + H.(X-0) + 0.5\ (\ F_{X+1} + F_{X-1}\ )\}$$
$$F_4 + G = min\ \{\ C(X) + H.(X+1) + 0.5\ (\ F_{X+2} + F_X\ )\}\ ,$$

где область минимизации определяется условиями $0 \le X \le 5$ и значениями индексов в диапазоне от 0 до 4; в развернутой форме

$$F_0 + G = min\{21+1+0.5\ (F_2+F_0),\ 23+2+0.5(F_3+F_1)\}\ , \qquad X = 4,\ 5;$$
$$F_1 + G = min\{19+1+0.5\ (F_2+F_0),\ 21+2+0.5(F_3+F_1),\ 23+3+0.5(F_4+F_2)\},\qquad X = 3,\ 4,\ 5;$$
$$F_2 + G = min\{17+1+0.5(F_2+F_0),\ 19+2+0.5(F_3+F_1),\ 21+3+0.5(F_4+F_2\ )\},\qquad X = 2,\ 3,\ 4;$$
$$F_3 + G = min\{15+1+0.5(F_2+F_0),\ 17+2+0.5(F_3+F_1),\ 19+3+0.5(F_4+F_2)\},\qquad X = 1,\ 2,\ 3;$$
$$F_4 + G = min\{0+1+0.5(F_2+F_0),\ 15+2+0.5(F_3+F_1),\ 17+3+0.5(F_4+F_2)\},\qquad X = 0,\ 1,\ 2.$$

Если в ранее рассмотренном примере детерминированного процесса оптимальные режимы производства в зависимости от начального запаса были равны (5, 5, 5, 0, 0), то решение этой системы приближением в поведениях дает значения (5, 5, 4, 3, 0).

9.3. Дихотомический выбор (задача о золотодобыче)

Если у вас развито воображение, представьте себе, что вы столь состоятельны, что смогли стать обладателями двух золотых приисков *A* и *B* (Р. Беллман [19] называет их *Анаконда* и *Бонанца*) с запасами золота соответственно *X* и *Y*, но по ряду причин смогли обзавестись лишь единственной добывающей машиной. Имеющаяся статистика золотодобычи на Клондайке показывает, что при работе в условиях прииска *A* машина в течение сезона с вероятностью P_1 добывает долю R_1 имеющегося запаса и с вероятностью $1-P_1$ ломается навсегда. При работе на прииске типа B значения вероятностей и доли добычи равны соответственно $P_2, R_2, 1-P_2$.

Требуется найти оптимальную перестановку машины между приисками, которая максимизирует *ожидаемый объем добычи* за N сезонов.

Рассмотрим эту красивую идеализированную задачу ради иллюстрации особенностей недетерминированного процесса и некоторых технических приемов, не бесполезных при решении других задач.

Обозначим через $F_k(X,Y)$ максимум ожидаемого суммарного объема добычи за k сезонов при начальных запасах X и Y (читатель должен еще раз вспомнить понятие математического ожидания для дискретных случайных величин). Поскольку есть лишь два выбора для установки машины в очередном сезоне, то

$$F_k(X,Y) = max \begin{bmatrix} A: P_1 \{ R_1 X + F_{k-1}((1-R_1)X, Y) \} \\ B: P_2 \{ R_2 Y + F_{k-1}(X, (1-R_2)Y) \} \end{bmatrix} \qquad (1)$$

$$F_1(X,Y) = max \begin{bmatrix} A: P_1 R_1 X \\ B: P_2 R_2 Y \end{bmatrix} \qquad (2)$$

Решение полученных рекуррентных соотношений обычным численным методом осложняется двухмерностью искомых функций и, если бы возникающие здесь функции не были линейными, то численное решение потребовало бы сохранения N двумерных таблиц, по крайней мере, для поведений. Здесь же на помощь приходит факт, что эти функции обладают свойством однородности $F(r\cdot X, r\cdot Y) = r\cdot F(X,Y)$, откуда можно сделать вывод, что оптимальная политика для точки (X,Y) остается оптимальной для всех точек луча из (0, 0), проходящего через указанную точку. Следовательно, оптимальная политика определяется не самими значениями X и Y, а лишь соотношением между ними.

Если учесть, что при $X >> Y$ оптимальным первым выбором является выбор A, а при $Y>>X$ – выбор B, то совокупность всех допустимых точек (X, Y) (положительный квадрант плоскости) разбивается на две области оптимального первого выбора (рис. 1) .

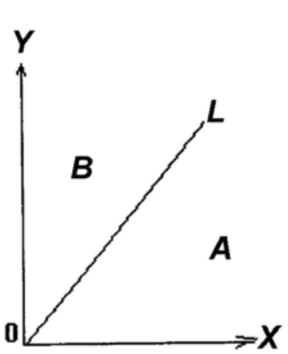


Рис. 1

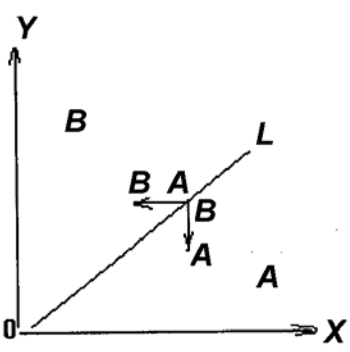


Рис. 2

Возникает вопрос о поиске граничной линии L, где оба выбора равноценны.

При k=1 она определяется элементарно уравнением

$$P_1 R_1 X = P_2 R_2 Y \ . \tag{3}$$

При k>1 воспользуемся следующими соображениями. Если в точке (X, Y) пограничной линии L использовать выбор A, то последующее состояние определяется точкой $((1–R_1)X, Y)$, которая лежит в области первого выбора B. При использовании в точке (X,Y) линии L первого выбора B возникает состояние, определяемое точкой $(X, (1–R_2)Y)$, для которой заведомо оптимален первый выбор A.

Следовательно, для точек граничной линии равноценны выборы «*AB + оптимальное продолжение*» и «*BA + оптимальное продолжение*» (рис. 2).

Для этих последовательностей выборов находим представления:

$$(AB+):\ F_k(X,Y) = P_1 [R_1X + F_{k-1}((1-R_1)X, Y)] =$$
$$= P_1 [R_1X + P_2 [R_2 Y + F_{k-2}((1-R_1)X, (1-R_2)Y)]]$$
$$= P_1 R_1 X + P_1 P_2 R_2 Y + P_1 P_2 F_{k-2}((1-R_1)X, (1-R_2)Y) ,$$
$$(BA+):\ F_k(X,Y) = P_2 [R_2 Y + F_{k-1} (X,(1-R_2)Y)] =$$
$$= P_2 [R_2 Y + P_1 [R_1 X + F_{k-2}((1-R_1)X, (1-R_2)Y)]] =$$
$$= P_2 R_2 Y + P_2 P_1 R_1 X + P_2 P_1 F_{k-2}((1-R_1)X, (1-R_2)Y) .$$

Сравнивая оба выражения, получаем искомое уравнение граничной линии при k>1:

$$P_1 R_1 X + P_1 P_2 R_2 Y = P_2 R_2 Y + P_2 P_1 R_1 X$$

или в более изящной форме:

$$\frac{P_1 R_1}{1-P_1} X = \frac{P_2 R_2}{1-P_2} Y \tag{4}$$

Отсюда возникает простой путь решения. На первом шаге проверяем исходное состояние (точку) на положение относительно граничной линии (4) и запоминаем соответствующий выбор. Отыскиваем координаты очередной точки в соответствии со сделанным выбором и проводим аналогичные действия. На последнем шаге (в оставшемся одношаговом процессе) суждение о выборе определяется линией (3).

Пример.

Пусть $N = 5$, $X = 10$, $Y = 20$, $P_1 = 1/3$, $P_2 = 1/2$, $R_1 = 3/4$, $R_2 = 2/3$.

Уравнение (4) граничной линии при $k > 1$ имеет вид

$$Y = \frac{9}{16} X .$$

Точка (10, 20) лежит выше граничной линии, что отвечает выбору B. Этот выбор порождает точку (10, (1–2/3) · 20), т. е. (10, 20/3), принадлежащую области B. Этот выбор, в свою очередь, дает точку (10, 20/9)$\in A$. Выбор A порождает точку (10/4, 20/9) $\in B$.

В итоге проделанных четырех выборов к последнему шагу возникает точка (10/4, 20/27), для которой

$$F_1\left(\frac{10}{4},\frac{20}{27}\right)=max\left[\frac{1}{3}\cdot\frac{3}{4}\cdot\frac{10}{4},\ \ \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{20}{27}\right]=\frac{5}{8}\ ,$$

что соответствует выбору A.

Таким образом, оптимальная последовательность выборов имеет вид $\{B\ B\ A\ B\ A\}$. Ожидаемый объем добычи равен $F_5(10,20)$ и его можно найти обратным ходом по ранее найденным состояниям:

A : $F_1(10/4, 20/27) = 0.625$,

B : $F_2(10/4, 20/9) = 1/2 \cdot [\ 2/3 \cdot 20/9 + F_1(10/4,20/27)] = 1.05$,

A : $F_3(10, 20/9) = 1/3 \cdot [3/4 \cdot 10 + F_2\ (10/4,20/9)] = 2.85$,

B : $F_4\ (10, 20/3) = ½ \cdot [\ 2/3 \cdot 20/3 + F_3\ (10,20/9)\ ] = 3.65$,

B : $F_5(10, 20) = 1/2 \cdot [\ 2/3 \cdot 20 + F_4\ (10,20/3)\ ] = 8.49$.

В случае процесса бесконечной длительности соответствующее функциональное уравнение имеет вид:

$$F(X,Y) = max \begin{bmatrix} A : P_1\ \{\ R_1X + F((1-R_1)X,Y)\ \} \\ B : P_2\ \{\ R_2Y + F(X,(1-R_2)Y)\ \} \end{bmatrix} \tag{5}$$

где $F(X,Y) < X+Y$ – максимум суммарного ожидаемого объема. Аналогичными рассуждениями можно показать, что области A и B первого выбора и здесь разделяются линией (4).

К сожалению, малейшее усложнение задачи многократно увеличивает трудоемкость решения. Так, если допустить, что машина выходит из строя не навсегда, а допускает ремонт к следующему сезону, то (1) преобразуется к виду:

$$F_k\ (X,Y) = max \begin{bmatrix} A : P_1\ \{\ R_1X + F_{k-1}((1-R_1)X,Y)\ \} + (1-P_1)F_{k-1}(X,Y) \\ B : P_2\ \{\ R_2Y + F_{k-1}(X,(1-R_2)Y)\ \} + (1-P_2)F_{k-1}(X,Y) \end{bmatrix} \tag{6}$$

Здесь при сохранении той же структуры оптимального первого выбора поиск граничной линии нетривиален.

Если обратиться к аналогичной *трихотомической* задаче

$$F(X,Y) = max \begin{cases} A : P_1\ [R_1X + F((1-R_1)X,Y)\ ] \\ B : P_2\ [R_2Y + F(X,(1-R_2)Y\ ] \\ C : P_3\ [R_3X + R_4Y + F((1-R_3)X,(1-R_4)Y)] \end{cases} \tag{7}$$

то можно обнаружить, что при определенных соотношениях между коэффициентами взаимное положение областей первого выбора может быть различным; например, как показано на рис.3 [19].

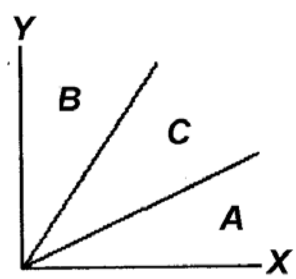


Рис. 3а

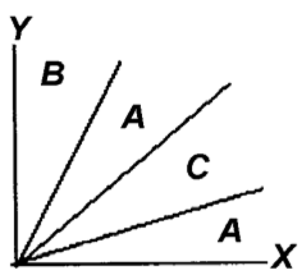


Рис. 3б

9.4. Марковские процессы принятия решений

Пусть некоторая система (техническая, биологическая, информационная) в любой фиксированный момент t может находиться в одном из n состояний и перейти из этого состояния в любое другое. Если вероятность $P_t(i,j)$ перехода в момент t из i-го состояния в j-е не зависит от предыстории системы, такая система называется *марковской*[13].

Если управление некоторым процессом идет без учета накапливаемого опыта, то система управления

Обозначив через $X_t(i)$ ожидаемую вероятность того, что в момент t система находится в i-м состоянии, находим ожидаемую вероятность нахождения системы в любом состоянии в любой последующий момент:

$$X_{t+1}(j) = \sum_{i=1}^{n} P_t(i,j)\, X_t(i), \qquad j = 1 \,..\, n.$$

Так, если вероятности перехода не зависят от t, определяются матрицей

$$P = \begin{Vmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{Vmatrix}$$

и в начальный момент система находится в состоянии 1, т. е. $X_0(1)=1$, $X_0(2)=0$, то:

$X_1(1) = 1/2\ X_0(1) + 1/3\ X_0(2) = 0.5, \quad X_1(2) = 0.\mathbf{5}$,
$X_2(1) = 1/2\ X_1(1) + 1/3\ X_1(2) = 0.\mathbf{4}167, \quad X_2(2) = 0.\mathbf{5}833$,
$X_3(1) = 1/2\ X_2(1) + 1/3\ X_2(2) = 0.\mathbf{40}28, \quad X_3(2) = 0.\mathbf{59}72$,
$X_4(1) = 1/2\ X_3(1) + 1/3\ X_3(2) = 0.\mathbf{40}05, \quad X_4(2) = 0.\mathbf{59}95$ и т. д.

Заметим, что при вероятностях перехода, не зависящих от времени, система обладает свойством *стационарности*, т.е. функция $X_t(j)$ при $t\to\infty$ асимптотически сходится к функции $X(j)$, удовлетворяющей уравнениям:

$$X(j) = \sum_{i=1}^{n} P(i,j)\, X(i), \qquad j = 1 \,..\, n.$$

Для нашего примера

$$X(1) = 1/2\, X(1) + 1/3\, X(2), \quad X(2) = 1/2\, X(1) + 2/3\, X(2),$$

откуда с учетом $X(1) + X(2) = 1$ получаем $X(1) = 0.4$, $X(2) = 0.6$.

Предположим, что в каждый момент времени выбор вероятностей перехода зависит от некоторой политики (выбора) q и переход сопровождается получением некоторого благоприятного эффекта $R_{ij}(q)$.

---

[13] По имени выдающегося русского математика Андрея Андреевича Маркова (1856–1922)

Обозначим через $F_k(i)$ ожидаемый эффект функционирования системы, находившейся в начальный момент в i-м состоянии, за k периодов при использовании оптимальной политики.

Руководствуясь принципом оптимальности, требующим независимо от начального состояния i и от начального выбора q далее действовать оптимально, т. е. гарантировать максимум ожидаемого эффекта в последующем процессе, приходим к рекуррентным соотношениям вида:

$$F_k(i) = \max_q \sum_{j=1}^{n} P_{ij}(q)[R_{ij}(q) + F_{k-1}(j)], \quad i = 1 \,..\, n, \; k \ge 1; \tag{1}$$

$$F_0(i) = 0, \quad i = 1 \,..\, n.$$

Для процессов большой длительности использование приведенных рекуррентных соотношений требует существенных затрат времени даже при машинной реализации процесса вычислений. Если учесть, что при независимости значений вероятностей и эффектов от времени процесс обладает свойством стационарности, то в предположении *регулярности* (возможности прямого или опосредствованного перехода из любого состояния в любое) полагаем для больших k

$$F_k(i) = F_i + k\,G, \tag{2}$$

где G – средний эффект за период и F_i – составляющая суммарного эффекта, определяемая начальным состоянием. Подставляя (2) в (1) и учитывая $\sum_{j=1}^{n} P_{ij}(q) = 1$, получаем систему функциональных уравнений:

$$F_i + G = \max_q \sum_{j=1}^{n} P_{ij}(q)[R_{ij}(q) + F_j], \quad i = 1 \,..\, n, \tag{3}$$

которую можно решать приближением в поведениях. Приведенную систему можно получить, если записать уравнение для бесконечношагового процесса с учетом дисконтирования, положить величину дисконтированного эффекта равной $F_i + G/(1-\alpha)$ и принять $\alpha = 1$.

Для иллюстрации марковского процесса принятия решений рассмотрим "задачу о такси" [21].

Таксист обслуживает окрестности трех городов (три возможных состояния) и может руководствоваться одним из трех выборов: ездить по городу в поисках случайного пассажира, ждать вызова по радио или поехать на стоянку и стать там в очередь.

Для каждого города (i) и каждого выбора (q) известны вероятности поездки в тот или иной город и соответствующие доходы, сведенные в таблице:

Город	Выбор	Вероятности перехода			Значения дохода		
i	q	j=1	j=2	j=3	j=1	j=2	j=3
1	1	1/2	1/4	1/4	10	4	8
	2	1/16	3/4	3/16	8	2	4
	3	1/4	1/8	5/8	4	6	4
2	1	1/2	0	1/2	14	0	18
	2	1/16	7/8	1/16	8	16	8
3	1	1/4	1/2	1/4	10	2	8
	2	1/8	3/4	1/8	6	4	2
	3	3/4	1/16	3/16	4	0	8

Возьмем за начальное поведение $q_0 = (1, 1, 1)$, т.е. во всех городах придерживаться первого выбора. Для выбранного поведения строим систему n уравнений с n+1 неизвестными

$$F_i + G = \sum_{j=1}^{n} P_{i\,j}(q_0)[\,R_{i\,j}(q_0) + F_j\,] \quad , \quad i = 1\,..\,n,$$

разрешимую с точностью до константы. Для нашего примера :

$$F_1 + G = 1/2\,[\,10 + F_1\,] + 1/4\,[\,4 + F_2\,] + 1/4\,[\,8 + F_3]$$
$$F_2 + G = 1/2\,[\,14 + F_1] + \qquad + 1/2\,[18 + F_3]$$
$$F_3 + G = 1/4\,[\,10 + F_1\,] + 1/2\,[\,2 + F_2\,] + 1/4\,[\,8 + F_3\,]$$

Полагая, например, $F_3 = 0$, получаем F_1 =0.95, F_2 = 7.16 и G = 9.32, т. е. выбранная политика дает средний доход за одну поездку, равный 9.32.

Вычисляем

$$T_i(q) = \sum_{j=1}^{n} P_{i\,j}(q)[\,R_{i\,j}(q) + F_j\,]$$

при всех i и q и найденных значениях F_i:

$$T_1(1) = 1/2\ [10+0.95\,] + 1/4\ [\,4+7.16\,] + 1/4\ [\,8+0\,]$$
$$T_1(2) = 1/16\,[\,8+0.95\,] + 3/4\ [\,2+7.16\,] + 3/16\,[\,8+0\,]$$
$$T_1(3) = 1/4\ [\,4+0.95\,] + 1/8\ [\,6+7.16\,] + 5/8\ [\,4+0\,]$$
$$T_2(1) = 1/2\ [14+0.95\,] + \qquad + 1/2\ [18+0\,]$$
$$T_2(2) = 1/16\,[\,8+0.95\,] + 7/8\ [16+7.16\,] + 1/16\,[\,8+0\,]$$
$$T_3(1) = 1/4\ [10+0.95\,] + 1/2\ [\,2+7.16\,] + 1/4\ [\,8+0\,]$$
$$T_3(2) = 1/8\ [\,6+0.95\,] + 3/4\ [\,4+7.16\,] + 1/8\ [\,2+0\,]$$
$$T_3(3) = 3/4\ [\,4+0.95\,] + 1/16\,[\,0+7.16\,] + 3/16\,[\,8+0\,].$$

Выбирая максимальное из значений $T_i(q)$ по q, получаем улучшенное поведение q = (1, 2, 2). Строим и решаем систему уравнений

$$F_1 + G = 1/2\ [\,10 + F_1] + 1/4\,[\,4 + F_2\,] + 1/4\ [\,8 + F_3\,]$$
$$F_2 + G = 1/16\,[\,8 + F_1] + 7/8\,[16 + F_2\,] + 1/16\,[\,8 + F_3]$$
$$F_3 + G = 1/8\ [\,6 + F_1] + 3/4\,[\,4 + F_2\,] + 1/8\ [\,2 + F_3] \quad ,$$

получая F_3 = 0, F_2 = –3.88, F_1 = 12.85, G = 13.15.

Попытка дальнейшего улучшения дает политику q =(2, 2, 2), для которой $F_3 = 0$, $F_2 = -1.18$, $F_1 = 12.86$, $G = 13.34$. Очередная попытка улучшения приводит к той же политике, откуда напрашивается вывод о том, что оптимальная политика состоит в использовании второго выбора во всех городах со средним ожидаемым доходом за одну поездку, равным 13.34. Значения F_k сами по себе не имеют значения, характеризуют относительный вклад начального состояния в общий ожидаемый эффект: в нашем случае при $F_3 = 0$ факт $F_2 < 0$, $F_1 > 0$ свидетельствует о предпочтительности исходить из города 1 и нежелательности в начальный момент оказаться в городе 2.

Существуют обобщения рассмотренного выше метода Ховарда на случай нерегулярности процесса, когда сеть возможных переходов между состояниями распадается на несколько подсетей, переходы между которыми невозможны [21] .

9.5. Задачи и упражнения

Рассмотрим ряд типовых задач, обнаруженных автором в различных источниках, связанных с марковскими процессами принятия решений.

1. Задача о рекламе

Фирма рекламирует свою продукцию с помощью радио, телевидения и газет. Недельные затраты на рекламу оцениваются соответственно в 200, 900 и 300 денежных единиц. Фирма оценивает недельный сбыт тремя состояниями: удовлетворительным (3), хорошим (2) и отличным (1). Известны недельные доходы при разных объемах сбыта и способах рекламы

способ	радио			телевидение			газеты		
состояние	1	2	3	1	2	3	1	2	3
1	400	520	600	1000	1300	1600	400	530	710
2	300	400	700	800	1000	1700	350	450	800
3	200	250	500	600	700	1100	250	400	650

и переходные вероятности между состояниями при различных видах рекламы

способ	радио			телевидение			газеты		
состояние	1	2	3	1	2	3	1	2	3
1	0.4	0.5	0.1	0.7	0.2	0.1	0.2	0.5	0.3
2	0.1	0.7	0.2	0.3	0.6	0.1	0	0.7	0.3
3	0.1	0.3	0.6	0.1	0.2	0.7	0	0.2	0.8

Убедитесь, что поиск способа рекламы, максимизирующего ожидаемый суммарный доход за определенный срок или средний недельный доход в процессе большой длительности, полностью аналогичен 9.4 .

2. Задача ремонта оборудования

Станок находится в одном из трех состояний: хорошем (1), удовлетворительном (2) или плохом (3) и дает доход от выпуска продукции, равный

250, 200 и 50 денежных единиц. Есть возможность использовать обычный или капитальный ремонты или замену на новый станок с затратами:

состояние	обычный ремонт	капитальный ремонт	замена на новый
1	10	15	30
2	50	60	100
3	150	180	200

Известны вероятности перехода между состояниями при разных ремонтах

состояние	обычный ремонт			капитальный ремонт		
	1	2	3	1	2	3
1	0.8	0.2	0	0.9	0.1	0
2	0.1	0.6	0.3	0.5	0.4	0.1
3	0	0.1	0.9	0	0.7	0.3

Вероятность того, что новый станок будет находиться в соответствующем состоянии, равна 0.8, 0.15 и 0.05.

3 Оптимальное распределение печатной продукции

Газеты доставляются для продажи в N киосках. Предполагая, что известна функция плотности распределения количества продаваемых в каждом киоске газет и что непроданные газеты возвращаются с соответствующей уценкой, определить оптимальное распределение газет по киоскам.

Покажите, что решение этой задачи аналогично рассмотренной выше задаче управления запасами при спросе, являющемся непрерывной случайной величиной.

Как можно найти оптимальный тираж газет?

4. Простейшие задачи об очередях

Перед человеком стоит очередь из N претендентов на обслуживание. Известна полезность R от выстаивания очереди и вероятность P того, что в единицу времени будет обслужен один человек. За каждую единицу времени ожидания человек терпит убыток C. Стоит ли занимать очередь?

Убедитесь, что $F_k = max\,[\,-C + P\,F_{k-1} + (1-P)\,F_k,\, 0\,]$, k 1..N, $F_0 = R$, где F_k – ожидаемый доход, получаемый от выстаивания очереди из k человек при оптимальной политике.

Покажите, что приведенные соотношения можно привести к виду

$$F_k = max\,[\,(\,P\,F_{k-1} - C\,)\,/\,(1-P)\,,\, 0\,]\,.$$

Выясните структуру оптимальной политики.

Убедитесь в правомерности приведенных ниже соотношений при оценке резона становиться в очередь из N жаждущих приобрести изделия от Versaci, имеющиеся в количестве L штук, в предположении известных вероятностей приобретения одного или двух изделий или отказа от покупки.

$$F_k\,(L>0) = max\,[P_1\,F_{k-1}(L-1) + P_2\,F_{k-1}\,(\,L-2) + (1 - P_1 - P_2)\,F_k(\,L),\, 0\,]\,,\ k=1..N\,;$$

$$F_k\,(L\le 0) = 0,\quad k=0..N\,;\quad F_0(L>0)=1$$

10. ЭЛЕМЕНТЫ ТЕОРИИ ИГР И СТАТИСТИЧЕСКИХ РЕШЕНИЙ

За окном XXI век. Вода превращается в вино только в выступлениях фокусников; желание «да будет свет» осуществляется вполне обыденно; человек, отрицающий шарообразность Земли через 500 лет после Колумба или печатающий сообщение в научном журнале о березовой ветке, выросшей на сосне, ассоциируется с клиентом психбольницы или проходимцем.

По мере накопления знаний в астрономии и физике, отдельных областях химии и биологии выявлены многие законы детерминированного характера, нарушение которых является событием невозможным или чрезвычайным. Здесь математика, вооруженная компьютером, избавила человека от необходимости проводить многочисленные экспериментальные взрывы, опытным путем подбирать критическую массу ядерного заряда или параметры оболочки реактора, гадать о дате очередного «парада планет» и т. п.

Познав физическую природу явления, человек строит его математическую модель и получает возможность если не управлять, то хотя бы давать четкий прогноз. Специалист, которому доверено проектирование новой гидростанции, управление распределением электроэнергии или выбор угла резания, опирается на соответствующие математические модели, и надежность его решений вне сомнений.

Однако, о природе многих явлений природы и человеческого мышления наши представления ничтожны. Какой-либо детерминизм представлений – база для принимаемых решений – исчезает и подменяется неопределенностью. «Аннушка уже разлила подсолнечное масло… и Берлиоза выкинуло на рельсы» (кто мог предугадать эту причинно-следственную связь, кроме Воланда). Гроссмейстер проигрывает партию из-за «зевка», о причине которого приходится только гадать. В каждой сотне операторов «отлаженной» программы присутствует хотя бы одна ошибка. Запланированный отпуск на Мальдивских островах безнадежно испорчен из-за невесть откуда взявшегося цунами. Разражается финансовый кризис из-за того, что какие-то банки были непредусмотрительны в своей кредитной политике.

Если некоторые явления в природе еще можно познать хотя бы на уровне вероятностных оценок, то психология человека, мотивация его поступков и суждений столь загадочны, что возникают сомнения в человеческом Разуме. Всякий экономист-академик (и не только) имеет свое, «единственно верное» мнение о действиях по выходу из кризиса или об использовании профицита баланса. Мудрые философы никогда не придут к определенному выводу о первичности курицы или яйца. О каких разумных управленческих решениях в таких условиях можно говорить?!

«Искусство выработки наилучших (в смысле тех или иных критериев) суждений старо, как род человеческий; именно это искусство является сущностью любой сферы человеческой деятельности …» [29].

При выработке суждений относительно последующих действий, приходится учитывать подчас случайный характер человеческих поступков, а также элемент случайности, присущий окружающей среде. Поэтому всякое научное предвидение тесно связано с использованием аппарата теории вероятностей и математической статистики.

Примечательно, что принимаемые рекомендации здесь не носят категорического характера, а звучат как «всё будет хорошо с вероятностью 0.95», «велика вероятность того, что ваши ожидаемые издержки не превысят …», «вы достигнете максимума ожидаемого успеха, если два дня в неделю на вас будет коричневый костюм» (ответ на вопрос «в какие дни?» остается за кадром) .

Наука о выработке таких суждений – математическая теория принятия решений – оформилась как самостоятельная математическая дисциплина «Теория игр и статистических решений» в 20–40 -е годы XX столетия в работе Джона фон-Неймана и О. Моргенштерна [28]. Большой вклад в ее развитие внесли А. Вальд (создатель теории последовательного анализа), блистательная группа американских математиков 50-х годов, многие из которых удостоены Нобелевских премий по экономике, и др. Тем не менее приходится констатировать, что пока в области приложений эта наука не может конкурировать, например, с математической физикой или небесной механикой.

10.1. Основные понятия теории игр

Теория игр занимается изучением т. н. конфликтных ситуаций, где сталкиваются интересы индивидов, партий, государств и т. п.

Как утверждал Г. Лейбниц, «...и игры заслуживают изучения; и если какой-нибудь проницательный математик посвятит себя их изучению, то получит много важных результатов, ибо нигде человек не показывает столько изобретательности, как в игре».

Разумеется, никакая математическая теория не может однозначно предсказать результат сегодняшней партии в преферанс или повышение акций ВТБ, но иногда возможен математический прогноз наилучшего поведения в подобных ситуациях.

Во избежание терминологической путаницы следует различать понятие ***игры*** как совокупности правил и индивидуальные ***партии*** (реализации) игры: двусмысленная фраза «я играю в шахматы» может означать знакомство с правилами этой популярной игры или факт пребывания за шахматной доской.

Столь же важно различать понятия ***выбор*** и ***ход***: игра состоит из последовательности ходов, тогда как партия – из последовательности выборов (в повседневности мы часто путаем эти понятия – когда Остап Бендер «сделал ход е2–е4», фактически он сделал соответствующий выбор из 20 допустимых. Самыми примитивными являются одноходовые игры, где всякая партия состоит из одного хода (например, дети показывают друг другу пальцы: если общее число показанных пальцев четное, выигрывает первый, в противном случае – второй).

Следует осмыслить и понятие ***стратегии*** – общих принципов, которым подчинены выборы; каждый игрок выбирает свою стратегию, если он в состоянии оценивать эффекты, достижимые в результате такого выбора. Другими словами, стратегия – это система правил, однозначно определяющая выбор игрока в зависимости от сложившейся ситуации. В одноходовых играх стратегия достаточно проста («всегда иди домой с цветами», « в половине случаев бери с собой зонтик», «если …, вступай в коалицию с соседом слева» и т. п.), в многоходовых (динамических) играх, где очередной выбор зависит от результата предыдущих, стратегия может оказаться много сложнее.

Каждая фиксированная стратегия игрока, где любой ситуации сопоставлен однозначно конкретный выбор, называется **чистой**. В реальности чаще используются т. н. **смешанные стратегии**, где какие-то выборы используются с некоторыми частотами.

Следует различать ***личные ходы***, где выбор производится конкретным игроком в зависимости только от его личной оценки ситуации, и ***случайные ходы***, где выборы делаются каким-то устройством, отличным от homo sapiens, случайным образом в соответствии с какими-то вероятностями.

Не претендуя на полноту, остановимся на классификации игр.

Как уже видно из предшествующих замечаний, игры могут быть **одно**- или **многоходовыми**.

Интересы участников игры (игроков) могут оказаться несовпадающими и даже противоположными. В последнем случае игра называется ***антагонистической***.

В игре могут участвовать два или более игроков. Случай игры с одним участником (пасьянс, управление физическим объектом и т.д.) в сущности является игрой двух лиц, где вторым участником выступает природа (судьба, рок, провидение).

Антагонистическую игру, где выигрыш одного коллектива равен проигрышу другого, называют ***игрой с нулевой суммой***.

Игроки могут в игре выступать каждый за себя или объединяться в группы. В последнем случае игра называется ***коалиционной***. Подобные игры представляют исключительный интерес для приложений, но результаты исследований в этой сфере достаточно скромны.

Игры, в которых игроки осведомлены о состоянии своем и партнеров, а также о прошлом поведении участников игры, относятся к категории игр ***с полной информацией*** (типичные примеры – шахматы, крестики-нолики и т. п.). Большинство же игр протекает в условиях неполной информации, где сведения о состоянии партнеров исчерпываются лишь вероятностными характеристиками (домино, карточные игры, игры против «природы»).

Если множество выборов определяется неким интервалом значений (угол атаки от 0 до 60°, случайное значение из (0,1) и т. п.), говорят о **непрерывных играх**. В противном случае можно говорить об игре на дискретном множестве выборов.

Простейшими среди игр являются одноходовые ***матричные игры двух лиц с нулевой суммой***.

Пусть в такой игре игрок 1 имеет m выборов и игрок 2 – n выборов. Если игрок 1 делает свой i-й выбор, а игрок 2 – свой j-й выбор, то выигрыш игрока 1 (проигрыш игрока 2) равен R_{ij}. Такая игра называется *матричной* и матрица $R = [\ R_{ij}\ /\ \ i=1..m,\ j=1..n\ ]$ называется *матрицей выигрышей* (платежной матрицей).

Выигрыш игрока 1 (проигрыш игрока 2) при оптимальной политике обоих игроков принято называть **ценой игры**.

При ведении игры игрок должен ориентироваться на оптимальную политику партнера и наказывать его за отступления от таковой.

Проведем рассуждения за игрока 1. Если Я намерен пользоваться i-м выбором, мой догадливый партнер для минимизации моего выигрыша сделает тот из своих выборов, который даст мне $min\ R_{ij}$. Соответственно, Я должен использовать тот выбор, который гарантирует мне выигрыш, не меньший

$$V_1 = \max_{i=1..m}\ \min_{j=1..n}\ R_{ij}\,.$$

Партнер, рассуждая аналогично, приходит к выводу о гарантированном проигрыше, не превышающем

$$V_2 = \min_{j=1..n} \max_{i=1..m} R_{ij}\,.$$

Если в матрице выигрышей существует элемент $R_{kl} = V_1 = V_2$, то говорят о наличии *оптимальной политики в пространстве чистых стратегий* и оптимальными выборами для игроков соответственно являются выборы k и l. Пару (k, l) называют *седловой* точкой и значение вышеуказанного элемента – ценой игры.

Пример 1. Пусть игра определяется матрицей

$$R = \begin{Vmatrix} 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 3 & 4 & 6 \\ 6 & 6 & 8 & 10 \end{Vmatrix}, \quad V_1 = max \begin{Vmatrix} 2 \\ 4 \\ 3 \\ 6 \end{Vmatrix} = 6, \ V_2 = min\,[\,6, 6, 8, 10\,] = 6.$$

Цена игры = 6, седловые точки – (4, 1) и (4, 2), то есть оптимальный выбор для игрока 1 – четвертый, для игрока 2 равнозначны первый и второй.

При анализе игр часто прибегают к попыткам обнаружить **доминировани**е между строками и столбцами. Так в примере 1 элементы четвертой строки больше элементов других строк: использование выбора 4 выгоднее других выборов при любой политике противника. Противник видит, что в такой ситуации использовать выборы 3 и 4 неразумно.

Использование доминирования позволяет уменьшить размеры изучаемой матрицы исключением «невыгодных» строк и столбцов.

Пример 2. Пусть игра определяется матрицей

$$R = \begin{Vmatrix} 5 & 4 & 3 & 7 & 6 \\ 6 & 2 & 4 & 3 & 5 \\ 2 & 7 & 3 & 5 & 4 \\ 7 & 3 & 4 & 4 & 4 \end{Vmatrix}, \quad V_1 = max \begin{Vmatrix} 3 \\ 2 \\ 2 \\ 3 \end{Vmatrix} = 3\,, \quad V_2 = min\,[\,7, 7\,, 4, 7, 6\,] = 4.$$

Обратите внимание, что значение $V_1 = 3$ определяет **гарантированный выигрыш** игрока 1 (он выиграет, по крайней мере, эту величину, если будет пользоваться первым или четвертым выбором). Значение $V_2 = 4$ определяет **гарантированный проигрыш** игрока 2 (можно гарантировать, что его проигрыш не превысит 4, если он будет пользоваться третьим из своих выборов).

Здесь равенство $V_1 = V_2$ не выполняется; оптимальной чистой стратегии для игроков нет и цена игры $V_1 \le V \le V_2$.

При отсутствии седловой точки среди чистых стратегий приходится искать таковую среди смешанных.

Если игрок 1 прибегает к своему выбору i с вероятностью P_i , а игрок 2 – к своему j-му выбору с вероятностью Q_j , то ожидаемый выигрыш игрока 1 (проигрыш игрока 2) равен $\sum_{i=1}^{m} R_{i\,j}\, P_i\, Q_j = \mathrm{P}^{\mathrm{T}} \mathrm{R}\, \mathrm{Q}$.

Основная теорема теории игр (*теорема Джона фон Неймана*) утверждает, что любая матричная игра с нулевой суммой всегда имеет седловую точку, т.е. существуют векторы P и Q такие, что

$$\max_{P} \min_{Q} P^T RQ = \min_{Q} \max_{P} P^T RQ = V,$$

(V - цена игры, ожидаемый выигрыш-проигрыш при оптимальной политике партнеров).

10.2. Матричные игры и линейное программирование

Очевидно, что если игрок 1 отступит от оптимальной политики, а игрок 2 будет действовать оптимально, то выигрыш игрока 1 будет меньше цены игры, и если игрок 2 отступит от оптимальной политики при сохранении оптимального поведения игроком 1, то его проигрыш превысит цену игры:

$$P^T R\, Q_{opt} \le V = P_{opt}^T R\, Q_{opt} \le P_{opt}^T R\, Q$$

Рассуждения игрока 1: мне хотелось бы сделать цену игры как можно большей, т.е. увеличить мой гарантированный выигрыш, и я должен подобрать систему значений P_i так, чтобы при любом выборе игрока 2 мой ожидаемый выигрыш был больше цены игры.

Рассуждения игрока 2: мне хочется уменьшить мой гарантированный проигрыш, т. е. цену игры, и мне надо подобрать значения Q_j так, чтобы при любом выборе игрока 1 мой проигрыш был меньше цены игры.

Отсюда возникают две задачи:

максимизировать V	минимизировать V
при условиях	при условиях
$\sum_{i=1}^{m} R_{i\,j} P_i \ge V$, $j=1\,..\,n$	$\sum_{j=1}^{n} R_{i\,j} Q_j \le V$, $i=1\,..\,m$
$\sum_{i=1}^{m} P_i = 1$	$\sum_{j=1}^{n} Q_j = 1$
$P_i \ge 0,\quad i = 1\,..\,m$	$Q_j \ge 0,\quad j = 1\,..\,n$

Легко видеть, что эти задачи образуют пару двойственных задач линейного программирования и решение матричной игры сводится к решению пары двойственных линейных программ.

Обратим внимание на то, что при увеличении элементов матрицы R на любую константу C цена игры увеличится на C и это изменение не окажет влияния на искомые вероятности выборов. Таким образом можно добиться, например, положительности элементов матрицы и, следовательно, цены игры.

В предположении $V > 0$ проведем замену переменных

$$X_i = P_i \ / \ V \quad , \qquad Y_j = Q_j \ / \ V .$$

Из равенств $\sum_{i=1}^{m} P_i = 1$, $\sum_{j=1}^{n} Q_j = 1$ можно получить, что

$$V = 1 / \sum X_i = 1 / \sum Y_j \quad .$$

Соответственно, поставленные задачи можно преобразовать к задачам с меньшим числом переменных:

минимизировать	максимизировать
$\sum_{i=1}^{m} X_i$	$\sum_{j=1}^{n} Y_j$
при условиях	при условиях
$\sum_{i=1}^{m} R_{i\,j} X_i \geq 1$, $j=1 \, . \, . \, n$	$\sum_{j=1}^{n} R_{i\,j} Y_j \leq 1$, $i=1 \, . \, . \, m$
$X_i \geq 0$, $i=1 \, . \, . \, m$	$Y_j \geq 0$, $j=1 \, . \, . \, n$

Например, для игры с матрицей

$$\begin{Vmatrix} 1 & 2 & 3 \\ 4 & 0 & 1 \\ 2 & 3 & 0 \end{Vmatrix}$$

возникают задачи:

максимизировать	минимизировать
$Y_1 + Y_2 + Y_3$	$X_1 + X_2 + X_3$
при	при
$Y_1 + 2\,Y_2 + 3\,Y_3 \leq 1$	$X_1 + 4\,X_2 + 2\,X_3 \geq 1$
$4\,Y_1 + Y_3 \leq 1$	$2\,X_1 + 3\,X_3 \geq 1$
$2\,Y_1 + 3\,Y_2 \leq 1$	$3\,X_1 + X_2 \geq 1$
$Y_1, \ Y_2, \ Y_3 \geq 0$	$X_1, \ X_2, \ X_3 \geq 0$

Решение этих задач симплексным методом дает оптимальные значения $X = \{ 11/37, 4/37, 5/37 \}$, $Y = \{ 8/37, 7/37, 5/37 \}$ и экстремумы целевых функций, равные 20/37.

Отсюда $V = 37/20$, $P = \{11/20, 4/20, 5/20\}$, $Q = \{8/20, 7/20, 5/20\}$.

Другими словами, при многократной реализации игры игрокам рекомендуется использовать свои выборы с соответствующими вероятностями (отступление от этого требования чревато неприятностями).

А как поступить в реальности с этими вероятностями?

Если бы вероятности оказались равными 0.25, 0.5 и 0.25, то достаточно бросить монету: выпадет «решка» – делай выбор 2, в противном случае брось монету еще раз, при выпадении «решки» делай выбор 1 и при орле – выбор 3 (или наоборот). В нашем случае можно включить компьютер, войти в любую знакомую программную среду и обратится к *датчику случайных чисел равномерного распределения в* (0,1)[14]. Если полученное число меньше 0.55 – делай выбор 1, при числе из интервала от 0.55 до 0.75 – выбор 2 и числе большем 0.75 – выбор 3.

10.3. Итеративный метод решения матричных игр

Как мы показали выше, игры могут решаться методами линейного программирования. Однако при больших размерах платежной матрицы даже в наши дни решение довольно трудоемкое и чревато большой вычислительной погрешностью – традиционная беда всех т. н. «точных методов».

Здесь мы рассмотрим *итеративный метод Брауна – Робинсон* (разумеется, мы не помышляем о решении вручную), допускающий простую программную реализацию.

Выполняем многократную реализацию игры на основе знания предыстории с последовательным совершенствованием стратегий.

Для примера возьмем задачу, которую мы только что решили.

Пусть игрок 1 случайно сделал выбор 1 с ожидаемыми выигрышами 1, 2, 3. Противник, стремясь минимизировать свой проигрыш, прибегнет к выбору 1 с ожиданием проигрыша 1, 4, 2.

Игрок 1 в стремлении максимизировать свой выигрыш прибегнет к выбору 2, что даст ему надежду на суммарный выигрыш (1+4, 2+0, 3+1). Но тогда

Шаг	Выбор i	Суммарный выигрыш			Выбор j	Суммарный проигрыш		
1	1	1	2	3	1	1	4	2
2	2	5	2	4	2	3	4	5
3	3	7	5	4	3	6	5	5
4	1	8	7	7	2	8	5	8
5	1	9	9	10	1	9	9	10
6	3	11	12	10	3	12	10	10
7	1	12	14	13	1	13	14	12
8	2	16	14	13	3	16	15	12
9	1	17	16	16	2	18	15	15
10	1	18	18	19	1	19	19	17

[14] Вообще-то эти числа, получаемые по определенному правилу, но подчиняющиеся определенным статистическим критериям, называют *псевдослучайными*.

его противник найдет среди этих значений меньшее и прибегнет к выбору 2 с ожидаемым суммарным проигрышем (1+2, 4+0, 2+3) и т. д.

Этот процесс реализуется достаточно большое число раз (см. в 5.1 характеристику методов Монте-Карло) с последующим поиском частоты использования выборов и усреднением выигрышей-проигрышей.

В результате 10 выборов для игрока 1 частоты составили 0.6, 0.2, 0.2; для игрока 2 – 0.4 , 0.3 и 0.3; оценка цены игры в диапазоне от 1.7 до 1.9 .

10.4. Многошаговые игры. Игры на выживание

Предыдущее рассмотрение игр проводилось в предположении, что игра является одноходовой и реализация игры может осуществляться большое число раз.

Однако в реальной многоходовой игре с ограниченными ресурсами е политика игроков (последовательность выборов) зависит от результата предыдущих выборов и от длительности игры.

Соответственно для матричной игры

$$F_k(A,B) = \max_P \min_Q \sum_{i=1}^{m} \sum_{j=1}^{n} P_i Q_j F_{k-1}(A+R_{ij}, B-R_{ij}) =$$

$$= \min_Q \max_P \sum_{i=1}^{m} \sum_{j=1}^{n} P_i Q_j F_{k-1}(A+R_{ij}, B-R_{ij}),$$

где $F_k(A, B)$ – ожидаемый выигрыш игрока 1 в k последовательных реализациях при начальных ресурсах A и B и использовании оптимальной политики.

Пусть общий начальный ресурс игроков $A + B = C$ и игра продолжается до разорения одного из игроков. Обозначим через $F(A)$ ожидаемую вероятность выживания (шансы не разориться) игрока 1 при его начальном ресурсе A и оптимальной политике обоих игроков.

Тогда

$$F(A) = \max_P \min_Q \sum_{i=1}^{m} \sum_{j=1}^{n} P_i Q_j F(A+R_{ij}) =$$

$$= \min_Q \max_P \sum_{i=1}^{m} \sum_{j=1}^{n} P_i Q_j F(A+R_{ij})$$

$$F(A \le 0) = 0, \qquad F(A \ge C) = 1 .$$

Если игра не обладает чистыми оптимальными стратегиями, то оптимальные значения вероятностей использования выборов соответствуют внутренним точкам множества планов ($0<P<1$, $0<Q<1$) и напрашивается мысль для поиска оптимальных P, Q прибегнуть к аппарату производных.

Пример. Рассмотрим игру на выживание с матрицей $\begin{Vmatrix} 2 & -1 \\ -2 & 1 \end{Vmatrix}$ при полном капитале игроков $C = 4$.

Здесь, в силу целочисленности данных, берем целочисленные значения A от 0 до 4. Если обозначить вероятности соответствующих выборов игроков через P, $1–P$, Q, $1–Q$, то $F(A \le 0) = 0$, $F(A \ge 4) = 1$,

$$F(1)= \max_P \min_Q [PQ\,F(3) + P(1–Q)\,F(0) +$$

$$+(1–P)Q\,F(–1) + (1–P)(1–Q)F(2)]=$$

$$= \max_P \min_Q [\,PQ\,F(3) + (1–P)(1–Q)\,\mathrm{F}(2)],$$

$$F(2)= \max_P \min_Q [PQ\,F(4) + P(1–Q)\,F(1) +$$

$$+(1–P)Q\,F(0) + (1–P)(1–Q)\,F(3)]=$$

$$= \max_P \min_Q [PQ + P(1–Q)\,F(1) + (1–P)(1–Q)\,F(3)],$$

$$F(3)= \max_P \min_Q [PQ\,F(5) + P(1–Q)\,F(2) +$$

$$+ (1–P)Q\,F(1) + (1–P)(1–Q)\,F(4)=$$

$$= \max_P \min_Q [PQ + P(1–Q)\,F(2) + (1–P)Q\,F(1) +(1–P)(1–Q)].$$

Находим частные производные и строим системы уравнений для поиска оптимальных значений $P(A)$, $Q(A)$:

1) $$\frac{\partial}{\partial P}F(1) = Q\,F(3) – (1–Q)\,F(2) = 0\ ,$$

$$\frac{\partial}{\partial Q}F(1) = P\,F(3) – (1–P)\,F(2) = 0;$$

2) $$\frac{\partial}{\partial P}F(2) = Q + (1–Q)\,F(1) – (1–Q)\,F(3) = 0,$$

$$\frac{\partial}{\partial Q}F(2) = P\ –\ P\,F(1) – (1–P)\,F(3) = 0;$$

3) $$\frac{\partial}{\partial P}\ F(3)\ = Q + (1–Q)\,F(2) – Q\,F(1) – (1–Q) = 0,$$

$$\frac{\partial}{\partial Q}F(3)\ = P\ – P\,F(2) + (1–P)\,F(1) – (1–P) = 0.$$

Решение приведенных систем дает

$$P(1) = Q(1) = \frac{F(2)}{F(2)+F(3)};$$

$$P(2) = \frac{F(3)}{1-F(1)+F(3)}\ ,\ Q(2) = \frac{F(3)-F(1)}{1-F(1)+F(3)};$$

$$P(3) = \frac{1-F(1)}{2-F(1)-F(2)}\ ,\ Q(3) = \frac{1-F(2)}{2-F(1)-F(2)}.$$

Подставляя полученные выражения в исходные выражения функций, вновь имеем нелинейную систему относительно $F(1)$, $F(2)$, $F(3)$:

$$F(1) = \frac{F(2)\cdot F(3)}{F(2)+F(3)}\ ,\ F(2) = \frac{F(3)}{1-F(1)+F(3)}\ ,\ F(3) = \frac{1-F(1)\cdot F(2)}{2-F(1)-F(2)}\ .$$

Решая эту систему, имеем оценки

$$F(1) = 0.3\ ,\ F(2) = 0.5\ ,\ F(3) = 0.7$$

и

$$P(1) = 0.41,\quad P(2) = 0.5,\quad P(3) = 0.59,$$
$$Q(1) = 0.41,\quad Q(2) = 0.3,\quad Q(3) = 0.41.$$

10.5. Многошаговые игры. Игры погони

Простейшим примером таких игр может служить задача для двух игроков, расположившихся на прямой на расстоянии d. На каждом шаге игры игроки могут одновременно смещаться влево или вправо при полной информации о позиции друг друга. После очередного шага игрок 2 уплачивает игроку 1 величину $G(S)$, где S – расстояние между ними. С вероятностью $A(d)$ игра может быть продолжена и с вероятностью $1-A(d)$ окончена.

Если обозначить через P_1, P_2, Q_1, Q_2 вероятности смещения игроков в ту или иную сторону, то

$$F(d) = G(d)+A(d)\cdot \max_P \min_Q [\, P_1 Q_1 F(d) + P_1 Q_2 F(d+2) +$$
$$+P_2 Q_1 F(d-2) + P_2 Q_2 F(d)\,] =$$
$$= G(d) + A(d) \min_Q \max_P [P_1 Q_1 F(d)+ \ldots\,]$$

Больший интерес может представить игра погони на плоскости или в пространстве, где устанавливается принципиальная возможность поимки одного игрока другим или отыскивается траектория, минимизирующая время поимки. Эти игры относятся к т. н. непрерывным многошаговым играм, решение которых сводится к дискретным моделям.

Нетрудно заметить, решение многошаговых игр приводит к необходимости решать нелинейные системы уравнений, что по сложности не уступает решению задач нелинейного программирования.

10.6. Статистические решения. Основные понятия

Выбор наилучших решений в условиях полной и неполной информации – одно из основных занятий людей. Принятие управленческих решений в условиях неполной или неточной информации сопряжено с неизбежным риском понести немалые убытки, причинить вред здоровью или вместо Сочи оказаться в «солнечном» Магадане в случае принятия ошибочного решения.

Когда мы знакомились с азбучными истинами теории игр, то предполагали, что участниками игры являются люди, способные принимать разумные решения. Другая ситуация возникает в т. н. **играх против природы**, где человек, разумный по предположению, противостоит непознанному им явлению. Девушка раскладывает пасьянс в надежде выяснить «любит – не любит» и расстраивается, если пасьянс не сходится. Собираясь в туристический поход, юноша укладывает вещи в рюкзак с учетом непредсказуемой погоды, надеясь получить максимум удовольствий, не превращаясь в рекордсмена по переноске тяжестей. Колумб плывет на запад, чтобы достичь Индии, и к неудовольствию королевы возвращается без золота и пряностей. Фермер сеет ячмень в надежде на прибыль, но урожай гибнет на корню; в последующие годы он повторяет эксперимент в надежде, что ему повезет.

Человек обычно сетует на судьбу и не благодарит ее, когда «повезло». Едва ли стоит связывать с природой (судьбой, высшими силами…) априорную злонамеренность или предрасположенность по отношению к человеку, хотя человек подчас делает многое, чтобы вывести окружающую среду из состояния равновесия.

Теория статистических решений может быть истолкована как теория поиска оптимального недетерминированного поведения в условиях неопределенности. Современная концепция статистического решения выдвинута А. Вальдом и считает поведение оптимальным, если оно **минимизирует риск в последовательных экспериментах**, т. е. математическое ожидание убытков статистического эксперимента. В такой постановке любая задача статистических решений может рассматриваться как игра двух лиц, в которой одним из игроков является «природа».

Если быть более точным в терминологии, то следует различать ***ситуацию риска*** и ***ситуацию неопределенности***.

Мы говорим о ***риске***, когда имеется статистическая информация о подобных решениях и существует возможность оценить вероятности, связанные с последствиями принятия решения.

Неопределённость существует тогда, когда нет соответствующей статистики, нет возможности объективно оценить указанные вероятности (иногда здесь говорят о дурной случайности). В таких ситуациях остается прибегнуть к экспертным оценкам или надеяться на собственное озарение. Возникает и проблема выбора критерия оптимальности, поскольку решение, оптимальное для каких-то условий, бывает неприемлемым в других и приходится искать некоторый компромисс.

Математизированная постановка задачи выбора решения в условиях неопределённости (риска) сводится к следующему.

Пусть задан некоторый *вектор* $S = (S_1, S_2, .., S_n)$, описывающий *n состояний внешней среды*, и вектор $X = (X_1, X_2, .., X_m)$, описывающий *m допустимых решений*. Требуется найти вектор $X^* =(0,0,..,0, X_i ,0,..,0)$, который обеспечивает оптимум некоторой ***функции полезности*** $W(X, S)$ по некоторому критерию K.

Функция полезности сопоставляет каждому состоянию S внешней среды и каждому предлагаемому решению X значение т. н. полезности (дохода, прибыли, эффективности …).

Информацию об указанной функции можно представлять матрицей размерности $m \times n$ с элементами $W_{ij} = F(X_i, S_j)$, где F - *решающее правило*.

Формирование решающего правила во многом предопределяет конечный результат расчетов (в случае его ошибочности едва ли принимаемое решение окажется наилучшим) и возможно лишь при достаточно четкой экономической постановке задачи.

Так арендуя помещение с M посадочными местами для увеселительного мероприятия при заранее неизвестном числе N посетителей, предприниматель стоит перед выбором: потерять возможную прибыль, если окажется посетителей больше чем мест в зале, или зря потратить деньги на наем просторного помещения при незначительном числе посетителей. Если стоимость билета равна k и затраты на аренду равны $a(M)$, то нетрудно сообразить, что функция полезности (дохода) равна $k\cdot \min (M,N) - a(M)$ и имеет смысл рассмотреть ее значения для возможных вариантов значений M и N.

Планируя выпуск новой продукции, необходимо **заблаговременно** закупить станки. Система оптовой торговли может поставить не более 50 станков; комплект поставки – 10 станков. Минимальный объем поставок – 20 станков. Производительность одного станка составляет 2 изделия в год, каждое из которых приносит доход 21.9 тыс.руб. Оптовая цена одного станка 4.775 тыс. руб., содержание станка – 3.6 тыс. руб. Затраты на подготовку производства составляют 25.5 тыс. руб. и не зависят от числа станков. Спрос на изделия прогнозируется в диапазоне от 20 до 100 .

Соответственно, вектор решений об объеме поставок $X = (20, 30, 40, 50)$, состояние же спроса можно описать вектором $S = [20, 40, 60, 80, 100]$ (едва ли резонно брать меньший шаг).

Если решающее правило сформулировать как «доход – издержки», то матрица полезности:

$$W(X, S.) = 21.9 \cdot min(2X, S) - 23.6\ X - 25.5\ - 4.775\ X$$

(последнее вычитаемое может быть модифицировано, если закупаемые станки будут эксплуатироваться не один год).

Можно привести множество подобных примеров принятия решения в условиях неопределенности (в реальной жизни неопределенность может быть более чем одномерной и определяться не только спросом на изделия, но и международной обстановкой на маршрутах доставки изделий заказчику).

Если ***в ситуации риска*** имеющиеся статистические данные позволяют оценить вероятность $P(S)$ того или иного состояния внешней среды, то можно найти математическое ожидание функции полезности и сделать выбор X_i, обеспечивающий его максимум:

$$W = \max_{i=1..m} \sum_{j=1}^{n} W_{i\,j} P_j \quad . \tag{1}$$

В ситуации неопределенности многообразие критериев (подходов к выбору наилучшего решения) несколько больше.

<u>Критерий Лапласа.</u> Когда невозможно выяснить вероятности возникновения того или иного состояния внешней среды, по **принципу недостаточного основания** (нет оснований полагать, что то или иное состояние возникает чаще других), им сопоставляют *равные значения* $p_i = 1/n$ и находят *средний эффект* для каждого из рассматриваемых вариантов решения, выбирая тот из них, для которого средний эффект максимален:

$$W = \max_{i=1..m} \frac{1}{n} \sum_{j=1}^{n} W_{i\,j} \quad . \tag{2}$$

<u>Критерий Вальда</u> (критерий наибольшей осторожности, или пессимистический критерий). *Для каждого* из рассматриваемых вариантов решения X_i выбирается *самый худший отклик среды* (наименьшее из W_{ij}) и среди них отыскивается гарантированный максимальный эффект:

$$W = \max_{i=1..m} \min_{j=1..n} W_{ij} \ . \qquad (3)$$

<u>Критерий Гурвица</u>[15]. Ориентация на самый худший исход является своеобразной перестраховкой, однако опрометчиво выбирать и излишне оптимистичную политику. Критерий Гурвица предлагает некоторый компромисс:

$$\mathrm{W} = \max_{i=1..m} [\,\alpha \max_{j=1..n} W_{ij} + (1-\alpha) \min_{j=1..n} W_{ij}\,] \ , \qquad (4)$$

где параметр $0 \le \alpha \le 1$ выступает как ***коэффициент оптимизма***.

К примеру, при $\alpha = 0$ (полный пессимизм) критерий Гурвица превращается в критерий Вальда, при $\alpha = 0{,}5$ шансы на успех и неудачу мы предполагаем равновероятными, при $\alpha = 0{,}8$ мы более радужно расцениваем свои шансы на успех. Выбор же $\alpha = 1$ вызывает определенные сомнения в трезвом подходе к решаемой проблеме. В некоторых сферах человеческой деятельности, например, при оценке сроков выполнения работ, предпочитают выбор $\alpha = 0{,}4$ (выводы делайте сами).

Особое место занимает ***<u>критерий Сэвиджа</u>***. При выборе решения по этому критерию:

1) матрице полезности сопоставляется новая матрица – ***матрица сожалений*** с элементами $D_{ij} = W_{ij} - \max\limits_{i} (W_{ij})$, которые отражают убытки от ошибочного действия, т. е. выгоду, упущенную в результате принятия i-го решения в j-м состоянии;

2) для матрицы D выбирается решение по пессимистическому критерию Вальда, дающее наименьшее значение максимального сожаления

$$W = \max_{i=1..m} \min_{j=1..n} D_{ij} \qquad (5)$$

(минимум упущенной выгоды при принятии данного решения) .

Вполне логично, что ***различные критерии приводят к различным выводам относительно наилучшего решения***. Каждый вывод

---

[15] Леонид (Леон) Гурвиц (1917–2008) – выходец из России, удостоенный в 2007 г. Нобелевской премии за создание основ теории механизмов распределения. Как отмечал Нобелевский комитет, теория, созданная им и развитая затем нобелевскими лауреатами Эриком Маскиным и Роджером Майерсоном, помогла «выявить эффективные торговые механизмы, схемы регулирования и процедуры голосования», а также значительно расширила знания об особенностях оптимального распределения экономических ресурсов.

звучит не категорически и сопровождается комментарием «если…»

Вместе с тем, возможность выбора критерия дает свободу лицам, принимающим экономические решения (если они, конечно, располагают достаточной информацией для постановки подобной задачи). Любой критерий должен согласовываться с намерениями решающего задачу и соответствовать его характеру, знаниям и убеждениям.

Еще один пример постановки и решения задачи.

Некая фирма, идя навстречу пожеланиям сельхозпроизводителей, которые нуждаются в хранении зерна, решила построить элеватор и эксплуатировать его в течение 5 лет (за эти годы хотелось бы рассчитаться за беспроцентные кредиты на строительство и получить заслуживающую внимания выгоду). Имеются типовые проекты элеватора мощностью на 20, 30, 40, 50 и 60 тысяч центнеров зерна.

Посевные площади сельхозрайона составляют 1430 га. Строительство элеватора мощностью 20 тыс. ц обойдется в 300 тыс. денежных единиц (ден. ед.) и эти затраты возрастают на 10 % с ростом мощности элеватора на 10 тыс. ц. Согласование проекта с районными властями, реклама будущего элеватора, строительство подъездных путей и вспомогательных сооружений обойдется в 185 тыс. ден. ед. Затраты на эксплуатацию элеватора мощностью 20 тыс. ц. составляют 10 тыс. д.е. и убывают на 10 % при увеличении мощности на 10 тыс. ц. За хранение зерна на счет элеватора вносится плата в размере 10 ден. ед. за 1 ц. Урожайность в данном районе колеблется от 14 до 20 ц с 1 га. Практика показывает, что зерновой запас элеватора расходуется за год и возобновляется при новом урожае. *Какой элеватор выгоднее построить*?

Если бы мы точно знали, каким будет урожай в каждом из пяти лет, то решение задачи можно было бы поручить добросовестному пятикласснику. Но в реальности, построив большой хороший элеватор, оснащенный автоматикой, и затратив значительную сумму, мы можем столкнуться с малым урожаем и, соответственно, с малым доходом от хранения. С другой стороны, построенный малый элеватор может не вместить большой урожай и будет упущена возможная выгода.

Примем типовые проекты элеваторов за **вектор допустимых решений**:

$$X = \{\, x_i \} = (\, 20, 30, 40, 50, 60\,) \qquad (\, i = 1 \div 5);$$

оценки урожайности в данном районе (здесь можно взять и другую сетку значений) примем за **вектор состояний внешней среды**:

$$S = \{\, S_j \} = (\, 14, 15, 16, 17, 18, 19, 20\,) \qquad (\, j = 1 \div 7)$$

и попытаемся построить **матрицу полезности** – эффективности принятия i -го решения в случае j -й урожайности.

Затраты на сооружение элеватора и инфраструктуру составляют 300000 + 30000 (x_i – 20) + 185000. Эти деньги по соглашению с кредитором придется возвратить равными долями в течение пяти лет. Ежегодные затраты на эксплуатацию элеватора равны 10000 – 1000 (x_i – 20).

Что касается доходной части, то здесь приходится учесть, что деньги мы получаем от реального урожая, который оказалось возможным поместить в элеваторе, Реальный урожай определяется величиной $1430S_j$ ц. Возможности элеватора – $1000x_i$ ц. Следовательно, наибольший сохраняемый объем зерна не превысит минимального из этих значений и плата за его хранение составит $10 \cdot min\,(1430S_j, 1000\cdot x_i)$.

Если просуммировать доходы-затраты за пять лет и разделить на 5, то мы получаем средний доход в год (принимаем его за функцию полезности) в виде матрицы с элементами

$$W_{ij} = 10 \cdot min\,(1430\,S_j, 1000x_i) -$$
$$-[60000 + 6000\,(x_i - 20) + 37000] - [10000 - 100\,(x_i - 20)]\,.$$

Выполнив несложные расчеты, заполним матрицу $\{W_{ij}\}$:

	s_1=14	s_2=15	s_3=16	s_4=17	s_5=18	s_6=19	s_7=20
x_1=20	**93000**	93000	93000	93000	93000	93000	93000
x_2=30	88200	**102500**	**116800**	**131100**	**145400**	**159700**	**174000**
x_3=40	83200	97500	111800	126100	140400	154700	169000
x_4=50	78200	92500	106800	121100	135400	149700	164000
x_5=60	73200	87500	101800	116100	130400	144700	159000

Если считать шансы на ту или иную урожайность равновероятными, то находим средние значения полезности $W_i(L) = \frac{1}{n}\sum_{j=1}^{n} W_{ij}$ для каждого из вариантов решения, например,

$$W_2 = (88200+102500+116800+131100+145400+$$
$$+159700+174000)\,/\,7 = 131100$$

и по критерию Лапласа устанавливаем оптимальность выбора проекта мощностью 30 тыс. ц с ожидаемой прибылью 131,1 тыс. ден. ед.

Если выбирать самый худший вариант по величине прибыли для каждой альтернативы (наименьшие значения $W_{ij}\,(B)$ полезности в строках матрицы W), можно из таких самых плохих оценок эффекта наших возможных выборов выбрать наилучший. Таким образом, по критерию Вальда обнаруживаем, что следует построить элеватор мощностью 20 тыс. ц и оправдываться, что даже в самом худшем случае здесь гарантирован максимум возможной прибыли 93 тыс. ден. ед.

Обратившись к оценкам *по критерию Гурвица* при трех различных уровнях оптимизма (α = 0,2; 0,5; 0,8), обнаруживаем целесообразность выбора проекта элеватора мощностью 30 тыс. ц с ожидаемой прибылью соответственно 105360, 131100, 156840 ден. ед.

	w$_i$(*L*)	**w$_i$(*B*)**	**α=0.2**	**α=0.5**	**α=0.8**
x_1=20	93000	**93000**	93000	93000	93000
x_2=30	**131100**	88200	**105360**	**131100**	**156840**
x_3=40	126100	83200	100360	126100	151840
x_4=50	121100	78200	95360	121100	146840
x_5=60	116100	73200	90360	116100	141840

При подходе с позиций *критерия Сэвиджа* (упущенных возможностей и последующего сожаления об этом) строим матрицу сожалений *D*, вычитая из столбцов матрицы полезности наибольшие значения, и применяем к ней пессимистический критерий Вальда, дающий наименьшее значение максимального сожаления.

	s_1=14	s_2=15	s_3=16	s_4=17	s_5=18	s_6=19	s_7=20	*min*
x_1=20	0	–9500	–23800	–38100	–52400	–66700	–81000	–81000
x_2=30	–4800	0	0	0	0	0	0	**–4800**
x_3=40	–9800	–5000	–5000	–5000	–5000	–5000	–5000	–9800
x_4=50	–14800	–10000	–10000	–10000	–10000	–10000	–10000	–14800
x_5=60	–19800	–15000	–15000	–15000	–15000	–15000	–15000	–19800

Для нашего примера по этому критерию оптимален проект элеватора мощностью 30000 ц. (прибегая к этому выбору, мы рискуем потерять прибыль до 4800 ден. ед.)

Таким образом, практически по всем критериям отдается предпочтение проекту 30000 ц, и лишь глубокий пессимист во взглядах на ожидаемый урожай отдаст предпочтение проекту 20000 ц с гарантией ожидаемой прибыли лишь в 93 тыс. ден. ед. и значительных упущенных возможностей. Остальные проекты рассматривать явно нецелесообразно.

11. ВВЕДЕНИЕ В СЕТЕВОЕ ПЛАНИРОВАНИЕ

Середина XX столетия знаменовалась исключительным усложнением управления и организации различных технических и научных разработок, собирающих воедино десятки или сотни конструкторских бюро, заводов, поставщиков и т. п. В такой обстановке усложнилась координация работ исполнителей и оценка хода выполнения работ. Одним из важнейших факторов стало максимальное сокращение сроков разработок с целью недопущения морального старения разрабатываемой системы.

Формулируя **закон необходимого разнообразия**, У. Р. Эшби указывал, что *для обеспечения процесса управления управляющая система должна обладать, по крайней мере, такой же сложностью, как и управляемая*.

Соответственно возникла необходимость создания системы, обеспечивающей возможность оценки текущего состояния и предсказания последующего хода разработки. Результатом исследований в этом направлении явилось создание систем, базирующихся на т. н. сетевых графиках, чему не в малой степени способствовало появление ЭВМ.

Первые системы такого рода получили название CPM (метод критического пути, впервые апробированный при управлении строительными работами) и PERT (метод оценки и обзора программ). Последняя впервые была применена для управления разработкой ракеты «Полярис», позволив сократить срок разработки, по мнению специалистов, на 2-3 года. Позднее методы сетевого планирования применялись при проектировании инженерных сооружений, постановке театральных спектаклей, организации переподготовки специалистов и так далее.

Было время, когда сетевое планирование было «модным» среди специалистов; сейчас им пользуются там, где оно действительно может оказаться полезным.

11.1. Понятие о сетевом графике

В терминологии теории графов *сетевым графиком называют конечный ориентированный граф без контуров, в котором имеются единственная вершина с отсутствующими прообразами и единственная вершина, не имеющая образов.*

Иначе *сетевым графиком можно назвать ориентированную транспортную сеть с одним входом и одним выходом, в которой нет путей с повторяющимися вершинами*.

Дуги (стрелки) указанного графа понимаются как некоторые *ра-*

боты. Вершины графа называются *событиями*.

Информация о сетевом графике некоторого проекта может задаваться в виде рисунка или различных списков. Если информация дана списком:

Работа	Последующие работы	Продолжительность	Работа	Последующие работы	Продолжительность
1	2, 4	3	6	5,9	4
2	8	4	7	6	3
3	5,9	5	8	-	4
4	6	3	9	8	2
5	-	3			

то ее можно представить в графическом виде (рис. 1); если ввести обозначения или нумерацию для вершин (событий), то можно использовать и графическое представление (рис. 2), где числа на дугах определяют продолжительность работ.

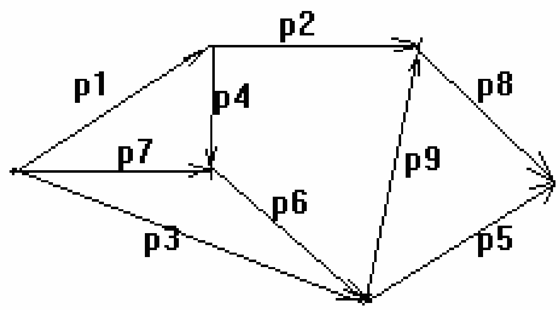


Рис . 1

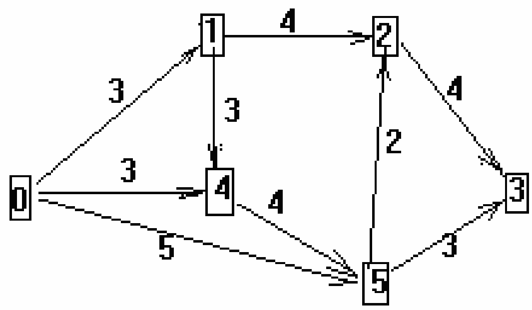


Рис . 2

При наличии нумерации вершин информацию о графике можно представить также перечнем работ с указанием начального и конечного событий и продолжительностей работ. Можно ту же самую информацию представить и в виде матрицы связей между событиями (элементы матрицы равны продолжительности соответствующих работ или не указаны при отсутствии таких работ).

Матричное задание особенно удобно и наглядно при обработке сетевого графика на ЭВМ .

Работа	Продолжител.	Работа	Продолжител.
0-1	3	4-5	4
1-2	4	0-4	3
0-5	5	2-3	4
1-4	3	5-2	2
5-3	3		

Событие	0	1	2	3	4	5
0		3			3	5
1			4		3	
2				4		
3						
4						4
5			2	3		

При рисовании сетевого графика (а при компьютерной обработке особенно) удобно использовать т.н. *фиктивные работы* – работы с нулевой продолжительностью, изображаемые пунктиром и служащие для указания порядка следования работ. Например, при наличии двух выходов их можно связать фиктивной работой.

Могут использоваться и *фиктивные события*. Так, если обнаружится, что пара событий связана более чем одной работой, то можно воспользоваться фиктивными событиями и работами (рис. 3).

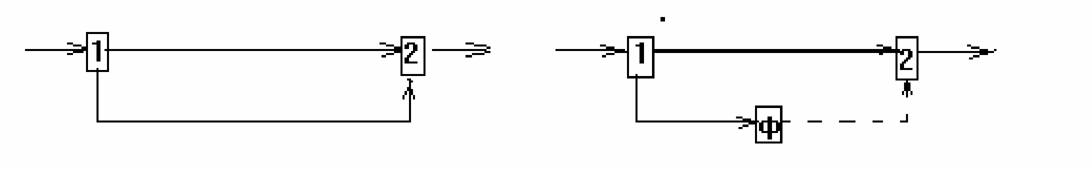


Рис. 3

Заметьте, что вышеуказанное требование отсутствия контуров неслучайно, т. к. наличие таковых создавало бы возможность возврата к повторению ранее выполненных работ.

В большинстве реальных проектов точный прогноз продолжительности работ T_{ij} невозможен, но на основании экспертизы могут быть предложены верхняя (*пессимистическая*) оценка B_{ij} , определяющая максимум продолжительности с учетом всех возможных срывов, и нижняя (*оптимистическая*) оценка A_{ij} , определяющая продолжительность работы в идеальных условиях. При наличии некоторого опыта может существовать и наиболее вероятная оценка M_{ij}. Соответственно, планируемая продолжительность работы (i, j) определяется по одной из формул:

$$T_{ij} = (2\,A_{ij} + 3\,B_{ij}) / 5, \quad T_{ij} = (A_{ij} + 4\,M_{ij} + B_{ij}) / 6 \quad .$$

При обработке сетевых графиков на объем вычислительных работ существенно влияет порядок просмотра (нумерация) событий.

Правило оптимальной нумерации связано с определением ранга событий.

Начальному событию (входу) сопоставляется ранг 0. Ранг 1 получают события, в которые приводят работы, начинающиеся только в событии ранга 0. Ранг 2 получают события, в которые приводят работы, начинающиеся только в событиях ранга 0 и 1 и т. д.

Последующая нумерация ведется в соответствии с рангами по возрастанию номеров (при одинаковом ранге порядок произволен; нумерация не обязательно сплошная, лишь бы соблюдалось условие $i < j$).

Для приведенного выше примера ранг 0 получает событие 0, ранг 1 – событие 1, ранг 2 – событие 4, ранг 3 – событие 5, ранг 4 – событие 2 и ранг 5 – событие 3.

11. 2. Критический путь и другие параметры сетевого графика

Если продолжительность работы принять за длину соответствующей дуги сетевого графика, то *критическим* можно назвать путь максимальной длины от входа до выхода графика. Длина этого пути определяет критическое время выполнения проекта, т. е. минимальное время, в пределах которого коллектив исполнителей в состоянии выполнить весь комплекс работ сетевого графика.

Пусть для определенности начальное событие имеет номер 0 и конечное – номер N. Обозначим через L_j длину пути наибольшей протяженности от события 0 до события j.

Согласно принципу оптимальности Р. Беллмана, здесь как бы обращенному в прошлое, получаем систему функциональных уравнений

$$L_j = \max_{(i\,j)} [\, L_i + T_{ij}\,] \;,\; j > 0 \;; \quad L_0 = 0 \quad .$$

Величина L_j соответствует *наиболее раннему* возможному времени T_j^0 наступления j –го события, т. е. самому раннему сроку завершения всех работ, предшествующих этому событию. Значение L_N определяет критическое время выполнения проекта $T_{крит}$.

Обозначим через M_i длину пути наибольшей протяженности от события i до события N . Тогда по тому же принципу оптимальности

$$M_i = \max_{(i\,k)} [\, T_{ik} + M_k\,] \;, i < N \;; \quad L_N = 0 \quad .$$

Величина $T_i^1 = T_{крит} - M_i$ соответствует *наиболее позднему* допустимому времени наступления i -го события, т.е. самому позднему сроку начала всех работ, последующих за этим событием. Совершенно очевидно, что для событий на критическом пути самое раннее и самое позднее времена их наступления будут совпадать.

Рассмотрим сетевой график (рис. 4) :

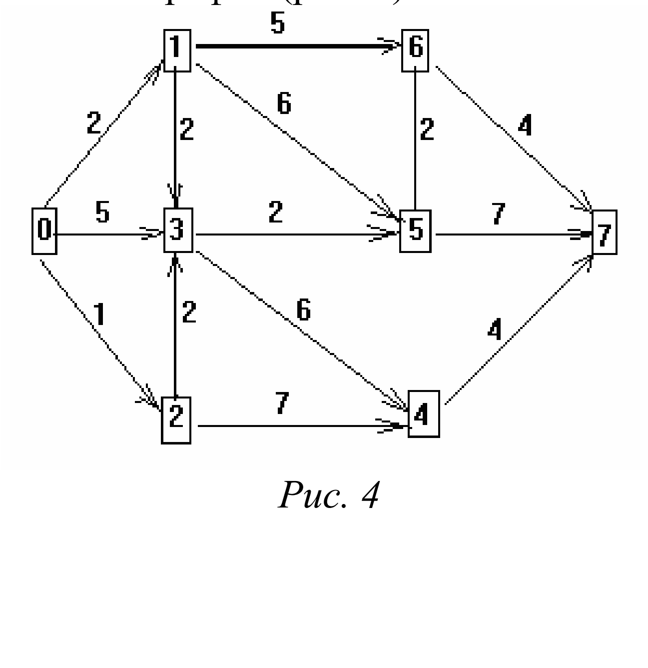


Рис. 4

Здесь рассчитываем значения L_j в порядке роста номеров:

$L_0 = 0$; $L_1 = 2$ (i=0); $L_2 = 1$ (i=0);

$L_3 = max\ [L_0 +5, L_1 +2, L_2 +2] = 5$ (i=0) и т. д.

Затем рассчитываем значения M_i в порядке убывания номеров:

$M_7 =0$; $M_6 =4$ (k=7); $M_5 =max\ [2+M_6, 7+M_7] = 7$ (k=7) ;

$M_4 =4$ (k=7); $M_3 =max\ [6+M_4, 2+M_5] =10$ (k=4) и т.д.

В итоге имеем информацию о наиболее ранних и наиболее поздних моментах наступления событий и индексы предшествующих и последующих событий в путях наибольшей длины, проходящих через данное событие.

Работа (i,j)	T_{ij}	Работа (i,j)	T_{ij}
0-1	2	2–4	7
0-2	1	3–4	6
0-3	5	3–5	2
1-3	2	4–7	4
1-5	6	5–6	2
1-6	5	5–7	7
2-3	2	6–7	4

Событие j	$L_j = T_j^0$	i пред.	Mj	k посл.	Tj^1
0	0	–	15	1,3	0
1	2	0	13	5	2
2	1	0	12	3	3
3	5	0	10	4	5
4	11	3	4	7	11
5	8	1	7	7	8
6	10	5	4	7	11
7	15	4,5	0	–	15

По информации из колонок 3 или 5 можно выявить критические пути с длиной 15: [0 – 1 –5 – 7] и [0– 3 – 4 –7] .

Очевидно, что работы, не лежащие на критических путях, обладают *резервами* времени – их выполнение при некоторых условиях может быть задержано на какое-то время .

Существуют 4 вида резервов :

– *полный резерв* $R_{ij} = T_j^1 - T_i^0 - T_{ij}$;

– *свободный резерв* $R_{ij} = T_j^0 - T_i^0 - T_{ij}$;

– *независимый резерв* $R_{ij} = max\ [\ T_j^0 - T_i^1 - T_{ij}\ , 0\ ]$;

– *частный резерв* $R_{ij} = T_j^1 - Ti^1 - T_{ij}$.

Так полный резерв работы можно понимать как время, на которое можно замедлить выполнение работы, если предшествующие работы завершатся к самому раннему возможному сроку, но комплекс последующих работ начнется в самый последний приемлемый момент.

Независимый резерв предполагает весьма жесткую гипотезу – завершение предшествующих работ к самому позднему, но начало последующих в самый ранний возможный срок.

Результаты обработки приведенного сетевого графика можно представить следующей таблицей:

Работа	Продол жител.	Раннее время		Позднее время		Р е з е р в ы			
		начала	конца	начала	конца	полн.	своб.	незав.	част.
0-1	2	0	2	0	2	0	0	0	0
0-2	1	0	1	2	3	2	0	0	2
0-3	5	0	5	0	5	0	0	0	0
1-3	2	2	4	3	5	1	1	1	1
1-5	6	2	8	2	8	0	0	0	0
1-6	5	2	7	6	11	4	3	3	4
2-3	2	1	3	3	5	2	2	0	0
2-4	7	1	8	4	11	3	3	1	1
3-4	6	5	11	5	11	0	0	0	0
3-5	2	5	7	6	8	1	1	1	1
4-7	4	11	15	11	15	0	0	0	0
5-6	2	8	10	9	11	1	0	0	1
5-7	7	8	15	8	15	0	0	0	0
6-7	4	10	14	11	15	1	1	0	0

Полученные данные позволяют выделить т. н. *подкритические* работы, т.е. работы, лежащие на путях, отличающихся по длине от критического не более чем на заданную величину. Основной характеристикой здесь может служить *полный резерв*.

Для нашего графика на уровне критичности 1 подкритическими будут работы 1–3, 3–5, 5–6, 6–7. Чтобы убедиться в этом, возьмем работу 5–6 и найдем путь максимальной длины, проходящий через нее, по индексам предшествующих и последующих событий. Так событию 5 в пути максимальной длины предшествует событие 1, а ему –событие 0. Событию 6 в таком пути последует событие 7. Длина пути 0–1–5–6–7 равна 14 и задержка на 1 при выполнении работ 5–6 или 6–7 сделает его критическим.

Однако полный резерв не совсем характеризует уровень критичности работ. Возьмем для примера два графика (рис. 5).

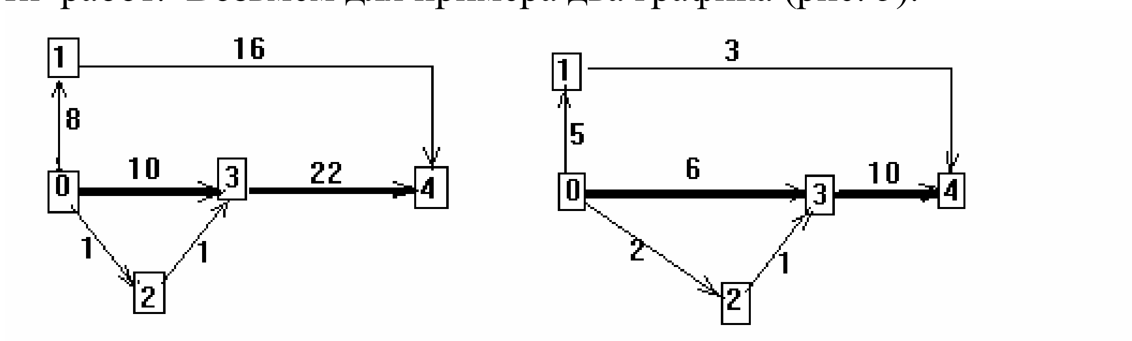


Рис . 5

В первом графике все некритические работы имеют одинаковый полный резерв, равный 8, но напряженность работ пути 0 – 1 – 4 составляет 24 единицы времени на интервале 32, тогда как напряженность работ пути 0 – 2 – 3 составляет 2 единицы на 10. Нет сомнения, что работы второго пути можно выполнять с большей «прохладцей», чем первого (отдыхаем 8 дней из 10 и 8 дней из 32 соответственно).

Во втором графике работы 0–2 и 2–3 имеют резерв 3, а работы 0–1 и 1–4 – резерв 8, но напряженность у них одинакова.

Поэтому всякая работа характеризуется и т. н. *коэффициентом напряженности*. Здесь отыскивается путь максимальной длины, проходящий через данную работу: при этом используются индексы предшествующих и последующих событий, которые мы находили при поиске T^0 и T^1. На этом пути ищутся ближайшие "слева" и "справа" события, принадлежащие критическому пути (путям), и определяется отношение длины пути между этими событиями, проходящего через данную работу, к длине соответствующего отрезка критического пути.

Так для рассмотренного выше сетевого графика выберем некритическую работу 1–3. Обнаруживаем, что через нее проходит путь (максимальной длины) 0 – 1 –3 – 4 – 7. Ближайшими соседями на критических путях 0 – 1 – 5 – 7 и 0 – 3 – 4 – 7 будут события 1, 7 и 0, 3 соответственно. Отсюда находим коэффициент напряженности

$$K_{13} = max\left|\frac{T_{13}+T_{34}+T_{47}}{T_{15}+T_{57}}, \frac{T_{01}+T_{13}}{T_{03}}\right| = max\left|\frac{12}{13}, \frac{4}{5}\right| = \frac{12}{13}.$$

Аналогично получаем

$$K_{35} = max\left|\frac{T_{03}+T_{35}}{T_{01}+T_{15}}, \frac{T_{35}+T_{57}}{T_{34}+T_{47}}\right| = max\left|\frac{7}{8}, \frac{9}{10}\right| = \frac{9}{10},$$

$$K_{16} = \max\left|\frac{T_{16}+T_{67}}{T_{15}+T_{57}}, \frac{T_{01}+T_{16}+T_{67}}{T_{03}+T_{34}+T_{47}}\right| = max\left|\frac{9}{13}, \frac{11}{15}\right| = \frac{11}{15},$$

$$K_{56} = K_{67} = max\left|\frac{T_{56}+T_{67}}{T_{57}}, \frac{T_{01}+T_{15}+T_{56}+T_{67}}{T_{03}+T_{34}+T_{47}}\right| = max\left|\frac{6}{7}, \frac{14}{15}\right| = \frac{14}{15},$$

$$K_{02} = K_{23} = max\left|\frac{T_{02}+T_{23}+T_{34}+T_{47}}{T_{01}+T_{15}+T_{57}}, \frac{T_{02}+T_{23}}{T_{03}}\right| = max\left|\frac{13}{15}, \frac{3}{5}\right| = \frac{13}{15},$$

$$K_{24} = max\left|\frac{T_{02}+T_{24}+T_{47}}{T_{01}+T_{15}+T_{57}}, \frac{T_{02}+T_{24}}{T_{03}+T_{34}}\right| = max\left|\frac{12}{15}, \frac{8}{11}\right| = \frac{12}{15}.$$

14.3. Линейная диаграмма проекта

Для небольших проектов с целью большей наглядности выполнения работ во времени после составления пронумерованного сетевого графика можно построить т. н. линейную диаграмму (рис. 6).

Работу (i, j) изображаем на диаграмме в виде отрезка так, чтобы его начало лежало на одной вертикали с самым правым концом работ, заканчивающихся в вершине i, что соответствует T_i^0. Самый правый конец всех отрезков соответствует $T_{крит}$.

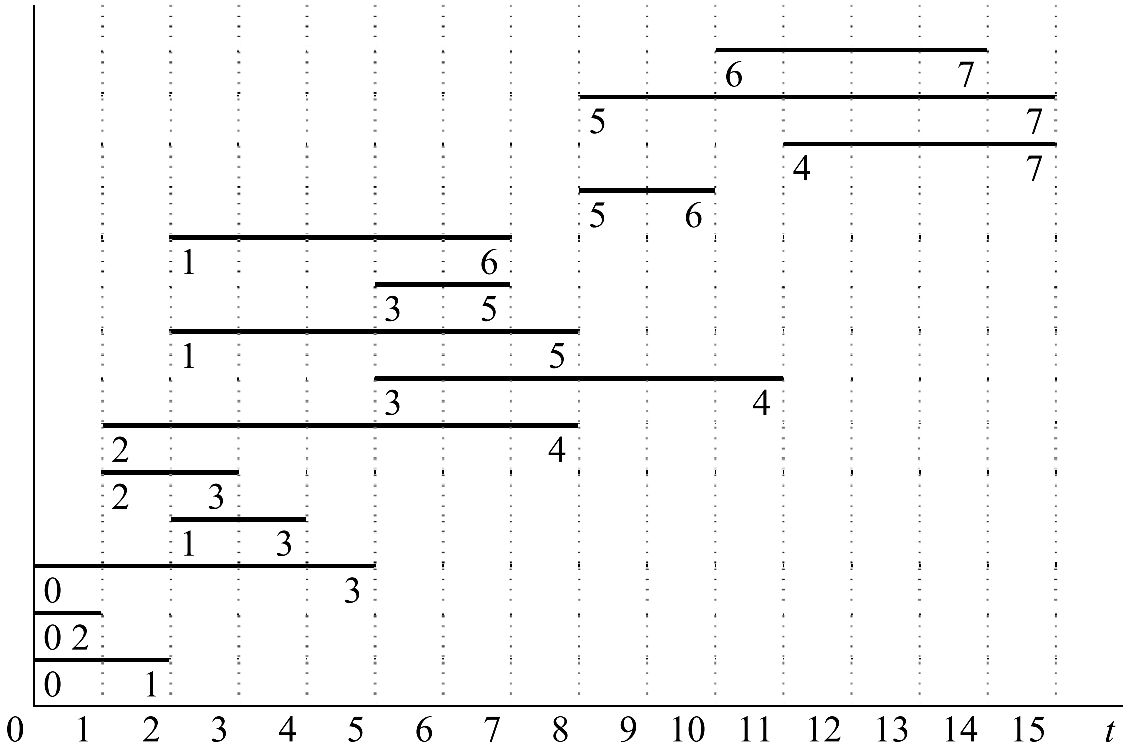


Рис. 6

Если сдвинуть отрезки вправо так, чтобы отрезки $i - j$ заканчивались самым левым концом отрезков с начальным индексом j, то эти самые левые концы будут соответствовать T_j^1. Величина такого сдвига определяет полный резерв работы.

Если сдвигать отрезок $i - j$ вправо без сдвига отрезков с начальным индексом j, то величина такого сдвига определит свободный резерв.

Существенным достоинством линейной диаграммы является возможность оценить загрузку исполнителей во времени. Так для t в интервале от 2 до 3 выполняется наибольшее количество работ (шесть). Если работы выполняются взаимозаменимыми исполнителями, то сдвинув работу 1–6 или 2–4, мы обнаружим возможность обойтись пятью исполнителями (непосредственно из сетевого графика это трудно увидеть).

11.4. Минимизация стоимости проекта при заданной продолжительности

Выполнение всякой работы связано с затратами. Достаточно часто, в условиях плохой организации или плохого планирования, замедление в выполнении некоторой работы чревато лишними затратами («исчезают» материалы со стройки, ряд исполнителей при недозагрузке работой находят время загружать себя горячительными напитками и др.). Однако столь же часто ускоренное выполнение работы связано с увеличением затрат (авралы, срочные поставки, привлечение специалистов экстра-класса и т. п.) и затраты являются обратной функцией от времени выполнения, которая выбирается обычно в виде:

1) линейный вариант $C_{ij} = -A_{ij} \cdot T_{ij} + B_{ij}$;

2) выпуклый вариант $C_{ij} = -A_{ij} / T_{ij}$.

В таких условиях может быть поставлена задача поиска *оптимального по стоимости безрезервного плана*, т.е. плана, в котором стоимость снижается удлинением выполнения работ до предельного допустимого времени.

Пусть проект требуется выполнить за время T, не большее $T_{крит}$.

Если обозначить через D_{ij} минимальное необходимое время выполнения работы $i-j$, а через T_j – момент наступления j -го события, то продолжительность работы $i-j$ принимаем равной $T_j - T_i$ и возникает задача:

минимизировать функцию

$$Z = \sum_{(i\,j)} [-A_{ij}(T_j - T_i) + B_{ij}]$$

при условиях

$$T_j - T_i \geq D_{ij} \text{ при всех } (i\,j) ;$$

$$T_0 = 0 , T_{вых} = T .$$

<u>Пример</u>. Пусть для рассмотренного выше сетевого графика заданы параметры стоимости и продолжительности работ:

$i-j$	A_{ij}	B_{ij}	D_{ij}	$i-j$	A_{ij}	B_{ij}	D_{ij}	$i-j$	A_{ij}	B_{ij}	D_{ij}
0–1	10	100	2	2–4	4	140	7	5–6	5	50	2
0–2	3	40	1	3–4	2	90	6	4–7	9	180	4
0–3	5	150	5	1–5	1	80	6	5–7	5	140	7
1–3	2	70	2	3–5	5	120	2	6–7	9	200	4
2–3	6	170	2	1–6	2	60	5				

и предельное время $T = 25$.

Минимизируемую функцию Z можно преобразовать к виду (обычное приведение подобных):

$$Z = \sum_{(i\,j)} B_{ij} - \sum_k A_k T_k ,$$

где
$$A_k = \sum_i A_{ik} - \sum_j A_{kj} ,$$

и задачу минимизации заменить задачей максимизации суммы $\sum_k A_k T_k$.

Для нашего примера

$A_0 = -18;\ A_1 = 5;\ A_2 = -7;\ A_3 = 6;\ A_4 = -3;\ A_5 = -4;\ A_6 = -2;\ A_7 = 23$

и возникает задача максимизации функции

$$-18\, T_0 + 5\, T_1\, -7\, T_2\, +6\, T_3\, -3\, T_4 -4\, T_5 -2\, T_6 +23\, T_7$$

при условиях

$T_1 - T_0 \geq 2$ $\quad T_3 - T_1 \geq 2$ $\quad T_4 - T_3 \geq 6$ $\quad T_6 - T_1 \geq 5$ $\quad T_7 - T_5 \geq 7$

$T_2 - T_0 \geq 1$ $\quad T_3 - T_2 \geq 2$ $\quad T_4 - T_1 \geq 6$ $\quad T_6 - T_5 \geq 2$ $\quad T_7 - T_6 \geq 4$

$T_3 - T_0 \geq 5$ $\quad T_4 - T_2 \geq 7$ $\quad T_5 - T_3 \geq 2$ $\quad T_7 - T_4 \geq 2$ $\quad T_0 = 0$, $T_7 \leq 25$

Решение этой задачи симплексным методом дает оптимальные времена наступления событий $T_1 = 11$, $T_2 = 1$, $T_3 = 15$, $T_4 = 21$, $T_5 = 17$, $T_6 = 19$ и значение $Z = 1146$.

Можно предложить решение задачи по *алгоритму И. А. Радчик* [32], являющемуся модификацией *венгерского метода Форда – Фалкерсона*.

Поставленная задача сводится к максимизации

$$F(T) = \sum_k A_k T_k$$

при условиях $\quad T_i - T_j \leq -D_{ij}$ при всех ij ;

$$T_n - T_0 \leq T\ , \quad T_0 = 0 \quad .$$

(индекс n соответствует выходу графика).

Сопряженная задача состоит в минимизации

$$G(X) = -\sum_{(i\,j)} D_{ij}\, X_{ij} + T\cdot X_{n0}$$

при ограничениях

$$\sum_j X_{ij} - \sum_k X_{ki} = A_i \quad \text{при всех} \quad i,$$

$$X_{ij} \geq 0 \quad \text{при всех} \quad ij\ , \qquad X_{n0} > 0\ .$$

Величину X_{ij} можно трактовать как количество вещества, протекающего по дуге $i - j$ в единицу времени, и значения A_i – как разницу между притоком и оттоком вещества в вершине i.

Соответственно возникает задача о потоке в сети, определенной сетевым графиком и дополненной дугой от выхода до входа.

На предварительном этапе отыскивается какое-нибудь допустимое решение исходной задачи. Например, оценки T_j берутся равными минимальным временам наступления событий (T_n берется равным T).

Затем строится матрица значений $R_{ij} = T_j - T_i - D_{ij}$ (значение $R_{n0} = 0$).

На каждом очередном шаге отыскивается путь от какого-то источника ($A_i > 0$) до какого-то стока ($A_i < 0$).

Для этого метим строки с $A_i > 0$ символом * (эти же метки переносим на столбцы). В строке *, например i_0, берем клетки с $R_{i0,j} = 0$ и отмечаем ранее неотмеченные столбцы индексами (i_0, $b_1 = A_{i0}$). Затем по значениям $X_{i0,j} < 0$ метим ранее неотмеченные столбцы индексами (i_0, $b_1 = min(A_{i0}, |X_{i0,j}|)$). Метки столбцов переносим на строки.

Затем выбираем отмеченную индексом (i_0, b_1) строку i и столбцы с значениями $R_{ij} = 0$ и значениями $X_{ij} < 0$ метим индексами (i, $b_2 = b_1$) или (i, $b_2 = min(b_1, |X_{ij}|)$). Метки столбцов переносим на строки и т.д., пока не будет отмечен некоторый сток или не обнаружится невозможность дальнейшего отмечания.

В первом случае, если некий j_0-й сток помечен индексом (k,b_s), отыскиваем величину $V = min(b_s, A_{j0})$ и обратным ходом по первому из индексов выявляем путь, который привел к этому стоку. Значения X_{ij} на этом пути увеличиваем на V и симметричные значения уменьшаем на V. Уменьшаем на V значение A_{i0} и увеличиваем на V значение A_{j0}.

Во втором случае отыскиваем величину H, равную минимальному из значений R_{ij}, лежащих на пересечении отмеченных строк и неотмеченных столбцов.

Если вход и выход помечены, вычитаем H из всех R_{ij}, находящихся в непомеченных столбцах, и прибавляем к находящимся в непомеченных строках. Уменьшаем на H и значения T_j для непомеченных столбцов. Если вход и выход не помечены, то вычитание H проводим для помеченных строк и помеченных столбцов, причем значения T_j для помеченных столбцов увеличиваем на H.

Алгоритм завершает работу, когда все A_j станут равными нулю, т.е. все источники опустошены и все стоки насыщены.

Строим начальную таблицу значений R_{ij} на основе ранее найденных оценок $T_j^{\,0}$ (значения X_{ij} равны 0):

		*		*		1, 5		*		
$i \setminus j$	0	1	2	3	4	5	6	7	Ai	
0		0	0	0					–18	
1				1		0	3		5	*
2				2	3				–7	
3					0	1			6	*
4								10	–3	
5							0	10	-4	1, 5
6								11	–2	
7	0								23	*
T_j	0	2	1	5	11	8	10	25		

Отмечаем «источники» 1, 3 и 7 символом *. В строке 1 обнаруживается $R_{15}=0$ и столбец 5 метим индексом (1.5) (из пункта 1 поток емкости $b_1=5$). Этот столбец соответствует стоку с потребностью 4, т.е. найден путь 1–5 с величиной потока $V=min(5, |-4|)$.

Увеличиваем на V значение A_5 и X_{15}, уменьшая A_5 и X_{51} (значения X_i записываем в этой же таблице под дробной чертой).

		*		*		1,1	5,1	*		
$i \setminus j$	0	1	2	3	4	5	6	7	A_i	
0		0	0	0					–18	
1				1		0/4	3		1	*
2				2	3				–7	
3					0	1			6	*
4								10	–3	
5		/–4					0	10	0	1,1
6								11	–2	5,1
7	0								23	*
T_j	0	2	1	5	11	8	10	25		

Здесь обнаруживается путь 1-5-6 интенсивности 1 и $V=min(1, |-2|)=1$. Увеличиваем X_{15}, X_{56}, A_6 и уменьшаем X_{51}, X_{65}, A_1.

	7, 23			*	3,6			*		
$i \setminus j$	0	1	2	3	4	5	6	7	A_i	
0		0	0	0					–18	7,23
1				1		0/5	3		0	
2				2	3				–7	
3					0	1			6	*
4								10	–3	3, 6
5		/–5					0/1	10	0	
6						/–1		11	–1	
7	0								23	*
T_j	0	2	1	5	11	8	10	25		

В полученной матрице обнаруживается путь 3 - 4 интенсивности 3 и путь 7-0 интенсивности 18 .

Корректируем X_{34}, X_{43}, X_{07}, X_{70} и A_4, A_3, A_0, A_7.

	7,5		0,5	*	3,3			*		
$i \setminus j$	0	1	2	3	4	5	6	7	Ai	
0		0	0	0				/–18	0	7,5
1				1		0/5	3		0	
2				2	3				–7	0,5
3					0/3	1			3	*
4				/–3				10	0	3,3
5		/–5					0/1	10	0	
6						/–1		11	–1	
7	0/18								5	*
T_j	0	2	1	5	11	8	10	25		

В полученной матрице обнаруживается путь 7–0–2 интенсивности 5 и корректируем X_{70} , X_{02} , X_{07} , X_{20} и A_7 , A_2 .

				*	3,3					
i \ j	0	1	2	3	4	5	6	7	A_i	
0		0	0/5	0				/–23	0	
1				1		0/5	3		0	
2	/–5			2	3				–2	
3					0/3	1			3	*
4				/–3				10	0	3,3
5		/–5					0/1	10	0	
6						/–1		11	–1	
7	0/23								0	
T_j	0	2	1	5	11	8	10	25		

Обнаружив невозможность достижения стоков 2 и 6, ищем в строках 3 и 4 в непомеченных столбцах минимальное из значений $R_{i\,j}$: $H = min(R_{35}\ , R_{47}\ ) = 1$. Вычитаем H из строк 3 , 4 и добавляем к столбцам 3, 4 (в том числе и к значениям T_3 , T_4) .

				*	3.3	3.3	5.3			
i \ j	0	1	2	3	4	5	6	7	Ai	
0		0	0/5	0				/–23	0	
1				1		0/5	3		0	
2	/–5			3	4				–2	
3					0/3	0			3	*
4				/–3				9	0	3.3
5		/–5					0/1	10	0	3.3
6						/–1		11	–1	5.3
7	0/23								0	
Tj	0	2	1	6	12	8	10	25		

В полученной матрице обнаруживается путь 3–5 –6 интенсивности 1 и корректируем X_{35}, X_{56} , X_{53}, X_{65}, A_3 , A_6 .

		5,2		*	3,2	3,2	5,2			
i \ j	0	1	2	3	4	5	6	7	A_i	
0		0	0/5	1				/–23	0	
1				2		0/5	3		0	5,2
2	/–5			3	4				–2	
3					0/3	0/1			2	*
4				/–3				9	0	3,2
5		/–5		/–1			0/2	10	0	3,2
6						/–2		11	0	5,2
7	0/23								0	
T_j	0	2	1	6	12	8	10	25		

В отличие от предыдущих таблиц здесь используем отмечание не только по R_{ij}=0, но и по $X_{ij} < 0$ (столбец 1). Добраться до стока 2 невозможно. Потому находим $H = min\, R_{i\,j}$ при i = 1, 3, 4, 5, 6 и j = 0, 2, 7 ($H = min$ [9, 10, 11] = 9) и вычитаем из отмеченных строк с добавлением к отмеченным столбцам.

	7,2	5,2	0,2	*	3,2	3,2	5,2	4,2		
$i \setminus j$	0	1	2	3	4	5	6	7	A_i	
0		9	0/5	10				/–23	0	7,2
1				2		0/5	3		0	5,2
2	/–5			12	13				-2	0,2
3					0/3	0/1			2	*
4				/–3				0	0	3,2
5		/–5		/–1			0/2	1	0	3,2
6						/–2		2	0	5,2
7	0/23								0	4,2
T_j	0	11	1	15	21	17	19	25		

Здесь обнаруживается путь 3 – 4 – 7 – 0 – 2 с интенсивностью 2 и после корректуры значения A_3 и A_2 обращаются в нуль, что служит признаком конца решения задачи.

Оптимальный безрезервный план определяется временами наступления событий :

$$T_0 = 0,\ T_1 = 11,\ T_2 = 1,\ T_3 = 15,\ T_4 = 21,\ T_5 = 17,\ T_6 = 19,\ T_7 = 25.$$

В рассмотренной задаче предполагалась ограниченность продолжительности работ снизу. При наличии ограничения сверху задача может быть поставлена в виде:

минимизировать функцию

$$Z = \sum_{(i\,j)} [\, - A_{i\,j}\ t_{ij} + B_{ij} \,]$$

при условиях $T_j - T_i \geq t_{ij}$ при всех $i\,j$,

$$D_{ij} \leq t_{ij} \leq W_{ij} \text{ при всех } i\,j\,,\quad T_0 = 0,\ T_{вых} = T\,.$$

Решение задачи дает *оптимальный по стоимости резервный план*. Если считать T переменной величиной, то поставленная задача становится задачей параметрического линейного программирования.

Определенный интерес представляет и задача минимизации времени выполнения проекта при заданной его стоимости.

Минимизировать T_n при условиях

$$T_j - T_i \geq t_{ij} \text{ при всех } i\,j\,,$$

$$D_{ij} \leq t_{ij} \leq W_{ij} \text{ при всех } i\,j\,,$$

$$T_0 = 0\,,\quad \sum_{(i\,j)} [\, - A_{ij}\, t_{ij} + B_{ij} \,] \leq C\,.$$

11.5. Проблемы применения систем сетевого планирования

Выше мы ориентировались на сетевой график некоторого самостоятельного проекта, на выполнение которого направлены усилия коллектива исполнителей. В реальности один и тот же коллектив выполняет работы "одновременно" по нескольким проектам и даже при идеальной отработке графиков каждого из них нет уверенности в выполнении всех проектов, т.к. выполнение многих тем приходится на один и тот же промежуток времени. При выполнении таких "многотемных" разработок наряду с оценкой работ по времени приходится учитывать трудоемкость работ (количество человеко-часов или количество специалистов в этой области), мощность подразделений исполнителей, возможность выполнения работы подразделениями.

Поэтому сначала для каждой темы разрабатывают сетевой график и проводят оценки не только по времени, но и по исполнителям. Работы из всех тем сортируют по подразделениям и накладывают на календарь. Оценивают возможности подразделения и, если все работы выполнить в данный период невозможно, часть из них переносят на более поздние сроки с соответствующими отметками в исходных графиках.

Сетевые графики могут иметь стохастическую структуру по оценке времени выполнения работ. Здесь по заданным пессимистической и оптимистической оценкам отыскивают математическое ожидание и дисперсию продолжительности работ:

$$T_{ij} = \frac{2A_{ij} + 3B_{ij}}{5}; \qquad D_{ij} = \left| \frac{B_{ij} - A_{ij}}{5} \right|^2 .$$

Первая из этих оценок носит дискуссионный характер, полагая пессимистов заслуживающими большего доверия (говорят, что пессимист – это хорошо информированный оптимист). После традиционной обработки графика оценивают дисперсию длины критического пути как сумму дисперсий составляющих его работ.

Предполагается, что продолжительность выполнения проекта (отдельной работы) имеет нормальное распределения. При длине критического пути в 38 дней и суммарной дисперсии 6.31, т. е. стандартном отклонении равном 2.51, вероятность того, что фактическое время выполнения проекта лежит в интервале от 35 до 42 дней, определится как разность значений функции нормального распределения при аргументах (35–38) / 2.51 и (42–38) / 2.51 и составит 0.822. Подробнее см. [32–34].

12. ВВЕДЕНИЕ В ТЕОРИЮ МАССОВОГО ОБСЛУЖИВАНИЯ

12.1. Понятие о задачах теории массового обслуживания

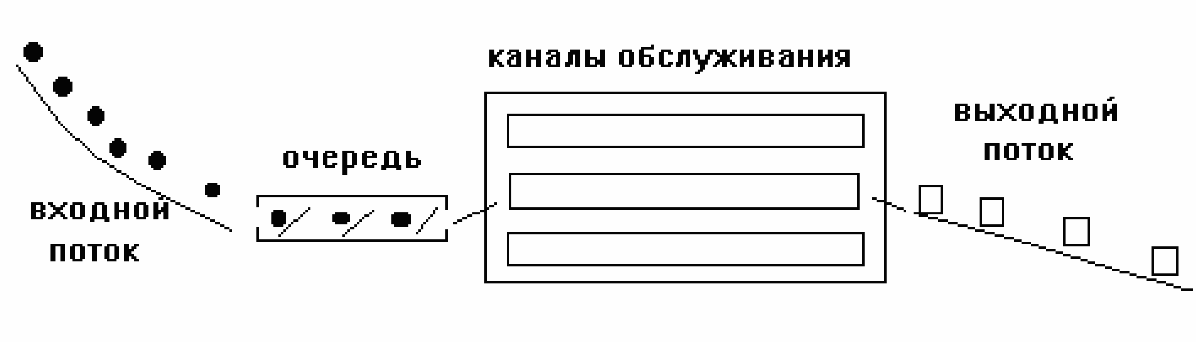
Буквально с момента рождения вам приходится сталкиваться с очередями.

Ваши родители сидят в очереди в ЗАГСе, чтобы официально зафиксировать факт вашего появления на свет, и это ожидание может завершиться через минуты или часы... Потом они становятся в очередь на ваше попадание в детский сад и эта очередь растягивается на годы... Вы набираете телефонный номер вашей подруги и слышите продолжительные гудки или вежливое приглашение оставить сообщение ... Не дозвонившись, вы решаете для экономии времени воспользоваться собственным лимузином и попадаете в традиционную «пробку»... Ваш самолет запросил посадку в Рио-де-Жанейро и, получив отказ, совершил посадку в Буэнос-Айресе... В старости отправляетесь на прием к офтальмологу в поликлинику по месту жительства и узнаете, что на ближайший месяц «очередей» нет...

Очереди возникают практически во всех системах массового обслуживания (С М О) и *теория массового обслуживания* (*теория очередей*) занимается оценкой функционирования системы при заданных параметрах и поиском параметров, оптимальных по некоторым критериям.

Эта теория представляет особый раздел теории случайных процессов и использует в основном аппарат теории вероятностей. Первые публикации в этой области относятся к 20-м гг. XX века и принадлежат датчанину А. Эрлангу, занимавшемуся исследованиями функционирования телефонных станций – типичных СМО, где случайны моменты вызова, факт занятости абонента или всех каналов, продолжительность разговора. В дальнейшем теория очередей нашла развитие в работах К. Пальма, Ф. Поллачека, А. Я. Хинчина, Б. В. Гнеденко, А. Кофмана, Р. Крюона, Т. Саати и других советских и зарубежных математиков.

В качестве основных элементов СМО следует выделить входной поток заявок, очередь на обслуживание, механизм обслуживания и выходящий поток. В роли заявок (требований, вызовов) могут выступать покупатели в магазине, телефонные вызовы, поезда при подходе к железнодорожному узлу, вагоны под разгрузкой, автомашины на станции техобслуживания, самолеты в ожидании разрешения на взлет, штабель бревен при погрузке на автотранспорт. Роль обслуживающих приборов (каналов, линий) играют продавцы или кассиры в магазине, таможенники, пожарные машины, взлетно-посадочные полосы, экзаменаторы, ремонтные бригады.



В зависимости от характеристик этих элементов СМО классифицируются следующим образом.

1. *По характеру поступления заявок.* Если интенсивность входного потока (количество заявок в единицу времени) постоянна или является заданной функцией от времени, поток называют *регулярным*. Если параметры потока независимы от конкретного момента времени, поток называют *стационарным*.

2. *По количеству одновременно поступающих заявок.* Поток с вероятностью одновременного появления двух и более заявок, равной нулю, называется *ординарным*.

3. *По связи между заявками.* Если вероятность появления очередной заявки не зависит от количества предшествующих заявок, имеем дело с потоком *без последействия*.

4. *По однородности заявок* выделяют *однородные* и *неоднородные потоки*.

5. *По ограниченности потока заявок* различают *замкнутые* и *разомкнутые* системы (система с ограниченной клиентурой называется замкнутой). Так универсальный магазин является разомкнутой системой, тогда как оптовый магазин с постоянными клиентами – замкнутая система.

6. *По поведению в очереди* системы делятся на системы *с отказами* (заявка покидает систему, если нет мест в очереди), *с ограниченным ожиданием* и *с ожиданием без ограничения времени*.

7. *По дисциплине выбора на обслуживание.* Здесь можно выделить системы с обслуживанием в порядке поступления, в случайном порядке, в порядке, обратном поступлению (последний пришел – первым обслужен) или с учетом приоритетов.

8. *По числу каналов обслуживания* системы разделяют на *одно- и многоканальные*.

9. *По времени обслуживания* выделяют системы *с детерминированным* и *случайным* временем.

10. *По количеству этапов обслуживания* различают *однофазные* и *многофазные* системы .

12.2. Основы математического аппарата анализа простейших СМО

Рассмотрим *стационарный поток однородных заявок* без последействия. Пусть $P_k(\tau)$ вероятность появления k заявок в интервале времени τ. Эта вероятность зависит только от τ и не зависит от начала отсчета времени, от поступления заявок в предыдущих временных интервалах . Пусть к тому же поток является ординарным, т. е. $P_k(dt)$ при $k>1$ бесконечно мала в сравнении с малым интервалом dt . Если обозначить через λ число заявок в единицу времени (интенсивность потока), то можно показать, что для такого *простейшего* потока

$$P_k(t) = \frac{(\lambda \cdot t)^k}{k!} e^{-\lambda t}. \qquad (1)$$

Формула (1) определяет *распределение Пуассона*. Для пуассоновского потока можно обнаружить, что промежутки времени T между поступлениями заявок распределены по экспоненциальному (показательному) закону

$$P(T < t) = 1 - e^{-\lambda t} \qquad (2)$$

(вероятность, что промежуток времени не превышает t) .

Естественно, что входной поток может описываться не только пуассоновским, но и другими распределениями (Эрланга, гиперэкспоненциальным и т. п.).

Аналогичная ситуация имеет место и для выходного потока. Чаще всего используется показательный закон распределения времени обслуживания:

$$P(t) = 1 - e^{-\mu t}, \qquad (3)$$

где $\mu = 1/t_{обсл}$ – интенсивность обслуживания (среднее число обслуживаний в единицу времени); $t_{обсл}$ – среднее время обслуживания заявки.

Пусть S – множество состояний системы и $P(l, t+\tau / i, t)$ – вероятность того, что система, находившаяся в момент t в состоянии i , в момент $t+\tau$ окажется в состоянии l. Для марковской системы (она привлекает нас отсутствием последействия) можно записать *уравнения Чепмена – Колмогорова*:

$$P(l, t+\tau / i, t) = \sum_{j \in S} P(j, t+\tau^* / i, t)\ P(l, t+\tau / j, t+\tau^*). \qquad (4)$$

Если под состояниями понимать число заявок, то эти уравнения можно записать в виде:

$$P_k(t+\tau) = \sum_{i+j=k} P_i(t)\,P_j(\tau)\ . \qquad (5)$$

Рассмотрим *случай разомкнутой системы* с простейшим входным потоком интенсивности λ и одним каналом обслуживания с интенсивностью μ.

Возьмем интервал времени $[\,t,\ t+dt]$. В силу разомкнутости системы множество состояний системы

$$S = \{\, S_0\, , S_1\, , S_2\, ,\ \ldots\, , S_k\, , S_{k+1}\, ,\ \ldots \},$$

где S_k — состояние, когда в системе находится k заявок.

Попробуем оценить вероятности перехода между состояниями с учетом, того, что вероятность появления заявки в этом интервале времени равна $\lambda\, dt$ и вероятность завершения обслуживания предшествующей заявки равна $\mu\, dt$.

Очевидно, что вероятность перехода $S_0 \to S_1$ равна $\lambda\, dt$ и вероятность перехода $S_1 \to S_0$ равна $1 - \lambda\, dt$. Если в системе присутствовали $k>0$ заявок (состояние S_k), то для перехода в состояние S_{k-1} необходимо, чтобы заявка была обслужена и не поступило новой заявки; отсюда вероятность перехода $S_k \to S_{k-1}$ равна $\mu\, dt\,(\,1 - \lambda\, dt\,) \cong \mu\, dt$. Для перехода из состояния S_k в состояние S_{k+1} необходимо, чтобы поступила новая заявка, но ни одна из ранее поступивших не была обслужена: вероятность перехода $S_k \to S_{k+1}$ равна $\lambda\, dt \cdot (1 - \mu\, dt) \cong \lambda\, dt$. Вероятность для системы остаться в том же состоянии составит $1 - (\lambda + \mu)dt$.

Тогда из (5) имеем

$$P_0\,(t+dt) = (1 - \lambda\, dt)\, P_0\,(t) + \mu\, dt\, P_1(t)\ ,$$

$$P_k(t+dt) = \lambda\, dt\, P_{k-1}(t) + (1 - \lambda\, dt - \mu\, dt)\, P_k\,(t) + \mu\, dt\, P_{k+1}(t)\ ,\quad k > 0\ .$$

При предельном переходе при $dt \to 0,$ получаем систему обыкновенных дифференциальных уравнений, для описания состояний СМО:

$$\begin{aligned} &\frac{d}{dt} P_0(t) = -\lambda\ P_0(t) + \mu\ P_1(t)\ , \\ &\frac{d}{dt} P_k(t) = \lambda\ P_{k-1}(t) - (\,\lambda + \mu\,)\, P_k(t) + \mu\, P_{k+1}(t)\, , k > 0 \end{aligned}\ . \qquad (6)$$

Решение (6) при заданных начальных условиях для непрофессионала в численном анализе может оказаться затруднительным (воспользоваться преобразованием Лапласа или прибегнуть к численному решению задачи Коши для системы обыкновенных дифференциальных уравнений большого порядка с помощью компьютерных библиотек).

Если ограничиться рассмотрением *установившегося* режима, признаком которого является существование предела

$$\lim_{t\to\infty} P_k(t) = P_k\,, \quad k = 0\,,1\,,2\,,\ldots\,, \tag{7}$$

система (6) приведется к бесконечной системе линейных алгебраических уравнений с трехдиагональной матрицей коэффициентов

$$\begin{aligned} &\lambda P_0 = \mu P_1 \\ &(\lambda + \mu)\,P_k = \lambda P_{k-1} + \mu P_{k+1}\,, \quad k = 1\,,2\,,\ldots \end{aligned} \tag{8}$$

Обозначив

$$\rho = \lambda/\mu\,, \tag{9}$$

имеем $P_1=\rho P_0$, $P_2=\rho^2 P_0$, $P_3=\rho^3 P_0$, ... , $P_k=\rho^k P_0$, ... , откуда с учетом $P_0 + P_1 + P_2 + P_3 + \ldots + P_k + \ldots = 1$ получаем при $\rho < 1$

$$P_0(1+\rho+\rho^2+\rho^3+\ldots+\rho^k+\ldots) = P_0\frac{1}{1-\rho} = 1\,.$$

Тогда

$$P_0 = 1-\rho \tag{10}$$

и

$$P_k = \rho^k P_0 \text{ для } \quad k = 1\,,2\,,\ldots \tag{11}$$

Обратите внимание на требование $\rho < 1$. Если это требование нарушено, то ни о каком установившемся режиме не может быть речи: очередь растет неограниченно (средняя продолжительность обслуживания больше среднего интервала времени между заявками).

Теперь обратимся к аналогичной *замкнутой системе* с числом заявок, не превышающим n . Здесь система уравнений (6) приведется к конечной системе

$$\begin{aligned} &\frac{d}{dt}P_0(t) = -\lambda\ P_0(t) + \mu\ P_1(t)\,, \\ &\frac{d}{dt}P_k(t) = \lambda\ P_{k-1}(t) - (\lambda+\mu)\,P_k(t) + \mu P_{k+1}(t)\,, k = 1,\ldots,n-1 \\ &\frac{d}{dt}P_n(t) = -\mu\ P_n(t) + \lambda\ P_{n-1}(t), \end{aligned} \tag{12}$$

которая для установившегося режима дает конечную систему линейных алгебраических уравнений

$$\begin{aligned} &\lambda P_0 = \mu P_1, \\ &(\lambda + \mu)\,P_k = \lambda P_{k-1} + \mu P_{k+1}\,, \quad k = 1\,,2\,,\ldots\,,n-1\,; \\ &\lambda P_{n-1} = \mu P_n\,. \end{aligned} \tag{13}$$

Решение этой системы $P_1=\rho P_0$, $P_2=\rho^2 P_0/2$, $P_3=\rho^3 P_0/(2\cdot 3)$, $P_4=\rho^4 P_0/(2\cdot 3\cdot 4)$, ... , $P_n=\rho^n P_0/\,n!$ дает

$$P_0 = \left[\sum_{k=0}^{n}\frac{\rho^k}{k!}\right]^{-1} \tag{14}$$

и
$$P_k = P_0\,\rho^k/\,k\,! \quad \text{для} \quad k = 1\,,2\,,\ldots\,,n\,. \tag{15}$$

Полученные выше решения можно обобщить на случай многоканальных систем с ограниченным ожиданием. Так, если СМО имеет N однотипных каналов обслуживания (интенсивность обслуживания равна $N{\cdot}\mu$), m мест в очереди и к тому же число n возможных заявок превышает $N{+}m$ (в противном случае нет проблем), возникает система

$$\frac{d}{dt}P_0(t) = -\lambda\ P_0(t) + \mu\ P_1(t);$$

$$\frac{d}{dt}P_k(t) = \lambda\ P_{k-1}(t) - (\lambda + k\mu)\, P_k(t) + (k+1)\mu\, P_{k+1}(t),\ k = 1,\ldots,N-1;$$

$$\frac{d}{dt}P_k(t) = \lambda\ P_{k-1}(t) - (\lambda + N\mu)\, P_k(t) + N\,\mu\ P_{k+1}(t),\ k = N,\ldots,N+m-1;$$

$$\frac{d}{dt}P_{N+m}(t) = -N\mu\ P_{N+m}(t) + \lambda\ P_{N+m-1}(t),$$

из которой получаем для установившегося режима

$$\lambda P_0 = \mu P_1,$$

$$(\lambda + k\mu)\, P_k = \lambda P_{k-1} + (k+1)\mu P_{k+1},\quad k = 1, 2, \ldots, N-1; \tag{16}$$

$$(\lambda + N\mu)\, P_k = \lambda P_{k-1} + N\mu P_{k+1},\quad k = N, N+1, \ldots, N+m-1;$$

$$\lambda P_{N+m-1} = N\mu P_{N+m}\ .$$

Решение этой системы дает

$$P_k = P_0\, \rho^k / k! \quad \text{для} \quad k = 1, 2, \ldots, N-1, \tag{17}$$

$$P_k = P_0\, \rho^k / (N^{k-N} N!) \quad \text{для} \quad k = N, N+1, \ldots, N+m, \tag{18}$$

$$P_0 = \left[\sum_{k=0}^{N} \frac{\rho^k}{k!} + \frac{\rho^{N+1}}{N \cdot N!} \cdot \frac{(\rho/N)^m - 1}{(\rho/N) - 1} \right]^{-1}. \tag{19}$$

Умение найти значения P_k дает возможность отыскать и ряд основных характеристик СМО.

12.3. Основные характеристики С М О

Значение P_0 определяет вероятность того, что все каналы обслуживания свободны (находятся в состоянии простоя) .

Значение P_k определяет вероятность того, что в системе (в очереди и на обслуживании) находятся k заявок. Если k не превышает числа каналов N, то все заявки находятся на обслуживании и очередь отсутствует; в противном случае все каналы заняты и $k{-}N$ заявок находится в очереди.

Вероятность $P_{отк}$ отказа в обслуживании определяется ситуацией занятости всех N каналов и всех m мест в очереди и равна P_{N+m} .

Среднее число занятых каналов $N_{зан}$ определяется математическим ожиданием дискретной случайной величины [31]:

$$N_{зан} = \sum_{k=1}^{N} k \cdot P_k + \sum_{k=N+1}^{N+m} N \cdot P_k = \rho \cdot \left[1 - \frac{\rho^{N+m}}{N! \cdot N^m} P_0 \right] \qquad (20)$$

(мы опускаем здесь достаточно простые преобразования).

Среднее число свободных каналов

$$N_{своб} = N - N_{зан}. \qquad (21)$$

Коэффициент простоя каналов

$$K_{прост} = N_{своб} / N. \qquad (22)$$

Коэффициент занятости каналов

$$K_{занят} = N_{зан} / N. \qquad (23)$$

Относительная пропускная способность (доля обслуженных заявок в общем числе поступавших в систему) определяется величиной

$$q = 1 - P_{отк}. \qquad (24)$$

Абсолютная пропускная способность (среднее число заявок, обслуживаемых в единицу времени) определяется величиной

$$A = \lambda \cdot q. \qquad (25)$$

Средняя длина очереди [30]

$$L_{очер} = \sum_{k=N+1}^{N+m} (k-N) P_k = \frac{\rho^{N+1}}{N! \cdot N} \cdot \frac{1 - (\rho/N)^m (m + 1 - m\rho/N)}{(1 - \rho/N)^2} P_0. \qquad (26)$$

Среднее число заявок, находящихся в системе, складывается из средних значений занятости каналов и длины очереди

$$L = N_{зан} + L_{очер}. \qquad (27)$$

Среднее время пребывания заявки в очереди равно

$$T_{очер} = L_{очер} / \lambda. \qquad (28)$$

Общее время пребывания заявки в очереди будет складываться из $T_{очер}$ и среднего времени обслуживания

$$T_{сист} = T_{очер} + q / \mu. \qquad (29)$$

Полученные характеристики дают возможность анализа замкнутых и разомкнутых систем с отказами (m=0), с очередью или с ожиданием ($m \to \infty$) при простейшем входном потоке и однотипных параллельных каналах обслуживания с показательным законом длительности обслуживания (в частности, с фиксированной длительностью).

12.4. Примеры систем с ограниченной очередью

Пример 1. Пусть на аэродром самолеты прибывают с интенсивностью 27 самолетов в час, время приземления составляет 2 минуты, допускается нахождение над аэродромом не более m = 10 самолетов. Нужно определить число N посадочных полос, гарантирующее вероятность отказа, не превышающую 0.05, и среднее время ожидания, не превышающее 5 минут.

Здесь λ=27 , μ = 30 , $\rho = \lambda / \mu = 0.9$.

Вероятность простоя диспетчеров службы посадки согласно (19) :

$$P_0 = \left[\sum_{k=0}^{N} \frac{0.9^{\,k}}{k!} + \frac{0.9^{\,N+1}}{N \cdot N!} \cdot \frac{(0.9/N)^{10} - 1}{(0.9/N) - 1} \right]^{-1}$$

Вероятность отказа в посадке равна

$$P_{отк} = P_0 \cdot 0.9^{\,N+10} / (\, N^{10} \cdot N\,!)\,.$$

Среднее время ожидания в воздухе согласно (28) и (26)

$$T_{очер} = L_{очер} / \lambda\,,$$

где

$$L_{очер} = \sum_{k=N+1}^{N+m} (k-N) \cdot P_k = \frac{\rho^{N+1}}{N!\cdot N} \cdot \frac{1-(\rho/N)^m (m+1-m\rho/N)}{(1-\rho/N)^2} \cdot P_0 =$$

$$= \frac{0.9^{N+1}}{N!\,N} \cdot \frac{1-(0.9/N)^{10}(11-10\cdot 0.9/N)}{(1-0.9/N)^2} \cdot P_0 \quad .$$

Выполняя арифметические действия при N=1 , обнаруживаем, что

$$P_0 \cong 0.14; \quad P_{отк} \cong 0.04; \quad L_{очер} \cong 0.045; \; T_{очер} \cong 0.9 \text{ мин}$$

и что одной посадочной полосы при указанных условиях вполне достаточно.

Пример 2. Пусть имеются станки, которые могут выходить из строя с частотой в среднем 2 раза за смену. Продолжительность ремонта одним оператором составляет около трех часов (оператор одновременно может ремонтировать лишь один станок и не переходит к другому, не отремонтировав предыдущий). Хотелось бы определить число операторов, при котором потери от простоя станков и оплаты лишнего числа операторов были бы минимальны.

Такую замкнутую систему можно представить системой с N каналами (операторами) и очередью с m местами ожидания (совпадает с числом станков). Если известны потери C_n от простоя станка в течение часа и оплата C_p часа работы оператора, то при семичасовой смене задача сводится к нахождению значения N , которое минимизировало бы значение

$$C_n \cdot T_{очер} + C_p \cdot 7N \;,$$

где $T_{очер}$ определяется (26) и (28) при $\lambda = 2/7$, μ=1/3, $\rho = \lambda/\mu = 6/7$.

Можно привести множество подобных задач для определения числа кассиров в универмаге, наилучшего с позиций минимума потерянных покупателей, для определения числа бригад грузчиков на железнодорожной станции, минимизирующего штраф за простой вагонов, для определения числа полос движения на проектируемой автомагистрали и т. п.

12.5. Дисциплина ожидания и приоритеты

Выше мы рассматривали простейший поток однотипных заявок с дисциплиной выборки на обслуживание в порядке поступления.

Можно показать, что и в случае случайного выбора на обслуживание полученные выше оценки не претерпят изменения, но их дисперсия (разброс относительно ожидаемой величины) возрастет. Очевидно, что среднее время сидения в очереди не изменится от того, что кто-то пройдет без очереди, но для отдельных клиентов время ожидания увеличится. Так отношение дисперсий времени ожидания в неупорядоченной и упорядоченной очереди имеет порядок $(2+\rho) / (2-\rho)$, где $\rho = \lambda/\mu$ (мы обычно предпочитаем систему с жесткой дисциплиной обслуживания из-за предсказуемости ее поведения и всякое «возмущение» в ее работе отрицательно действует на нашу психику).

Существует множество систем, в которых присутствует $N>1$ входных потоков с различной интенсивностью λ_i ($i = 1, .., N$), время обслуживания заявок которых распределено по показательному закону с параметрами μ_i.

Здесь при условии пуассоновости входных потоков можно считать, что суммарный поток будет пуассоновским с интенсивностью $\Lambda=\sum\lambda_i$; функция распределения времени обслуживания заявок суммарного потока в одноканальной системе

$$S(t) = \sum_{i=1}^{N} \frac{\lambda_i}{\Lambda}(1-e^{-\mu_i t});$$

среднее время ожидания определяется формулой Полачека – Хинчина:

$$W = \frac{\Lambda/2 \cdot \int_0^\infty t^2 dS(t)}{1-R_N}, \quad R_N = \Lambda \int_0^\infty t\, dS(t) \quad ,$$

которая для данного случая дает $R_N = \sum [\, \lambda_i / \mu_i] = \sum\rho_i$ и в случае стационарности режима ($R_N<1$)

$$W = \frac{\sum_{i=1}^{N} \rho_i/\mu_i}{1-R_N} \quad .$$

Определенный интерес представляют системы, где каждому входному потоку сопоставлено целое число k – показатель приоритета потока (наивысший приоритет определяется k=1).

Если обслуживание заявки не прерывается ни при каких условиях и выбор на обслуживание происходит с учетом приоритета (при одинаковом приоритете выбирается первый пришедший в систему), то такая система называется *системой с относительными приоритетами.*

Можно показать [31], что поскольку время ожидания заявки с приоритетом k складывается из времени завершения обработки требования, вошедшего в канал, времени обслуживания ранее поступивших требований приоритета от 1 до k–1 и ранее поступивших требований с приоритетом k, то его среднее значение равно

$$W_k = \frac{\sum_{i=1}^{N} \rho_i / \mu_i}{(1-R_k)(1-R_{k-1})} \quad ; \quad R_k = \sum_{i=1}^{k} \rho_i \quad (k=1..N) \ .$$

На этой основе можно определить среднюю длину очереди заявок k -го приоритета $L_k = \lambda_k W_k$ и среднее число таких заявок в системе $L_k+\rho_k$. Показано [37], что введение приоритетов улучшает функционирование системы, если более высокое преимущество присваивается заявкам с меньшей длительностью обслуживания. Если учитывать стоимостные характеристики, то более высокое преимущество предоставляется заявкам с большим значением $C_k \cdot \mu_k$, где C_k - средняя стоимость ожидания.

Существуют системы с абсолютными приоритетами, где появление заявки более высокого уровня прерывает обслуживание текущей заявки, которая вернется в очередь и потом снова поступит на обслуживание с места прерывания (или с начала). Здесь среднее время ожидания заявки с приоритетом k

$$W_k = \frac{\sum_{i=1}^{N} \rho_i / \mu_i}{(1-R_k)(1-R_{k-1})} + \frac{R_{k-1}}{\mu_k (1-R_{k-1})} \quad ; \quad R_k = \sum_{i=1}^{k} \rho_i \quad (k = 1 \ldots N) \ .$$

В [30] рекомендуется для минимизации затрат на пребывание заявок в очереди в системах с относительными и абсолютными приоритетами, равных

$$\Phi = \sum_{k=1}^{N} \alpha_k \lambda_k W_k \ ,$$

где α_k – издержки на ожидание заявки k -го приоритета в единицу времени, более высокий приоритет давать заявкам с наибольшим значением $\alpha_k \mu_k$.

Исключительно сложно установить разумные приоритеты в случае многофазных систем, где заявка проходит обслуживание в нескольких последовательных подсистемах [37, 38]. Здесь относительно простые выводы удается сделать лишь для случая двух подсистем, и для получения выводов для более сложных систем приходится прибегать к моделированию.

12.6. Моделирование систем массового обслуживания и метод Монте-Карло

До сих пор мы рассматривали системы, для которых удавалось описать результаты обследования в аналитическом виде. Однако, многие реальные системы (многофазные, с оригинальными приоритетами, с признаками нестационарности, с непуассоновскими входными потоками, с непоказательным распределением длительности обслуживания и т. д.) не поддаются такому решению.

Суть математического моделирования системы заключается в следующем.

Время функционирования системы разделяется на достаточно большое количество подинтервалов (единиц времени, в течение которых не может возникнуть более одной заявки или завершиться выполнение более одной заявки). Для каждого такого подынтервала последовательно моделируется факт появления новой заявки (да/нет), проверяется наличие свободного канала (закончено ли обслуживание какой-то заявки) и загрузка его заявкой из очереди, проверяется наличие мест в очереди с последующим выводом (принять в очередь/отказать в обслуживании) и т. д. При этом фиксируется число отказов, время ожидания заявок в очереди и в системе вообще, число заявок в очереди в каждый момент и другие значения, которые позволяют найти вероятность отказа, распределение времени ожидания и среднее время, вероятность простоя каналов и т. п. Для надежности выводов такое разовое моделирование повторяется достаточно много раз.

Очевидно, что ни о каком ручном моделировании не может быть речи (объем работы здесь слишком велик для нормального индивида). Здесь приходится использовать метод статистических испытаний (Монте-Карло), предполагающий наличие компьютера с датчиком псевдослучайных чисел, равномерно распределенных в интервале от 0 до 1. Мы уже отмечали выше, что псевдослучайные числа получаются по какому-либо алгоритму, но в совокупности подчиняются всем законам проверки на случайность (мы не останавливаемся на методах их получения, так как есть отличные программные датчики во всех системах программирования).

В процессе моделирования часто возникает необходимость генерации случайных чисел с законом распределения, отличным от вышеуказанного.

Пусть R – случайные числа с равномерным законом распределения в (0,1) и X – создаваемые случайные числа с плотностью распределения $p(X)$. Между ними можно установить соотношение [39]

$$R = \int_{-\infty}^{x} p(X)\, dX \quad .$$

Так для показательного распределения $p(X) = \lambda\, exp(-\lambda x)$ для $X>0$ легко установить, что $X = -\, ln(1-R)/\lambda$. Для равномерного распределения в интервале (a, b) с очевидностью – $X=a+(b-a)R$.

Получение дискретных случайных чисел сводится к поиску наименьшего значения X, при котором

$$R \leq e^{-\lambda} \sum_{k=0}^{x} \lambda^{k} / k! \quad .$$

Если взятие интеграла и представление X через R составит затруднение, можно воспользоваться методом Неймана.

Здесь при неограниченности области значений X усекаем ее до некоторого интервала $[a,b]$; например для нормального распределения концы интервала берем отстоящими от среднего на 3–4 стандартных отклонения. Затем генерируется пара случайных чисел R_1 и R_2; если $R_2 \leq (b-a)\cdot p[a+(b-a)R_1]/ p^*$, где $p^* = max[(b-a)\cdot p(x)]$, то берем $X = a+(b-a)R_2$ и в противном случае берем следующую пару случайных чисел.

Таким путем мы можем моделировать интервалы времени между заявками входного потока, продолжительность обслуживания заявки, вероятность выхода канала из строя и т.п.

Вопрос о числе N отдельных реализаций системы мы уже обсуждали при рассмотрении методов нелинейного программирования.

Мы прибегаем к методам Монте-Карло там, где другие методы терпят фиаско (моделирование сложных систем массового обслуживания, реальных технологических процессов, больших систем сетевого планирования, вычисление интегралов кратности 10 и выше, поиск экстремумов «нехороших» функций с большим числом переменных и др.).

Уже давно существуют многочисленные примеры успешного моделирования вполне реальных СМО (см. библиографию [37]). Сегодня такое моделирование стало рутинным методом прогноза в разнообразных сферах – от космонавтики до сельского хозяйства.

ЗАКЛЮЧЕНИЕ

Завершая обзор методов исследования операций, мы должны отметить, что он далек от полноты. Если на методах линейного программирования мы остановились достаточно подробно, лишь опустив те подходы, которыми пользовались в детские годы ЭВМ ради экономии нескольких ячеек памяти или нескольких миллисекунд, то о методах нелинейного программирования ограничились, как говорится, «верхушкой айсберга». Мы уделили значительное внимание динамическому программированию, знакомство с идеологией которого, по нашему мнению, способствует формированию предусмотрительного, ответственного специалиста, от решений которого в будущем может оказаться зависимой судьба людей. Мы достаточно поверхностно, в рамках общих представлений, дали минимальный обзор подходов к анализу стохастических систем принятия решений и систем массового обслуживания (этой тематике посвящены тысячи серьезных публикаций, доступных лишь специалистам в сфере случайных процессов). По аналогичным причинам мы оставили «за бортом» многие результаты, полученные за последние 30–40 лет в теории игр. Мы не затронули и значительный практический опыт исследования в последнее пятидесятилетие.

Некоторые современные авторы настаивают на включении в «Исследование операций» таких изящных математических дисциплин как теория графов и вариационное исчисление. Есть опасность предложения объять под этим термином всю прикладную математику, вплоть до уравнений математической физики.

Читатель может заметить, что приведенный ниже список литературы преимущественно содержит ссылки на издания 60–70-х годов. Это не только дань уважения автора к тем статьям и книгам, по которым он учился сам и рекомендовал своим первым слушателям полвека назад. Дело в том, что идеология методов исследования операций заложена именно в эти годы в книгах Д. Данцига и С. Гасса, Р. Беллмана, С. Дрейфуса и Р. Ховарда, Л. Канторовича, Е. Г. Гольштейна и Д. Б. Юдина, Л. Форда и Д. Фалкерсона, Г. Вагнера и Х. Таха и др.

Некоторые из этих изданий устарели из-за многословия (попыток объяснять серьезные вопросы математически безграмотному читателю) и неспособности увидеть компьютерные перспективы, но в большинстве остаются непревзойденными по изяществу изложения, по математической культуре в сочетании с доступностью для непрофессионала – немногочисленная учебная литература последних лет, за ничтожным исключением, излагает результаты отцов и дедов, не всегда отдавая им должное.

В последние десятилетия существенно расширились возможности исследования операций не столько за счет создания новых методов, сколько за счет фантастического увеличения возможностей компьютера. Так целочисленная линейная программа, на решение которой автору 40 лет назад потребовалось до трех часов машинного времени, сейчас решается в считанные минуты.

Соответственно, появилось достаточно много компьютерных разработок, позволяющих пользователю, не слишком искушенному в области математики и не умеющему писать программы даже на уровне обычных алгоритмических языков, успешно решать задачи исследования операций. Например, в популярной среде электронных таблиц Excel предусмотрена стандартная процедура «Поиск решения» [42], позволяющая достаточно просто записать ограничения на значения переменных (ячеек) и потребовать найти сочетание этих значений так, чтобы содержимое т.н. целевой ячейки приняло максимальное, минимальное или конкретное значение. Для реализации такого поиска можно выбрать итерационный подход (модифицированный метод Ньютона или метод сопряженных градиентов) и задать начальные приближения для искомых переменных [43]. Очевидно, что успех в подобном решении зависит от удачного выбора начального приближения и для задач большой размерности чаще всего получается сообщение о том, что решение с заданной точностью за указанное число шагов оказалось недостижимым. Тем не менее современные компьютерные пакеты существенно облегчают поиск оптимальных решений.

Не сбылись мечты поколения 60-х о «безлюдных шахтах» или сельскохозяйственных комбайнах, «гуляющих сами по себе», хотя уже сегодня роботы заменили человека во многих сферах, подчас даже в сфере принятия решений. Уходит время (к сожалению, не столь быстро), когда можно было управлять без серьезного (количественного!) анализа на основе волюнтаризма личности или толпы. Требования «максимального удовлетворения жизненных потребностей трудящихся в текущем квартале» или «поддержки отечественного производителя», не поверенные числом, были и остаются уделом демагогов и фигляров.

Очевидно, что исследование операций и математическое моделирование не подменяют *опытного здравомыслящего* руководителя, думающего о завтрашнем дне, а создают базу для размышлений, способствуют минимизации ошибок, весьма дорого обходящихся рядовому человеку.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Данциг, Дж. Б. Воспоминания о началах линейного программирования. 1994 / пер. с англ. В. Зацепина [Электронный ресурс: index/ html http://webcenter.ru/~zwb/origins. htm#1#1].

2. Данциг, Д. Линейное программирование, его применения и обобщения. – М. : Прогресс, 1966.

3. Вершик**, А. М.** О Л. В. Канторовиче и линейном программировании [Электронный ресурс: http://www.mmonline.ru/articles.php?mid =1893&topic=207]

4. Канторович, Л. В. Математические методы организации и планирования производства. – Л. : Изд-во ЛГУ, 1939 (В сб. "Применение математики в экономических исследованиях". – М. : Соцэкгиз, 1959).

5. Гасс С. Линейное программирование. – М. : Физматгиз, 1961.

6. Гасс С. Путешествие в Страну Линейного Программирования / пер. с англ. Ю. Н. Сударева; предисл. Ю. В. Овсиенко. – М. : Мир, 1971.

7. Гейл, Д. Теория линейных экономических моделей. – М. : ИИЛ, 1963.

8. Канторович, Л.В. Математическое оптимальное программирование в экономике./ Л. В. Канторович, А. Б. Горстко. – М. : Знание, 1968.

9. Гольштейн, Е. Г. Линейное программирование / Е. Г. Гольштейн, Д. Б. Юдин. – М. : Наука, 1969.

10. Гольштейн, Е.Г. Новые направления в линейном программировании. / Е. Г. Гольштейн, Д. Б. Юдин. – М. : Сов. радио, 1966.

11. Рубинштейн, Г. Ш. Двойственность в математическом программировании и некоторые вопросы выпуклого анализа // Успехи математических наук. – 1970. – Т. 25. – № 5.

12. Бирман, С. Оптимальное планирование. – М. : Экономика, 1968.

13. Форд, Л.. Потоки в сетях / Л. Форд, Д. Фалкерсон. –М. : Мир, 1966.

14. Ху, Т. Целочисленное программирование и потоки в сетях. – М. : Мир, 1974.

15. Прим, Р. К. Кратчайшие связывающие сети и некоторые обобщения // Кибернетический сборник, 2. – М. : ИИЛ, 1961.

16. Берж, К. Теория графов и ее применения. – М. : ИИЛ, 1962.

17. Хедли, Д. Нелинейное и динамическое программирование. – М. : Мир, 1967.

18. Кюнци, Г. П. Нелинейное программирование / Г.П. Кюнци, В. Крелле. – М. : Сов. радио, 1965.

19. Беллман, Р. Динамическое программирование.– М. : ИИЛ, 1960.

20. Беллман, Р. Прикладные задачи динамического программирования. / Р. Беллман, С. Дрейфус. – М. : Наука, 1965.

21. Ховард, Р.. Динамическое программирование и марковские процессы. – М. : Сов. радио, 1964.

22 Калихман, И. Динамическое программирование в примерах и задачах. / И. Калихман, М. Войтенко. – М. : Высш. школа, 1979.

23. Даффин, Р. Геометрическое программирование. / Р. Даффин, Э. Питерсон, К. Зенер. – М.: Мир, 1972.

24. Бекишев, Г.А. Элементарное введение в геометрическое программирование / Г.А. Бекишев, М. И. Кратко. – М. : Наука, 1980.

25. Вагнер, Г. Основы исследования операций. – М. : Мир, 1973.

26. Таха, Х. Введение в исследование операций: в 2-х кн. – М. : Мир, 1985.

27. Моудер, Д. Исследование операций. Т.1 / Д. Моудер, С. Элмаграби. – М. : Мир, 1981.

28. Нейман, Дж. Теория игр и экономическое поведение / Дж. Нейман, О. Моргенштерн. – М.: ИИЛ, 1970. –708 с.

29. Карлин, С. Математические методы в теории игр, программировании и экономике. – М.: Мир, 1964.

30. Льюс, Р. Игры и решения. / Р. Льюс, Г. Райфа. – М. : ИИЛ, 1960.

31. Костевич, Л. С. Теория игр. Исследование операций. / Л. С. Костевич, А. А. Лапко – Минск : Вышэйшая школа, 1982.

32. Зуховицкий, С.И. Математические методы сетевого планирования. /С. И. Зуховицкий, А. И. Радчик. – М. : Наука, 1965.

33. Кофман, А. Сетевые методы планирования. / А. Кофман, Г. Дебазей. – М. : Прогресс, 1968.

34. Абрамов, С.А. Сетевые методы планирования и управления.. / С.А.Абрамов и др. – М. : Сов. радио, 1965.

35. Саати, Т.Л. Элементы теории массового обслуживания и ее приложения. – М. : Сов. радио, 1965.

36. Хинчин, А. Я. Работы по теории массового обслуживания. – М. : Физматгиз, 1963.

37. Кофман, А. Массовое обслуживание (теория и приложения). / А. Кофман, Р. Крюон. – М. : Мир, 1965 .

38. Кокс, Д. Теория очередей / Д. Кокс, У. Смит. – М. : Мир, 1966.

39. Бусленко, Н.П. Метод статистических испытаний (Монте-Карло) / Н.П. Бусленко, Ю. А. Шрейдер. – М. : Физматгиз, 1961.

40. Тынкевич, М.А. Экономико-математические методы (исследо-

вание операций): учеб. пособие. – 2-е изд., испр. и доп. – Кемерово, 2000 [Электронный ресурс: http:/vtit/kuzstu/ru/books/shelf/book1/ index/html].

41. Тынкевич, М. А. Введение в теорию графов. – Кемерово, 2002 [Электронный ресурс: http:/vtit/kuzstu/ru/books/shelf/book1/ index/html].

42. Терпугов, А.Ф. Экономико-математические модели: учеб. пособие. – Барнаул: Алтайск. экон.-юрид. ин-т, 1999.

43. Курицкий, Б. Поиск оптимальных решений средствами Excel 7.0 в примерах. – Спб. : 1997.

44. Акулич, И.Л. Математическое программирование в примерах и задачах. – М. : Высш. шк., 1996 [Электронный ресурс: http://listlib.narod.ru /vichteh/aAkulich.html].

45. Дегтярев, Ю. И. Исследование операций. – М. : Высш. шк., 1986.

46. Кудрявцев, Е.М.. Исследование операций в задачах, алгоритмах и программах. – М. : Радио и связь, 1984.

47. Экономико-математические методы и модели: учеб. пособие / под ред А. В. Кузнецова. – Минск : БГЭУ, 2000.

48. Афанасьев, М. Ю. Исследование операций в экономике: модели, задачи, решения: учеб. пособие. / М. Ю. Афанасьев, Б. П. Суворов. – М. : ИНТРА-М, 2003.

49. Афанасьев, М. Ю. Прикладные задачи исследования операций: учеб. пособие /М. Ю. Афанасьев, Б. П. Суворов. –М. : ИНТРА-М, 2006.

50. Кремер, Н. Ш. (ред.). Исследование операций в экономике. – М. : ЮНИТИ-ДАНА , 2004.

51. Зайченко, Ю.П. Исследование операций. – Киев, 2002–2004 [Электронный ресурс http://iasa.org.ua/iso?lang=rus&ch=0]

52. Иглин, С. П. Математические расчеты на базе MatLab. – СПб. : БХВ-Петербург, 2005.

53. Аллен, Р. Математическая экономия. – М. : Изд-во иностр. лит., 1963.

54. Интрилигатор, М. Математические методы оптимизации и экономическая теория. – М. : Прогресс, 1975.

55. Ланкастер, К. Математическая экономика. – М. : Сов. радио, 1972.

ПРИЛОЖЕНИЕ
Создатели методов исследования операций

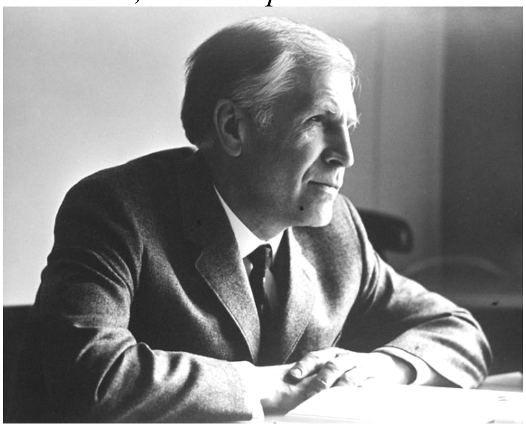
Джон фон Нейман
(*1903-1957*)

Леонид Витальевич Канторович
(*1912 – 1986*)

Джордж Б. Данциг
(*1914 —2005*)

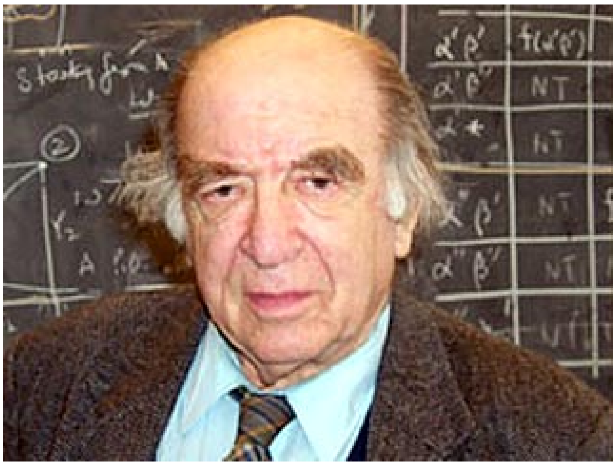
Вручение Л.В. Канторовичу диплома и Нобелевской медали королем Швеции Карлом XVI, 10 декабря 1975 года

Лауреат Нобелевской премии в области экономики за 1972 год - Кеннет Джозеф Эрроу (*род.1921г.*)



Академик АН СССР
Лев Семенович Понтрягин
(*1908-1988*)

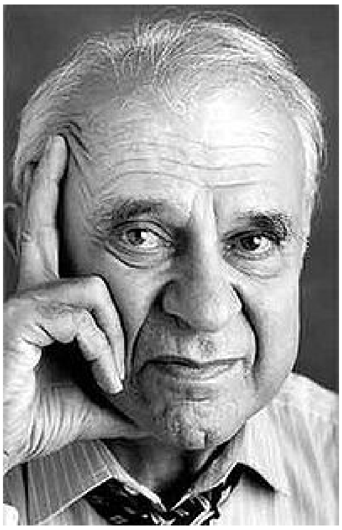
Геннадий Шлемович Рубинштейн (*1923–2004*)



Лауреат Нобелевской премии в области экономики - Леонид Гурвиц (*1917-2008*)

Ричард Э.Беллман (*1920 – 1984*)

Ральф Э. Гомори (*род.1929*)



Лауреат Нобелевской премии в области экономики за 1973 год Василий Васильевич Леонтьев (*1905-1999*)

Заслуженный деятель науки РФ Евгений Г. Гольштейн (ЦЭМИ РАН)

О Г Л А В Л Е Н И Е

Тынкевич Моисей Аронович

Экономико-математические методы (Исследование операций) Учебное пособие

Редактор З. М. Савина

Подписано в печать 31.05.2011. Формат 60×84/16
Бумага офсетная. Отпечатано на ризографе
Уч.-изд. л. 13,00. Усл. печ. л. 12,79
Тираж 300 экз. Заказ № 15

Кузбасский государственный технический университет
650000, Кемерово, ул. Весенняя, 28
Типография Кузбасского государственного технического университета
650000, Кемерово, ул. Д. Бедного, 4а

Тынкевич Моисей Аронович

Родился 28 февраля 1937 года в г. Новосибирске. Окончив в 1954 г. среднюю школу №24 г. Кемерово с серебряной медалью, поступил на механико-математический факультет Томского университета и в 1959 г. окончил его среди первых за Уралом 24 выпускников по новой специальности «Вычислительная математика».

В 1959-1966 гг. работал на кафедре вычислительной математики Томского университета, вел курс программирования для ЭВМ того времени («Стрела», «Урал», «М-20») и подготовил первое учебное пособие по программированию для ЭВМ М-20 (изд. ТГУ, 1965), руководил дипломными работами и практикой первых сибирских математиков-вычислителей, вел первые спецкурсы по исследованию операций.

С 1966 г. – старший преподаватель кафедры экономики и организации производства Кузбасского политехнического института, с 1969 г. старший преподаватель новой в вузе кафедры вычислительной техники и промэлектроники. Читал курсы программирования и экономико-математических методов, в начале 70-х принимал участие в подготовке экономистов по специализации «Математическое обеспечение АСУ».

В 1974 г. в Совете Томского университета защитил диссертацию на тему «Количественный анализ электромагнитных полей в движущихся средах при сосредоточенных источниках», в 1976 г. утвержден ВАКом в ученой степени кандидата физико-математических наук и в 1978 г. в звании доцента.

В настоящее время профессор кафедры вычислительной техники и информационных технологий.

Почетный работник высшего образования России (1997). Награжден медалью «За особый вклад в развитие Кузбасса» (2001), медалью «За достойное воспитание детей» (2010).

Ведет занятия по курсам:
1. Численные методы анализа.
2. Исследование операций в экономике.
3. Экономико-математические методы и модели.

Публикации: более 75 научных работ и учебных пособий, нескольких циклов методических разработок.