

Министерство сельского хозяйства Российской Федерации
ФГБОУ ВПО «Красноярский государственный аграрный университет»

С.Ю. Журавлев

МЕТОДЫ ОПТИМИЗАЦИИ

Рекомендовано научно-методическим советом федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Красноярский государственный аграрный университет» для внутривузовского использования в качестве учебно-методического пособия для студентов, обучающихся по направлению подготовки 110800.62 «Агроинженерия»

Красноярск 2014

ББК 40.7
Ж 91

Рецензенты:

В.А. Терсков, д-р техн. наук, проф. каф. математики и информатики
Красноярского института железнодорожного транспорта
(филиала ИрГУПС)

Н.В. Петровский, канд. техн. наук, директор ГНУ КНИИСХ
Россельхозакадемии

Ж 91 **Журавлев, С.Ю.**
Методы оптимизации: учеб.-метод. пособие / С.Ю. Журавлев;
Краснояр. гос. аграр. ун-т. – Красноярск, 2014. – 132 с.

Дается описание решения различных оптимизационных задач с использованием классических методов оптимизации.

Предназначено для студентов, обучающихся по направлению 110800.62 «Агроинженерия», профили: «Технологическое оборудование для хранения и переработки сельскохозяйственных продуктов», «Технические системы в агробизнесе», «Технический сервис в агропромышленном комплексе».

ББК 40.7

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
КЛАССИФИКАЦИЯ ЗАДАЧ ОПТИМИЗАЦИИ.....	6
Практическая работа № 1. ЧИСЛЕННЫЕ МЕТОДЫ ПОИСКА БЕЗУСЛОВНОГО ЭКСТРЕМУМА.....	11
Практическая работа № 2. КЛАССИЧЕСКИЕ МЕТОДЫ БЕЗУСЛОВНОЙ ОПТИМИЗАЦИИ.....	36
Практическая работа № 3. МЕТОДЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ В ОБЛАСТИ НАХОЖДЕНИЯ ОПТИМАЛЬНЫХ РЕШЕНИЙ ОПРЕДЕЛЕННЫХ ПРАКТИЧЕСКИХ ЗАДАЧ.....	73
ЗАКЛЮЧЕНИЕ.....	130
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	131

ВВЕДЕНИЕ

Учитывая прикладную направленность подготовки специалистов в технических вузах, основное внимание в учебно-методическом пособии уделено практическим вопросам оптимизации. Как показывает практика компьютерного моделирования и решения реальных оптимизационных задач, обычно представленные в учебной литературе универсальные «библиотечные» методы нелинейного программирования часто оказываются малоэффективными и требуют адаптации при практической реализации. Кроме того, важнейшее значение приобретают элементы сопутствующих вычислительных технологий, связанные с формализацией конкретной прикладной проблемы и разработкой сценария оптимизации. Для изучения соответствующих вопросов, обычно выпадающих из стандартных учебных курсов, целесообразно излагать теорию и методы оптимизации «на фоне» некоторой предметной области, а не в абстрактном виде. Это позволяет, с одной стороны, глубже понять возникающие проблемы, а с другой – получить наглядные и конкретные интерпретации применяемых технологий.

Понятно, что под эту простую схему подпадают многие ситуации, как связанные с нашей повседневной жизнью, так и возникающие внутри технических, экономических, программных, эко- и других систем. Обычно речь идет о некотором наилучшем в определенном смысле, т. е. «оптимальном», управлении, что и определяет необходимость создания теории управления, базирующейся на принципах оптимальности. Помимо обеспечения оптимальности самого управляющего воздействия, алгоритмизация и компьютерная реализация большинства этапов процесса управления сводятся к решению целой цепочки различных оптимизационных задач. В этом смысле можно утверждать, что теория конечномерной (параметрической) оптимизации является основным алгоритмическим фундаментом современной теории управления.

Для полного усвоения представленных в книге материалов предполагается знакомство с курсами высшей математики (содержащими, в частности, достаточно полные сведения по линейной и матричной алгебре), численного анализа, теории вероятностей (в частности, корреляционной теории) и основ теории управления.

Если вы владеете этим материалом в стандартном для технических вузов объеме, то никаких проблем возникнуть не должно. Для облегчения использования пособия и согласования применяемых терминов и обозначений даны необходимые для данного курса общие сведения из теории оптимизации и сформулированы характерные постановки оптимизационных задач.

Целью учебно-методического пособия является приобретение студентами знаний и навыков в области применения различных математических инструментов нахождения оптимальных или рациональных подходов к принятию управляющих воздействий и механизмов на основе математических моделей оптимизации.

В пособии приведены теоретические основы и методические указания для выполнения практических работ по методике решения различных задач в области оптимизации.

КЛАССИФИКАЦИЯ ЗАДАЧ ОПТИМИЗАЦИИ

Математическая модель задачи оптимизации может быть представлена следующим образом:

$$F(X) \rightarrow \underset{X \in D \subseteq S}{\text{opt}} .$$

$F = \{f_1, f_2, \dots, f_k\}$ – оптимизируемая функция (целевая функция, целевой функционал, критерий качества и т. п.), *численно* выражает степень достижения целей функционирования оптимизируемого объекта.

$X = (x_1, x_2, \dots, x_n)^T$ – вектор независимых переменных, его компоненты – неизвестные задачи оптимизации (переменные оптимизации) – являются *управляемыми* входами объекта оптимизации.

D – множество допустимых значений неизвестных, определяемое налагаемыми на неизвестные ограничениями (допустимая область, допустимое множество).

S – пространство оптимизации.

$$\min F(X) = -\max[-F(X)]$$

$$F = \{f_1, f_2, \dots, f_k\},$$

$k = 1$ – задача *однокритериальной, или скалярной* оптимизации,
 k – количество критериев оптимизации;

$k > 1$ – задача *многокритериальной, или векторной* оптимизации.

$$F(X) \rightarrow \underset{X \in D \subseteq S}{\text{opt}}$$

$$X = (x_1, x_2, \dots, x_n)^T,$$

$n = 1$ – задача *одномерной* (иногда *линейной*) оптимизации,
 n – количество переменных, входящих в целевую функцию;

$n > 1$ – задача *многомерной* оптимизации, или *задача* оптимизации со *многими переменными*.

Задачи оптимизации исходя из допустимой области

$D = S$ – задача *безусловной* оптимизации, или *задача оптимизации без ограничений* (какие-либо ограничения на неизвестные отсутствуют).

$D \subset S$ – задача *условной* оптимизации, или *задача с ограничениями* (т. е. в задаче не все значения переменных допустимы).

Задачи оптимизации исходя из пространства оптимизации

$S = \mathbf{R}^n$ – задача оптимизации с *непрерывными переменными*.

$S = \mathbf{Z}^n$ – задача *целочисленной оптимизации*.

$S = \mathbf{B}^n$ – задача *булевой оптимизации* (частный случай задачи целочисленной оптимизации, при которой переменные могут принимать только два значения – ноль и единица). Если при этом $f(X)$ принимает значения из \mathbf{R}^n , то – задача *псевдобулевой оптимизации*.

Если значение целевой функции зависит от некоторых комбинаций объектов из конечного набора, их размещения или способа упорядочения, то такие задачи называются задачами *комбинаторной оптимизации*.

Задачи целочисленной и комбинаторной оптимизации объединяются понятием задач *дискретной оптимизации*.

В задачах *смешанной оптимизации* могут одновременно присутствовать переменные нескольких или даже всех типов (наиболее известный частный случай – задачи *смешанного целочисленного* программирования с целочисленными и непрерывными переменными).

Задачи оптимизации, определяемые свойствами целевой функции

Целевая функция имеет более одного локального экстремума – задача *глобальной, или многоэкстремальной* оптимизации (если требуется найти все локальные экстремумы или наилучший из них).

В задачах *локальной* оптимизации требуется найти один локальный экстремум (единственный для одноэкстремальной целевой функции или любой для многоэкстремальной).

Целевая функция и/или функции, описывающие ограничения, заданы не аналитически (в виде компьютерных программ, имитационных моделей, человеко-машинных процедур или как выход реальной системы) – это задачи *оптимизации с неявными функциями* (поисковые задачи оптимизации).

Все функции, входящие в постановку задачи, записываются в явном аналитическом виде – это задача *математического программирования*.

Общая формулировка задачи математического программирования:

$$f(X) \rightarrow \min_x (\max)$$

при ограничениях

$$h_i(X) = 0, i = 1, \dots, m,$$

$$g_i(X) \geq 0, i = m + 1, \dots, p.$$

Все функции, входящие в постановку, являются непрерывно дифференцируемыми – это задача *дифференцируемой оптимизации*, иначе – задача *недифференцируемой оптимизации*.

Целевая функция выпукла, функции-ограничения образуют выпуклую допустимую область – это задача *выпуклой оптимизации*.

Целевая функция сепарабельна, ограничения линейны – задача *сепарабельного программирования*.

Целевая функция квадратичная, ограничения линейны – задача *квадратичного программирования*.

Все функции общего вида – общая задача *нелинейного программирования*.

Целевая функция и функции-ограничения являются линейными относительно независимых переменных – задача *линейного программирования*.

Более узкие постановки задачи линейного программирования – *транспортная задача*, *задача о назначениях*, *задача целочисленного линейного программирования* и т. п.

Параметрические задачи оптимизации (параметрическое программирование) – функции и коэффициенты, входящие в постановку задачи, зависят от некоторого параметра или параметров.

Вся исходная информация задачи оптимизации определена однозначно – *детерминированные задачи оптимизации*.

Все или некоторые параметры модели носят вероятностный характер – *принятие решения в условиях риска (стохастические задачи, стохастическое программирование, стохастическая аппроксимация)*.

Неопределенность данных имеет не вероятностный характер – *оптимизация в условиях неопределенности (нечеткое математическое программирование, или нечеткая оптимизация)*.

Конечномерные (математическое программирование) и бесконечномерные (вариационное исчисление)

На плоскости xOy даны две точки $A_1(x_1, y_1)$ и $A_2(x_2, y_2)$. Найти кривую кратчайшей длины, соединяющую эти точки.

$$S = \int_{x_1}^{x_2} \sqrt{1 + y'^2} dx \quad ,$$
$$y(x_1) = y_1, \quad y(x_2) = y_2.$$

Статические и динамические (оптимальное управление)

Задача оптимального управления

Дана система, поведение которой описывается дифференциальным уравнением:

$$\frac{dx}{dt} = f(x, u, t),$$

где x – вектор фазовых координат; u – вектор управления; t – время. На вектора x и u наложены ограничения: $x \in X$, $u \in U$. Система рассматривается на интервале $t \in [0, T]$.

Требуется определить вектор-функции $u(t)$ и $x(t)$, доставляющие минимум функционалу $J = J(x, u)$ при переводе из начального состояния $(x(0), 0)$ в конечное состояние $(x(T), T)$.

$$J(x, u) = \int_0^T F(x, u, t) dt,$$

где F – дифференцируемая функция своих аргументов.

Классификация методов оптимизации

Один из способов классификации методов оптимизации состоит в соотнесении их оптимизационным задачам, для решения которых они предназначены.

По типу информации о производных, требуемой для организации процесса оптимизации, методы подразделяются:

- на методы нулевого порядка, требующие только вычислений значений функции в точках пространства оптимизации и не требующие аналитического вида производных;
- методы первого порядка (градиентные), требующие кроме значений функции в точке еще и аналитическое задание производных первого порядка для вычисления градиента;
- методы второго порядка (ньютоновские), для работы которых требуются еще и производные второго порядка.

Другая классификация:

- методы *прямого поиска*;
- методы *линейной аппроксимации*;
- методы *квадратичной аппроксимации*.

По степени математической обоснованности методы делят на *эвристические* и *рациональные*.

Методы оптимизации подразделяют на *детерминированные* и *стохастические*. Стохастические алгоритмы используют элементы случайности при выборе направления или длины шага в процессе оптимизации.

Практическая работа № 1

ЧИСЛЕННЫЕ МЕТОДЫ ПОИСКА БЕЗУСЛОВНОГО ЭКСТРЕМУМА

Цель работы: изучить теоретические основы и получить навыки решения задач оптимизации с помощью численных или итерационных методов.

Итерационные методы оптимизации функций одной переменной

Методы деления интервала

С помощью численных (итерационных) методов можно, например, определять минимум функции в некотором интервале $a < x < b$, в котором, как предполагается, лежит точка минимума. При этом может вычисляться только значение функции в выбранных точках данного интервала (т. е. не используется производная). Такой подход имеет важные приложения на практике, когда может быть, например, неизвестен явный вид функции. Методами поиска определяется достаточно малый интервал, в котором находится минимум, при этом осуществляется наименьшее количество вычислений функции (так как затраты на вычисления могут быть весьма велики).

Для всех методов, приводимых ниже, предполагается, что функция унимодальна на выделенном интервале, т. е. имеет только одну точку минимума. Прежде чем применять эти методы, необходимо выделить интервал, содержащий искомую точку минимума, и убедиться, что она там единственная. В противном случае предлагаемые методы будут определять только один локальный минимум.

Для определения интервала, содержащего точку минимума, обычно используют следующую процедуру. Выбирают две стартовые точки x и y такие, что $y = x + s$. Затем, если $f(x) > f(y)$, определяют следующие точки:

$$x^{k+1} = x^k + s$$

до тех пор, пока не будет получено $f(x^k) < f(x^{k+1})$. В этом случае полученный интервал $[x, x^{k+1}]$ «накрывает» искомую точку минимума.

Если же $f(x) < f(y)$, то выбирается противоположное направление. Точки строятся по правилу

$$x^{k+1} = x^k - s$$

до тех пор, пока не будет получено $f(x^k) < f(x^{k+1})$.

Основной проблемой при применении описанной процедуры является правильный выбор величины s . Во многих прикладных задачах переменная изменяется в пределах многих степеней десятки (например, от 0,001 до 10000). Тогда при неправильном выборе величины s минимизация может потребовать слишком больших затрат (при очень малом s – на «накрытие» точки минимума, а при большом – на последующее за «накрытием» сужение отрезка до требуемой точности).

Для того чтобы преодолеть эту проблему, был предложен подход с удвоением шага. В этом случае определение верхней границы интервала осуществляется по правилу

$$x^{k+1} = x^0 + 2^k s.$$

После того, как точка минимума была «накрыта», нижняя граница интервала определяется тем же самым процессом, но с изменением знака перед s и в обратном направлении. Иногда такой процесс используют не только для «накрытия» точки минимума, но и для ее отыскания.

И последнее, что следует здесь учесть – это возможность того, что целевая функция вообще является постоянной. Для учета этого обстоятельства необходимо вводить максимальную длину шага, которая не должна быть превышена в процессе определения отрезка, содержащего точку минимума.

Метод бисекций

Этот метод позволяет на каждом шаге уменьшать вдвое длину интервала, содержащего минимум, причем значения целевой функции вычисляются с этой целью в двух определенных точках. Если выполнено n вычислений целевой функции, то длину интервала можно уменьшить в $2(n - 3)/2$ раза.

Сначала разбивается исходный интервал (a^0, b^0) . Берется середина интервала

$$c^0 = (a^0 + b^0)/2,$$

затем две точки

$$e^0 = (a^0 + c^0)/2 \text{ и } d^0 = (c^0 + b^0)/2,$$

что дает пять точек, удаленных друг от друга на одно и то же расстояние:

$$\delta^0 = (b^0 - a^0)/4.$$

С учетом унимодальности легко видеть, что всегда можно исключить два из четырех подынтервалов (поскольку точка минимума не может в них содержаться) и что остаются только два смежных подынтервала (a^1, c^1) и (c^1, b^1) .

Таким образом, все сводится к той же задаче, но уже на отрезке (a^1, b^1) , длина которого в два раза меньше, чем у исходного.

На следующем шаге понадобится только два дополнительных вычисления целевой функции в точках $e^1 = (a^1 + c^1)/2$ и $d^1 = (c^1 + b^1)/2$ и т. д.

Глобальная сходимость метода бисекций с очевидностью следует из самого определения этого метода.

Метод Фибоначчи

Пусть нужно определить минимум с наименьшим возможным интервалом неопределенности (точность), но при этом можно выполнить только n вычислений функции. Как следует выбрать точки, в которых вычисляется функция?

Очевидно, что надо сделать так, чтобы значения функции, полученные в предыдущих экспериментах, определяли положение последующих точек.

Предположим, что имеется интервал неопределенности (x_1, x_3) и известно значение функции $f(x_2)$ внутри этого интервала. Если можно вычислить значение функции всего один раз в точке x_4 , то где следует поместить точку x_4 , чтобы получить наименьший возможный интервал неопределенности?

Пусть $x_2 - x_1 = L$, $x_3 - x_2 = R$, $L > R$. Эти значения фиксированы, если известны x_1, x_2, x_3 .

Если x_4 находится в интервале (x_1, x_2) , то:

1) если $f(x_4) < f(x_2)$, то новым интервалом неопределенности будет (x_1, x_2) длиной $x_2 - x_1 = L$;

2) если $f(x_4) > f(x_2)$, то новым интервалом неопределенности будет (x_4, x_3) длиной $x_3 - x_4$.

Поскольку неизвестно, какая из этих ситуаций будет иметь место, выберем x_4 таким образом, чтобы минимизировать наибольшую из длин $x_3 - x_4$ и $x_2 - x_1$. Достигается это тем, что x_4 помещается в (x_1, x_3) симметрично относительно x_4 . В этом случае $x_3 - x_4 = x_2 - x_1$. В любом другом случае размещения x_4 может быть так, что полученный интервал будет больше L .

Если окажется, что можно выполнить еще одно вычисление, то следует применить описанную процедуру к интервалу (x_1, x_2) или (x_3, x_4) . Последовательное применение такой процедуры и образует итерационный процесс приближения к точке минимума, называемый поиском Фибоначчи.

Числа Фибоначчи определяются следующим образом:

$$F_0 = 1, F_1 = 1, F_k = F_{k-1} + F_{k-2}, \\ k = 2, 3, \dots,$$

то есть $(1, 1, 2, 3, 5, 8, 13, 21, \dots)$.

Если начальный интервал (a, b) имеет длину $L = b - a$, ε – точность вычисления значения функции, то конечный интервал неопределенности будет равен

$$L_n = \frac{L_1}{F_n} + \varepsilon \frac{F_{n-2}}{F_n},$$

т. е. уменьшится в F_n раз (пренебрегая ε), и это наилучший гарантированный результат.

Отметим, что в начале поиска, когда известны только a и b , необходимо правильно установить первую внутреннюю точку. Она размещается на расстоянии L_2 от любого конца интервала:

$$L_2 = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n \varepsilon}{F_n},$$

т. е. тоже определяется с помощью чисел Фибоначчи.

Метод «золотого сечения»

Не всегда можно заранее определить, сколько раз придется вычислять функцию. В методе Фибоначчи это нужно знать для определения положения начальной точки.

Метод «золотого сечения» почти столь же эффективен, как и метод Фибоначчи, однако при этом не требуется знать n – количество вычислений функции, определяемое вначале. В этом методе интервал неопределенности на каждом шаге делится на две части так, что отношение целого к большей части равно отношению большей части к меньшей, т. е. равно так называемому «золотому сечению»:

$$\tau = \frac{(1 + \sqrt{5})}{2} \approx 1.618033989.$$

Поиск методом «золотого сечения» является предельным случаем поиска Фибоначчи и при достаточно больших n дает результат, лишь на 17 % худший. Зато не требует начального задания количества вычислений функции.

Интерполяционные методы

Рассмотренные выше методы позволяли найти малый интервал, в котором находится минимум функции, причем от самой функции мало что зависело, ее свойства вообще не принимались во внимание. Это выглядит довольно странным, что и приводит к идее о том, что возможен и другой подход: используется несколько значений функции в определенных точках для аппроксимации функции обычным полиномом, по крайней мере, в небольшой области. Затем положение минимума функции аппроксимируется положением минимума полинома, поскольку последний вычислить проще. При этом характер поведения оптимизируемой функции учитывается при выборе вида полинома и при его построении, в чем и состоит отличие от методов деления интервала.

Метод квадратичной интерполяции

Если известны значения функции $f(x)$ в трех точках α , β , γ , равные f_α , f_β , f_γ , то функция $f(x)$ может быть аппроксимирована квадратичной функцией $\varphi(x) = Ax^2 + Bx + C$, где A , B , C определяются из системы уравнений:

$$A\alpha^2 + B\alpha + C = f_\alpha,$$

$$A\beta^2 + B\beta + C = f_\beta,$$

$$A\gamma^2 + B\gamma + C = f_\gamma,$$

т. е. из системы трех *линейных* уравнений с тремя неизвестными.

Так как квадратичная функция $\varphi(x)$ имеет минимум в точке $-B/2A$, если $A > 0$, то точка минимума $f(x)$ может быть аппроксимирована значением

$$\delta = \frac{1}{2} \frac{(\beta^2 - \gamma^2)f_\alpha + (\gamma^2 - \alpha^2)f_\beta + (\alpha^2 - \beta^2)f_\gamma}{(\beta - \gamma)f_\alpha + (\gamma - \alpha)f_\beta + (\alpha - \beta)f_\gamma}.$$

Опишем алгоритм. Пусть $f(x)$ – унимодальная функция одной переменной, A – начальная аппроксимация положения минимума, длина шага H является величиной того же порядка, что и расстояние от точки A до истинного минимума (условие, которое не просто удовлетворить).

Алгоритм квадратичной интерполяции:

1. Вычислить $f(A)$, $f(A + H)$.
2. Если $f(A) < f(A + H)$, то взять в качестве третьей точки $(A - H)$ и вычислить $f(A - H)$. В противном случае взять точку $(A + 2H)$ и вычислить $f(A + 2H)$.
3. Используя эти три точки, найти аппроксимацию точки минимума δ и вычислить $f(\delta)$.
4. Если разность между $f(\delta)$ и предыдущим наименьшим значением функции меньше заданной точности, то процедура заканчивается.
5. Если процедура не завершилась на шаге 4, то точка с наибольшим значением отбрасывается и выполняется переход на шаг 3. Но если, оставив точку с наибольшим значением, мы определим конечные границы интервала, в котором лежит минимум, то следует действительно оставить эту точку, отбросить другую и вернуться на шаг 3.

В примере на рисунке 1 оставляем x_1 , x_2 , x_3 , а не x_2 , x_3 , x_4 , хотя x_1 и дает наибольшее значение функции.

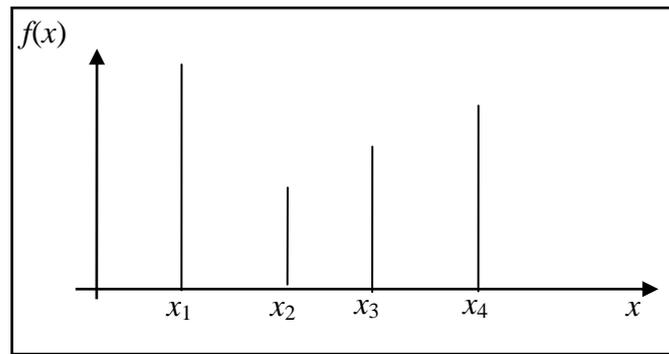


Рисунок 1 – Выбор границ интервала

Описанная процедура квадратичной интерполяции называется методом Пауэлла и является частным (но наиболее часто используемым) случаем так называемого интерполяционного метода Лагранжа, который предусматривает аппроксимацию целевой функции полиномами любого порядка (а не только второго). Наиболее эффективна эта процедура при оптимизации квадратичных или близких к ним функций. В других случаях этот метод может быть не столь эффективным.

Для оптимизации неквадратичных функций предложена модификация этого метода, состоящая в том, что для интерполяции выбираются не три, а *четыре* точки. При этом две внутренние точки используются по очереди с крайними точками для построения *двух* аппроксимаций точки минимума, выполняемых по алгоритму Пауэлла. Затем их взвешенная средняя выбирается для новой итерации. Такая процедура повышает надежность определения точки минимума для функций, не являющихся квадратичными.

Метод кубической интерполяции

При решении реальных задач редко приходится иметь дело с функциями одной переменной. Однако методы одномерной оптимизации важны при многомерной оптимизации, которая выполняется в основном по следующему правилу: зафиксировать некоторую точку, выбрать подходящее направление, выполнить одномерную оптимизацию из выбранной точки в выбранном направлении. Поэтому методы интерполяции полезны для выполнения линейного поиска при оптимизации функций многих переменных. В этом случае обычно требуется найти минимум функции $f(x)$ в точках прямой x_2, x_3, x_4 , где x_0 – заданная точка, \bar{d} – определенное направление. Значение функции $f(X_0 + \lambda\bar{d})$ на этой прямой являются значениями функции одной переменной λ : $\varphi(\lambda) = f(X_0 + \lambda\bar{d})$.

Рассмотрим метод Дэвидона, в котором функция аппроксимируется кубическим полиномом и который обеспечивает большую точность определения точки минимума. Обсуждение метода проведем в виде, пригодном для многомерной оптимизации.

Для интерполяции используются значения функции и ее производной, вычисленные в двух точках.

Пусть минимизируется функция $f(x)$ на прямой $X_0 + \lambda \bar{d}$, т. е. функция

$$\varphi(h) = f(X_0 + h\bar{d}) = f(x_1^0 + hd_1, x_2^0 + hd_2, \dots, x_n^0 + hd_n).$$

$$\frac{\partial \varphi}{\partial h} = \nabla f(X_0 + h\bar{d})^T \bar{d} = g(X_0 + h\bar{d})^T \bar{d}.$$

Предполагаем, что известны следующие значения:

$$\varphi(p) = \varphi_p, \varphi(q) = \varphi_q, \frac{\partial \varphi}{\partial h}(p) = G_p, \frac{\partial \varphi}{\partial h}(q) = G_q.$$

Эту информацию можно использовать для построения кубического полинома $a + bh + ch^2 + dh^3$, который и будет аппроксимировать функцию $\varphi(h)$.

Если выбрать $p = 0$ (для простоты), то уравнения, определяющие a, b, c, d , выглядят так:

$$\begin{cases} a = \varphi_p, \\ a + bq + cq^2 + dq^3 = \varphi_q, \\ b = G_p, \\ b + 2cq + 3dq^2 = G_q. \end{cases}$$

Эти уравнения имеют решение:

$$a = \varphi_p, b = G_p, c = -(G_p + Z)/q, d = (G_p + G_q + 2Z)/3q, \\ Z = \frac{3(\varphi_p - \varphi_q)}{q} + G_p + G_q.$$

Стационарные точки кубического полинома являются решением уравнения

$$G_p - 2(G_p + Z)\frac{h}{q} + (G_p + G_q + 2Z)\left[\frac{h}{q}\right]^2 = 0.$$

Точкой минимума кубического полинома является

$$r = q \left[\frac{Z + W - G_p}{G_q - G_p + 2W} \right], \quad W = [Z^2 - G_p G_q]^{\frac{1}{2}}. \quad (1)$$

Как уже отмечалось, это все выполняется при $p = 0$. Выбор q осуществляется из следующих соображений.

Если $G_p < 0$, то следует выбирать $q > 0$, т. е. сделать шаг в направлении убывания функции $\varphi(h)$. В противном случае $q < 0$. Значение q должно быть таким, чтобы интервал $(0, q)$ содержал минимум. Это будет верно, если $\varphi(q) > \varphi(p)$ или $G_q > 0$ (даже при $\varphi(q) < \varphi(p)$). Если ни одно из этих условий не выполнено, то надо удвоить q , повторяя это до тех пор, пока полученный интервал не будет содержать минимум.

Остается задача определения начальной величины q . Значение q , которое было бы одинаково приемлемо для всех задач, выбрать весьма сложно. Дэвидон, Флетчер и Пауэлл предложили выбирать q следующим образом:

$$q = \min \left\{ \eta, -\frac{2(\varphi_p - \varphi_m)}{G_p} \right\},$$

где φ_m – оценка наименьшего значения $\varphi(h)$, $\eta = \text{const}$ (2 или 1).

Алгоритм кубической интерполяции:

1. Найти $\varphi_p = f(X_0)$ и $G_p = (g(X_0))^T \bar{d}$.
2. Проверить, выполняется ли условие $G_p < 0$. Если не выполняется, производить поиск в направлении $-\bar{d}$, иначе – в направлении \bar{d} . Выбрать q в соответствии с рекомендациями (при этом необходимо «угадать» φ_m).
3. Вычислить $\varphi_q = f(X_0 + q\bar{d})$.
4. Если $G_q > 0$ или $\varphi_q > \varphi_p$, то интервал, содержащий минимум, найден. Иначе положить $q = 2q$ и перейти к шагу 3.
5. Использовать выражение (1) для аппроксимации точки минимума на интервале $(0, q)$ значением r .

6. Если $\left| \frac{\partial \varphi}{\partial h} \right| = \left| [g(X_0 + h\bar{d})]^T \bar{d} \right| = |G_0| < \varepsilon$, где ε – заданная точность,

то остановиться.

7. Вернуться на шаг 5, используя интервал $(0, r)$, если $G_r > 0$, либо используя интервал (r, q) , если $G_r < 0$.

Рассмотренный метод кубической интерполяции является одним из интерполяционных методов Эрмита, приспособленным специально для оптимизации. Этот метод требует знания производных функций и используется в основном в так называемых *квазиньютоновских* методах многомерной оптимизации, которые и сами по себе используют производные.

Итерационный метод

Хорошо известный *метод секущих* отыскания нуля функции первоначально назывался методом *regula falsi*, или *regula falsorum*. Этот метод правильно предсказывает положение нуля функции, если она зависит линейно от своего аргумента. Если даны две точки a и b и значения функции в них, то формула предсказания положения нуля функции имеет следующий вид:

$$c = a - f(a) \frac{b - a}{f(b) - f(a)}.$$

Так как процесс отыскания точки минимума целевой функции можно интерпретировать как отыскание нуля ее производной, то метод секущих может быть легко использован для целей оптимизации. Формула предсказания положения нуля производной (т. е. точки минимума) принимает вид

$$c = a - f'(a) \frac{b - a}{f'(b) - f'(a)}.$$

Если $f'(c) \neq 0$, то процедура может быть продолжена итеративно с использованием уменьшенного интервала $[a, c]$, если $f'(c)$ и $f'(b)$ имеют одинаковый знак, или $[c, b]$, если $f'(c)$ и $f'(a)$ имеют одинаковый знак. Если $f'(a)$ и $f'(b)$ имеют одинаковый знак, то точка c должна лежать вне отрезка $[a, b]$. Если $f'(c)$ имеет опять тот же знак, то c заменяет ту точку, которая дает наибольшее значение производной. Если $f'(c)$ имеет другой знак, то можно продолжать процедуру *regula falsi*

итеративно до тех пор, пока не будет получена точка с нулевой (или достаточно малой) производной.

Метод работает надежно только в случае, если функция почти линейна или хотя бы строго выпукла. Кроме того, стартовые точки должны быть выбраны в окрестности точки минимума. Иначе процедура может выдать в качестве ответа точку максимума, так как при движении к нему производная тоже будет уменьшаться.

Итерационный метод Ньютона-Рафсона

Интерполяционная формула Ньютона для улучшения приближенного решения уравнения $f(x)=0$

$$x_{k+1} = x_k - f(x_k)/f'(x_k)$$

использует только одну точку, но требует оценки значения функции и ее производной. Если функция линейна по x , то ноль функции будет указан точно, в противном случае будет получено только улучшенное значение, и процесс должен быть повторен. Как и в предыдущем случае, эта рекуррентная формула может быть применена для поиска нуля производной, т. е. для определения точки минимума самой функции. Соответствующая итерационная формула, называемая также *правилом Ньютона-Рафсона*, имеет вид

$$x_{k+1} = x_k - f'(x_k)/f''(x_k).$$

Если целевая функция не является квадратичной, то понадобится несколько итераций до тех пор, пока не будет удовлетворен критерий останова.

Существенным преимуществом алгоритма является то, что требуется только одна точка, и то, что для квадратичной функции достаточно одной итерации. Однако имеются и недостатки:

- если первая и вторая производные оцениваются приближенно по конечным разностям, то эффективность алгоритма снижается не только из-за ошибок округления, но и из-за неточности аппроксимации. Это особенно верно в окрестности точки минимума, где первая производная становится существенно малой;

- не различаются точки минимумов, максимумов и перегибов. Уже стартовая точка должна выбираться как можно ближе к искомому минимуму;

– если целевая функция имеет порядок выше второго, то итерации метода Ньютона-Рафсона могут не сойтись. Условием сходимости является положительность второй производной для всех точек, встречающихся в процессе оптимизации.

Вообще говоря, надежность и скорость сходимости всех интерполяционных алгоритмов существенно зависит от того, насколько целевая функция соответствует модельной функции, положенной в основу метода. В некоторых случаях методы деления интервала оказываются лучше, чем интерполяционные. Например, поиск Фибоначчи эффективнее кубической интерполяции в случае логарифмической функции и в случае полинома высокой степени. Поэтому разрабатываются новые и более сложные схемы одномерной оптимизации, базирующиеся, в основном, на всевозможных комбинациях интерполяционных подходов и методов деления отрезка.

Применение необходимых и достаточных условий безусловного экстремума эффективно для решения ограниченного числа задач. Для решения большинства практических задач они не могут быть рекомендованы по следующим причинам: целевая функция может не иметь непрерывных производных, использование необходимого условия связано с решением системы нелинейных алгебраических уравнений, возможны случаи, когда целевая функция задана неявно.

Численные методы оптимизации относятся к классу итерационных методов, порождающих последовательность точек в соответствии с предписанным набором правил.

Рассмотрим методы одномерной оптимизации. Возьмем унимодальную функцию $f(x)$ на интервале $[a_0, b_0]$. Последовательная стратегия поиска минимума этой функции заключается в выборе начального интервала неопределенности, уменьшения интервала неопределенности, проверке условия окончания.

Решение задач оптимизации методом Фибоначчи

Метод Фибоначчи применяется для поиска безусловного экстремума функции одной переменной. Для построения эффективного метода одномерной оптимизации, работающего по принципу последовательного сокращения интервала неопределенности, следует задать правило выбора на каждом шаге двух внутренних точек. Желательно, чтобы одна из них всегда использовалась в качестве внутренней и для следующего интервала. Тогда количество вычислений функции сократится вдвое.

В методе Фибоначчи реализована стратегия, обеспечивающая максимальное сокращение интервала неопределенности при заданном количестве вычислений функции. Эта стратегия опирается на числа Фибоначчи, которые определяются по формуле

$$F_0 = F_1 = 1, F_k = F_{k-1} + F_{k+1}, k = 1, 2, \dots$$

Выпишем несколько первых чисел 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89,

Точки вычисления функции находятся с использованием чисел Фибоначчи, количество вычислений функции задается. Поиск заканчивается, когда длина текущего интервала неопределенности оказывается меньше установленной величины.

Алгоритм:

Шаг 1. Задать начальный интервал неопределенности $L_0 = [a_0, b_0]$, допустимую длину конечного интервала l .

Шаг 2. Найти N из условия $F_N \geq \frac{L}{l}$ и числа Фибоначчи $F_0, F_1, \dots, F_N, k = 0$.

Шаг 3. Вычислить $y_0 = a_0 + \frac{F_{N-2}}{F_N}(b_0 - a_0), z_0 = a_0 + \frac{F_{N-1}}{F_N}(b_0 - a_0)$.

Шаг 4. Вычислить $f(y_k), f(z_k)$.

Шаг 5. Сравнить $f(y_k)$ и $f(z_k)$.

Если $f(y_k) \leq f(z_k)$, то положить

$$a_{k+1} = a_k, b_{k+1} = z_k, z_{k+1} = y_k, y_{r+1} = a_{k+1} + \frac{F_{N-k-3}}{F_{N-k-1}}(b_{k+1} - a_{r+1}).$$

Перейти к шагу 6.

Если $f(y_k) > f(z_k)$, то положить

$$a_{k+1} = y_k, b_{k+1} = b_k, y_{k+1} = z_k, z_{k+1} = a_{k+1} + \frac{F_{N-k-2}}{F_{N-k-1}}(b_{k+1} - a_{k+1}).$$

Шаг 6. Проверить условие окончания.

Пример 1.

Методом Фибоначчи найти минимум функции $f(x) = 2x^2 - e^x$.

Количество вычислений функции – 10 (должны получить 10 интервалов), первоначальный отрезок неопределенности $[0, 1]$.

$$y_0 = a_0 + \frac{F_{N-2}}{F_N} (b_0 - a_0), \quad N = 10, [0,1], \quad z_0 = a_0 + \frac{F_{N-1}}{F_N} (b_0 - a_0).$$

Числа Фибоначчи:

F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
1	1	2	3	5	8	13	21	34	55	89

$$a_0 = 0, b_0 = 1,$$

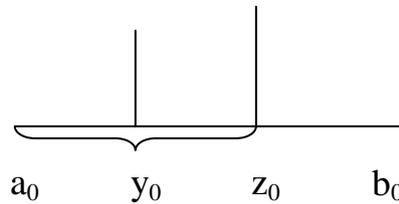
$$y_0 = 0 + \frac{F_8}{F_{10}} (1 - 0) = \frac{F_8}{F_{10}} = \frac{34}{89} = 0,382,$$

$$z_0 = 0 + \frac{F_9}{F_{10}} (1 - 0) = \frac{F_9}{F_{10}} = \frac{55}{89} = 0,617,$$

$$f(y_0) = 2(0,382)^2 - e^{0,382} = 0,291 - 1,465 = -1,174,$$

$$f(z_0) = 2(0,617)^2 - e^{0,617} = 0,761 - 1,853 = -1,092,$$

$$[0; 0,617].$$



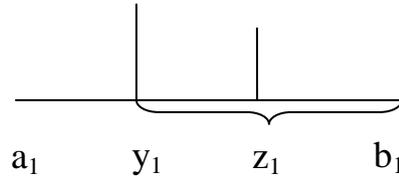
$$a_1 = a_0 = 0, b_1 = z_0 = 0,617, z_1 = y_0 = 0,382,$$

$$y_1 = a_1 + \frac{F_{N-3}}{F_{N-1}} (b_1 - a_1) = 0 + \frac{F_7}{F_9} (0,617 - 0) = \frac{21}{55} (0,617) = 0,235,$$

$$f(y_1) = 2(0,235)^2 - e^{0,235} = 0,110 - 1,264 = -1,154,$$

$$f(z_1) = f(y_0) = -1,174,$$

$$[0,235; 0,617].$$



$$a_2 = 0,235, \quad b_2 = 0,617,$$

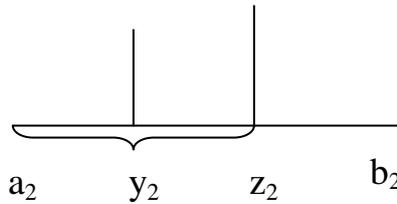
$$y_2 = z_1 = 0,382,$$

$$\begin{aligned} z_2 &= a_2 + \frac{F_{N-3}}{F_{N-2}} (b_2 - a_2) = 0,235 + \frac{F_7}{F_8} (0,617 - 0,235) = 0,235 + \frac{21}{34} (0,382) = \\ &= 0,235 + 0,235 = 0,470, \end{aligned}$$

$$f(z_2) = f(0,470) = 2(0,47)^2 - e^{0,47} = 0,4418 - 1,599 = -1,158,$$

$$f(y_2) = f(0,382) = -1,174,$$

$$[0,235; 0,470].$$



$$a_3 = 0,235, \quad b_3 = 0,470,$$

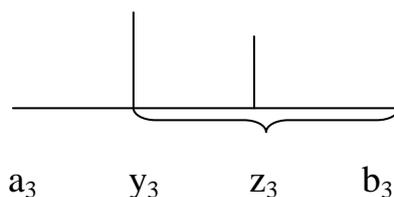
$$z_3 = y_2 = 0,382,$$

$$\begin{aligned} y_3 &= a_3 + \frac{F_{N-5}}{F_{N-3}} (b_3 - a_3) = 0,235 + \frac{F_5}{F_7} (0,470 - 0,235) = 0,235 + \frac{8}{21} (0,235) = \\ &= 0,235 + 0,089 = 0,324, \end{aligned}$$

$$f(y_3) = 2(0,324)^2 - e^{0,324} = 0,209 - 1,382 = -1,173,$$

$$f(z_3) = f(y_2) = -1,174,$$

$$[0,324; 0,470].$$



$$a_4 = 0,234, \quad b_4 = 0,470,$$

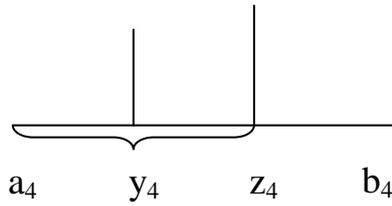
$$y_4 = z_3 = 0,382,$$

$$\begin{aligned} z_4 &= a_4 + \frac{F_{N-5}}{F_{N-4}} (b_4 - a_4) = 0,324 + \frac{F_5}{F_6} (0,47 - 0,324) = 0,324 + \frac{8}{13} (0,146) = \\ &= 0,324 + 0,089 = 0,413, \end{aligned}$$

$$f(z_4) = f(0,413) = 2(0,413)^2 - e^{0,413} = 0,341 - 1,511 = -1,170,$$

$$f(y_4) = f(z_3) = -1,174,$$

$$[0,324; 0,413].$$



$$a_5 = 0,324, \quad b_5 = 0,413,$$

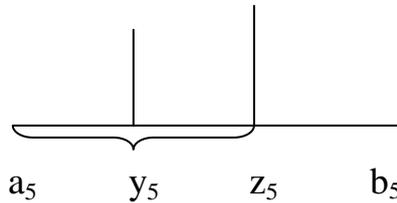
$$z_5 = y_4 = 0,382,$$

$$y_5 = a_5 + \frac{F_{N-7}}{F_{N-5}} (b_5 - a_5) = 0,324 + \frac{F_3}{F_5} (0,413 - 0,324) = 0,324 + \frac{3}{8} 0,089 = 0,357,$$

$$f(y_5) = 2(0,357)^2 - e^{0,357} = 0,254 - 1,429 = -1,175,$$

$$f(z_5) = f(y_4) = -1,174,$$

$$[0,324; 0,382].$$



$$a_6 = 0,324, \quad b_6 = 0,382,$$

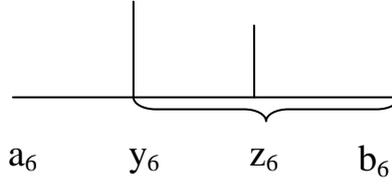
$$z_6 = y_5 = 0,357,$$

$$\begin{aligned} y_6 &= a_6 + \frac{F_{N-8}}{F_{N-6}} (b_6 - a_6) = a_6 + \frac{F_2}{F_4} (b_6 - a_6) = 0,324 + \frac{2}{5} (0,382 - 0,324) = \\ &= 0,324 + 0,023 = 0,347, \end{aligned}$$

$$f(y_6) = 2(0,347)^2 - e^{0,347} = 0,240 - 1,414 = -1,174,$$

$$f(z_6) = f(y_5) = -1,175,$$

$$[0,347; 0,382].$$



$$a_7 = 0,347, \quad b_7 = 0,382,$$

$$y_7 = z_6 = 0,357,$$

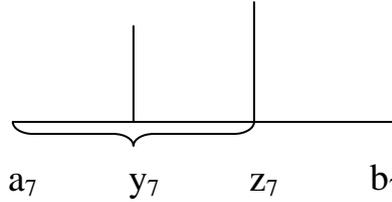
$$z_7 = a_7 + \frac{F_{N-8}}{F_{N-7}} (b_7 - a_7) = 0,347 + \frac{F_2}{F_3} (0,382 - 0,347) = 0,347 + \frac{2}{3} 0,035 =$$

$$= 0,347 + 0,023 = 0,370,$$

$$f(z_7) = 2(0,37)^2 - e^{0,37} = 0,273 - 1,447 = -1,174,$$

$$f(y_7) = f(z_6) = -1,175,$$

$$[0,347; 0,370].$$



$$a_8 = 0,347, \quad b_8 = 0,370,$$

$$z_8 = y_7 = 0,357,$$

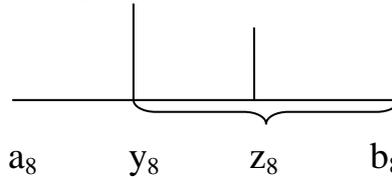
$$y_8 = a_8 + \frac{F_{N-10}}{F_{N-8}} (b_8 - a_8) = a_8 + \frac{F_0}{F_2} (b_8 - a_8) = 0,347 + \frac{1}{2} 0,023 =$$

$$= 0,347 + 0,011 = 0,358,$$

$$f(y_8) = 2(0,358)^2 - e^{0,358} = 0,256 - 1,430 = -1,174,$$

$$f(z_8) = f(y_7) = -1,175,$$

$$[0,358; 0,370].$$



$$a_9 = 0,358, \quad b_9 = 0,37,$$

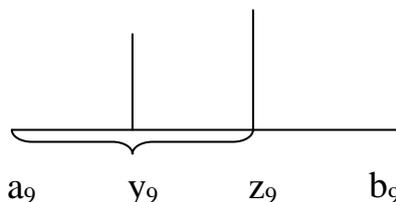
$$y_9 = z_8 = 0,357,$$

$$z_9 = a_9 + \frac{F_{N-10}}{F_{N-9}} (b_9 - a_9) = 0,358 + \frac{F_0}{F_1} (0,370 - 0,358) = 0,358 + \frac{1}{1} 0,012 = 0,370,$$

$$f(z_9) = f(0,370) = 2(0,370)^2 - e^{0,370} = 0,273 - 1,447 = -1,174,$$

$$f(y_9) = f(z_8) = -1,175,$$

$$[0,358; 0,370].$$



Пример 2.

Методом Фибоначчи найти минимум функции $f(x) = \sin(x)$.

Количество вычислений функции – 10 (должны получить 10 интервалов), первоначальный отрезок неопределенности $[\frac{\pi}{2}, \frac{5\pi}{2}]$.

$$y_0 = a_0 + \frac{F_{N-2}}{F_N} (b_0 - a_0), \quad N = 10, [0,1], \quad z_0 = a_0 + \frac{F_{N-1}}{F_N} (b_0 - a_0).$$

F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
1	1	2	3	5	8	13	21	34	55	89

$$a_0 = \frac{\pi}{2}, \quad b_0 = \frac{5\pi}{2},$$

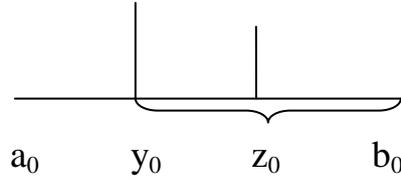
$$y_0 = \frac{\pi}{2} + \frac{F_8}{F_{10}} \left(\frac{5\pi}{2} - \frac{\pi}{2} \right) = \frac{\pi}{2} + \frac{34}{89} \cdot 2\pi = 0,5 \pi + 0,764 \pi = 1,264 \pi,$$

$$z_0 = \frac{\pi}{2} + \frac{F_9}{F_{10}} \left(\frac{5\pi}{2} - \frac{\pi}{2} \right) = \frac{\pi}{2} + \frac{55}{89} \cdot 2\pi = 0,5 \pi + 1,235\pi = 1,735 \pi,$$

$$f(y_0) = \sin(1,264 \pi) = -0,737,$$

$$f(z_0) = \sin(1,735 \pi) = -0,739,$$

$$[1,264 \pi; \frac{5\pi}{2}].$$



$$a_1 = 1,264 \pi, \quad b_1 = \frac{5\pi}{2},$$

$$y_1 = z_0 = 1,735 \pi,$$

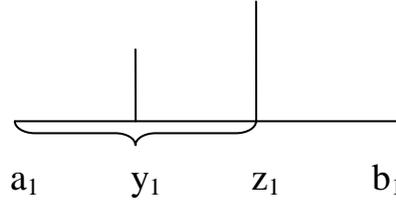
$$z_1 = a_1 + \frac{F_{N-2}}{F_{N-1}} (b_1 - a_1) = 1,264 \pi + \frac{F_8}{F_9} \left(\frac{5\pi}{2} - 1,264 \pi \right) = 1,264 \pi + \frac{34}{55} (1,236 \pi) =$$

$$= 1,264 \pi + 0,764 \pi = 2,028 \pi,$$

$$f(z_1) = \sin(2,028 \pi) = 0,087,$$

$$f(y_1) = f(z_0) = -0,739,$$

$$[1,264 \pi; 2,028 \pi].$$



$$a_2 = 1,264 \pi, \quad b_2 = 2,028 \pi,$$

$$z_2 = y_1 = 1,735 \pi,$$

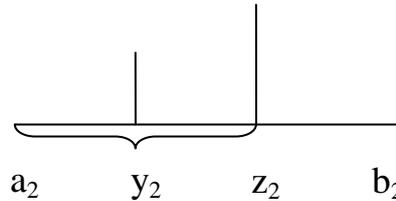
$$y_2 = a_2 + \frac{F_{N-4}}{F_{N-2}} (b_2 - a_2) = 1,264 \pi + \frac{F_6}{F_8} (b_2 - a_2) = 1,264 \pi + \frac{13}{34} (2,028 \pi -$$

$$- 1,264 \pi) = 1,264 \pi + 0,292 \pi = 1,556 \pi,$$

$$f(y_2) = \sin(1,556 \pi) = -0,9845,$$

$$f(z_2) = f(y_1) = -0,739,$$

$$[1,264 \pi; 1,735 \pi].$$



$$a_3 = 1,264 \pi, \quad b_3 = 1,735 \pi,$$

$$z_3 = y_2 = 1,556 \pi,$$

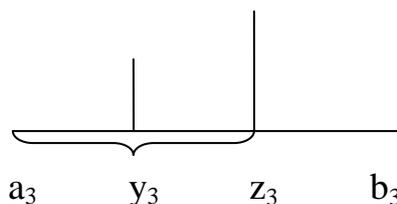
$$y_3 = a_3 + \frac{F_{N-5}}{F_{N-3}}(b_3 - a_3) = 1,264 \pi + \frac{F_5}{F_7}(b_2 - a_2) = 1,264 \pi + \frac{8}{21}(1,735\pi -$$

$$-1,264 \pi) = 1,264 \pi + 0,179\pi = 1,443 \pi,$$

$$f(y_3) = \sin(1,443 \pi) = -0,984,$$

$$f(z_3) = f(y_2) = -0,9845,$$

$$[1,264 \pi; 1,556 \pi].$$



$$a_4 = 1,264 \pi, b_4 = 1,556 \pi,$$

$$z_4 = y_3 = 1,443 \pi,$$

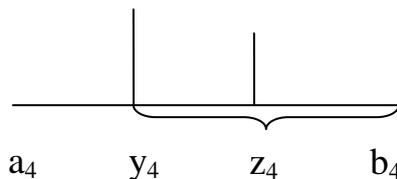
$$y_4 = a_4 + \frac{F_{N-6}}{F_{N-4}}(b_4 - a_4) = 1,264 \pi + \frac{F_4}{F_6}(b_4 - a_4) = 1,264 \pi + \frac{5}{13}(1,556 \pi -$$

$$-1,264 \pi) = 1,264 \pi + 0,112 \pi = 1,376 \pi,$$

$$f(y_4) = \sin(1,376 \pi) = -0,925,$$

$$f(z_4) = f(y_3) = -0,984,$$

$$[1,376 \pi; 1,556 \pi].$$



$$a_5 = 1,376 \pi, b_5 = 1,556 \pi,$$

$$y_5 = z_4 = 1,443 \pi,$$

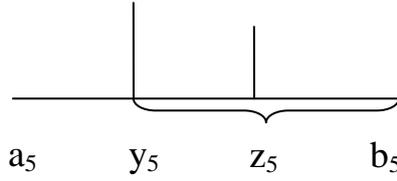
$$z_5 = a_5 + \frac{F_{N-6}}{F_{N-5}}(b_5 - a_5) = 1,376 \pi + \frac{5}{8}(1,556 \pi - 1,376 \pi) = 1,376 \pi + 0,112 \pi =$$

$$= 1,488 \pi,$$

$$f(z_5) = \sin(1,488 \pi) = -0,9992,$$

$$f(y_5) = f(z_4) = -0,984,$$

$$[1,443 \pi; 1,556 \pi].$$



$$a_6 = 1,443 \pi, \quad b_6 = 1,556 \pi,$$

$$y_6 = z_5 = 1,488 \pi,$$

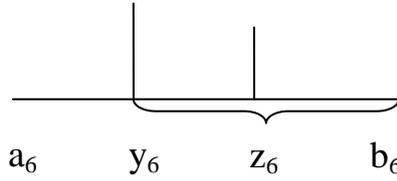
$$z_6 = a_6 + \frac{F_{N-7}}{F_{N-6}} (b_6 - a_6) = 1,443 \pi + \frac{3}{5} (1,556 \pi - 1,443 \pi) = 1,443 \pi +$$

$$+ 0,067 \pi = 1,51 \pi,$$

$$f(z_6) = \sin(1,51 \pi) = -0,9995,$$

$$f(y_6) = f(z_5) = -0,9992,$$

$$[1,488 \pi; 1,556 \pi].$$



$$a_7 = 1,488 \pi, \quad b_7 = 1,556 \pi,$$

$$y_7 = z_6 = 1,51 \pi,$$

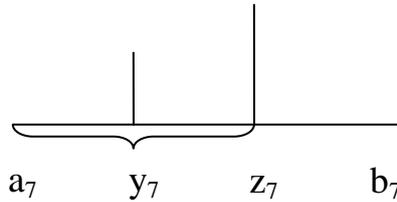
$$z_7 = a_7 + \frac{F_{N-8}}{F_{N-7}} (b_7 - a_7) = 1,488 \pi + \frac{2}{3} (1,556 \pi - 1,488 \pi) = 1,488 \pi +$$

$$+ 0,045 \pi = 1,533 \pi,$$

$$f(z_7) = \sin(1,533 \pi) = -0,994,$$

$$f(y_7) = f(z_6) = -0,999,$$

$$[1,488 \pi; 1,533 \pi].$$



$$a_8 = 1,488 \pi, \quad b_8 = 1,533 \pi,$$

$$z_8 = y_8 = 1,51 \pi,$$

$$y_8 = a_8 + \frac{F_{N-10}}{F_{N-8}} (b_8 - a_8) = 1,488 \pi + \frac{1}{2} (1,533 \pi - 1,488 \pi) = 1,488 \pi +$$

$$+ 0,022 \pi = 1,510 \pi,$$

$$f(y_8) = f(z_8) = -0,994.$$

Процесс остановился, так как $y_8 = z_8$, следовательно, точкой, в которой достигается минимум $f(x) = \sin(x)$ на отрезке $[\frac{\pi}{2}, \frac{5\pi}{2}]$, является точка $x = 1,51 \pi$.

Примеры использования метода квадратичной интерполяции

Метод применяется для поиска безусловного минимума функции одной переменной.

Задается начальная точка и с помощью пробного шага находятся три точки. Затем строится интерполяционный полином второй степени, проходящий через три точки. В качестве приближения точки минимума берется точка минимума полинома. Процесс поиска заканчивается, когда полученная точка отличается от наилучшей из трех опорных точек не более, чем на заданную величину.

Алгоритм поиска:

Шаг 1. Задать начальную точку, величину шага, малое положительное число x , Δx , ε .

Шаг 2. Вычислить $x_2 = x_1 + \Delta x$.

Шаг 3. Вычислить $f(x_1) = f_1$, $f(x_2) = f_2$.

Шаг 4. Сравнить $f(x_1), f(x_2)$.

Если $f(x_1) > f(x_2)$, $x_3 = x_1 + 2\Delta x$, $F(x_1) \leq f(x_2)$, $x_3 = x_1 - \Delta x$.

Шаг 5. Вычислить $f(x_3) = f_3$.

Шаг 6. Найти $F_{\min} = \{f_1, f_2, f_3\}$.

Шаг 7. Вычислить точку минимума по трем точкам:

$$x^* = \frac{1}{2} \frac{(x_2^2 - x_3^2)f_1 + (x_3^2 - x_1^2)f_2 + (x_1^2 - x_2^2)f_3}{(x_2 - x_3)f_1 + (x_3 - x_1)f_2 + (x_1 - x_2)f_3}.$$

Шаг 8. Проверить выполнение условия окончания

$$\left| \frac{F_{\min} - f(x^*)}{f(x^*)} \right| < \varepsilon.$$

Если условие выполнено, то процедура закончена.

Если нет, то выбрать наилучшую точку и две точки по обе стороны от нее. Обозначить эти точки в естественном порядке и перейти к шагу 6.

Приведем пример применения метода квадратичной интерполяции для функции $f(x) = 2x^2 - e^x$. Ищется минимум функции $f(x)$, начальная точка $x_1 = 1$, величина шага $\Delta x = 0,5$, $\varepsilon = 0,001$.

$$x_1 = 0,5 \quad f(x_1) = -1,14872.$$

$$x_2 = 1 \quad f(x_2) = -0,71828.$$

$$x_3 = 1,5 \quad f(x_3) = 1,83110.$$

$$x^* = 0,04702 \quad f(x^*) = -1,04372.$$

Лучшая точка x_1 , переобозначим точки и сделаем следующий шаг.

$$x_1 = 0,04702 \quad f(x_1) = -1,04372.$$

$$x_2 = 0,5 \quad f(x_2) = -1,14872.$$

$$x_3 = 1 \quad f(x_3) = -0,71828.$$

$$x^* = 0,37459 \quad f(x^*) = -1,17376.$$

Лучшая точка x^* , переобозначим точки и сделаем следующий шаг.

$$x_1 = 0,04702 \quad f(x_1) = -1,04372.$$

$$x_2 = 0,37459 \quad f(x_2) = -1,17376.$$

$$x_3 = 0,5 \quad f(x_3) = -1,14872.$$

$$x^* = 0,36150 \quad f(x^*) = -1,17411.$$

Лучшая точка $x^* = 0,36150$, переобозначим точки и сделаем следующий шаг.

$$x_1 = 0,04702 \quad f(x_1) = -1,04372.$$

$$x_2 = 0,36150 \quad f(x_2) = -1,17411.$$

$$x_3 = 0,37459 \quad f(x_3) = -1,17376.$$

$$x^* = 0,357937 \quad f(x^*) = -1,174138.$$

Лучшая точка $x^* = 0,357937$, переобозначим точки и сделаем следующий шаг.

$$x_1 = 0,04702 \quad f(x_1) = -1,04372.$$

$$x_2 = 0,357937 \quad f(x_2) = -1,174138.$$

$$x_3 = 0,36150 \quad f(x_3) = -1,17411.$$

$$x^* = 0,3575219 \quad f(x^*) = -1,174138.$$

Процесс поиска завершен, так как достигнута заданная точность.

Метод конфигураций

Метод применяется для поиска безусловного экстремума функции многих переменных.

Метод представляет собой комбинацию исследующего поиска и ускоряющегося поиска по образцу. Исследующий поиск ориентирован на выявление локального поведения целевой функции и определение направления ее убывания. Полученная информация используется при поиске по образцу. В качестве множества направлений поиска выбирается множество координатных направлений. Если значение функции в пробной точке меньше значения функции в исходной точке, то шаг считается удачным. В противном случае необходимо вернуться в предыдущую точку и сделать шаг в противоположном направлении. После перебора всех координат исследующий поиск завершается. Далее производится поиск по образцу, заключающийся в движении по направлению от старой точки к новой.

Алгоритм метода:

Шаг 1. Задать начальную точку x^0 , величину шага по координатным направлениям Δ_i , ε .

Шаг 2. Осуществить исследующий поиск по выбранному координатному направлению.

Если $f(y_1^i, \dots, y_i^i + \Delta_i, \dots, y_n^i) < f(y_1^i, \dots, y_i^i, \dots, y_n^i)$, $y^{i+1} = y^i + \Delta_i d_i$, перейти к шагу 3. Если $f(y_1^i, \dots, y_i^i - \Delta_i, \dots, y_n^i) < f(y_1^i, \dots, y_i^i, \dots, y_n^i)$, $y^{i+1} = y^i - \Delta_i d_i$, перейти к шагу 3.

Шаг 3. Проверить условия:

– если $i < n$, то положить $i = i + 1$ и перейти к шагу 2;

– если $i = n$, проверить $f(y^{n+1}) < f(x^k)$, перейти к шагу 4, если $f(y^{n+1}) \geq f(x^k)$, перейти к шагу 5.

Шаг 4. Провести поиск по образцу.

Положить $x^{k+1} = y^{n+1}$, $y^1 = x^{k+1} + \lambda(x^{k+1} - x^k)$, $i = 1$, $k = k + 1$ и перейти к шагу 2.

Шаг 5. Проверить условие окончания.

Приведем пример применения метода конфигураций для функции $f(x) = 100x_1^2 + x_2^2$. Ищется минимум функции $f(x)$, начальная точка $x^0 = (0, 10)$, величина шага $\Delta = 0,5$, $\varepsilon = 0,1$.

$$x^0 = (0, 10) \quad f(x^0) = 100.$$

Исследующий поиск:

$$x^1 = (0, 9.5) \quad f(x^1) = 90,25.$$

Поиск по образцу:

$$x^3 = (0, 8) \quad f(x^3) = 64.$$

Поиск по образцу:

$$x^4 = (0, 4.5) \quad f(x^4) = 20,25.$$

Поиск по образцу:

$$x^5 = (0, -2) \quad f(x^5) = 4.$$

Исследующий поиск:

$$x^6 = (0, -1.5) \quad f(x^6) = 2,25.$$

Исследующий поиск с уменьшением шага:

$$x^7 = (0, 0) \quad f(x^7) = 0.$$

Наименьшее значение функции достигается в точке $(0, 0)$ с точностью $\varepsilon = 0,1$.

Содержание отчета

1. Титульный лист.
 2. Цель и задачи работы.
 3. Результаты выполнения работы.
 4. Ответы на контрольные вопросы.
- Защитить отчет у преподавателя.

Контрольные вопросы

1. Перечислите основные методы численной оптимизации.
2. Область применения итерационных методов оптимизации.
3. Назовите основные особенности интерполяционных методов.
4. Основное назначение метода Фибоначчи.
5. Перечислите основные шаги метода квадратичной интерполяции.
6. Основное назначение метода конфигураций.

Практическая работа № 2

КЛАССИЧЕСКИЕ МЕТОДЫ БЕЗУСЛОВНОЙ ОПТИМИЗАЦИИ

Цель работы: изучить теоретические основы и получить практические навыки решения задач безусловной оптимизации.

Оптимизация функций многих переменных

Разнообразные методы многомерной оптимизации различают обычно по виду информации, которая необходима им в процессе работы:

- *методы прямого поиска* (методы нулевого порядка), которым нужны только значения целевой функции;
- *градиентные методы* (методы первого порядка), которым дополнительно требуются частные производные первого порядка целевой функции;
- *ньютоновские методы* (методы второго порядка), которые используют и частные производные второго порядка.

Рекуррентная схема большинства методов многомерной оптимизации основана на выражении

$$\mathbf{X}^{k+1} = \mathbf{X}^k + \lambda_k s^k.$$

Они отличаются друг от друга способом выбора длины шага λ_k и направления поиска s^k . Первый из них – это скаляр, а второй – вектор единичной длины.

Как видно, методы многомерной оптимизации сводятся в той или иной форме к методам одномерной оптимизации. Исторически первыми подходами были прямые обобщения одномерных методов (например поиск Фибоначчи или интерполяционные методы) на многомерные задачи. Эти подходы не были особенно удачными в связи с тем, что их трудоемкость росла экспоненциально с ростом размерности задачи («проклятие размерности» по Беллману).

Все обсуждаемые ниже методы предполагают ту или иную степень гладкости целевой функции. Все они не дают гарантированной сходимости к глобальному минимуму, а в лучшем случае сходятся к одному из локальных или иногда только к седловой точке.

Методы прямого поиска

Методы прямого поиска являются методами, которые используют только значения функций и не используют никакой внутренней модели целевой функции. Вместо этого направления спуска и длины шагов выбираются эвристически или по некоторой схеме, которая ориентируется на какую-то внутреннюю модель, но не всегда оптимальным образом. Таким образом, появляется риск не найти улучшения на каком-либо шаге. Однако неудачи могут быть изучены, и поведение метода может измениться в соответствии с полученной информацией. Такой характер поведения алгоритмов прямого поиска послужил причиной, по которой за ними закрепилось общее название методов *проб и ошибок*. На разработку таких методов было затрачено много усилий, и некоторые из методов выдержали проверку временем и являются эффективными процедурами, нашедшими широкое применение и используемыми до сих пор. Основное достоинство этих методов – не теоретическое доказательство сходимости или скорости сходимости, а простота реализации, доказанность их работоспособности на практике и отсутствие сложных подготовительных этапов, вроде аналитического вычисления первых и вторых производных. Кроме того, при оптимизации сложных функций далеко не всегда возможно аналитическое определение производных, не говоря уже о разрывных функциях.

В случае же выпуклой или квадратичной унимодальной функции методы прямого поиска в общем работают хуже методов первого и второго порядка, обсуждаемых в другом разделе.

Покоординатный спуск

Простейшим методом прямого поиска является метод покоординатного спуска. Из начальной точки A производится поиск минимума вдоль направления оси x_1 и определяется точка B , в которой касательная к линии уровня параллельна оси x_1 . Затем производится поиск минимума из точки B в направлении оси x_2 и определяется точка C . Продолжая спуск вдоль осей x_3, x_4, \dots, x_n , затем снова вдоль x_1, x_2, \dots, x_n , приходим в точку минимума. Для поиска вдоль направления $x_i, i = 1, \dots, n$ может быть использован любой из одномерных методов, рассмотренных в предыдущем разделе.

Теоретически метод покоординатного спуска эффективен для унимодальной функции, но на практике он оказывается слишком медленным. Метод хорошо работает на гладких, сепарабельных функциях и особенно на квадратичных, линии уровня которых похожи на концентрические круги, т. е. в случае эллипсоидов они должны иметь главные оси примерно равными по длине. Метод притягателен своей простотой, и были сделаны многочисленные попытки улучшить его сходимость в общем случае, которые не привели к большим успехам. Поэтому были разработаны более сложные методы, использующие больше информации на основании уже полученных значений функции. Одной из таких модификаций покоординатного спуска является поиск по образцу, разрешающий шаги по диагонали координатной плоскости.

Метод Хука-Дживса (поиск по образцу)

Этот метод впервые предложен в 1960 г., но и в настоящее время широко используется на практике. Хук и Дживс предложили логически простую стратегию поиска, использующую априорные сведения и в то же время отвергающую устаревшую информацию относительно характера поведения целевой функции. Поиск состоит из последовательности шагов исследующего поиска вокруг базисной точки, после которого, в случае успеха, следует поиск по образцу.

Алгоритм Хука-Дживса

Шаг А. Выбрать начальную базисную точку B_1 и шаг длиной h_j для каждой переменной $x_j, j = 1, \dots, n$.

Шаг Б. Вычислить $f(x)$ в базисной точке B_1 с целью получения сведений о локальном поведении функции. Эти сведения будут использованы для нахождения подходящего направления поиска по образцу, с помощью которого можно надеяться достичь большего убывания значений функции. Информация о локальном поведении $f(x)$ находится следующим образом:

1. Вычисляется значение функции $f(x)$ в базисной точке B_1 .
2. Каждая переменная по очереди изменяется прибавлением длины шага.

Вычисляются значения функции $f(B_1 + h_j \bar{e}_j)$, где \bar{e}_j – единичный вектор в направлении оси x_j . Если это приводит к уменьшению значения функции, то B_1 заменяется на $B_1 + h_j \bar{e}_j$. В противном случае

вычисляется значение $f(B_1 - h_j \bar{e}_j)$, если значение функции уменьшилось, то B_1 заменяется на $B_1 - h_j \bar{e}_j$. Если ни один из проделанных шагов не приводит к уменьшению значения функции, то точка B_1 остается без изменения и рассматриваются изменения в направлении следующей оси и т. д. После рассмотрения всех n переменных получают новую базисную точку B_2 .

3. Если $B_2 = B_1$, т. е. уменьшение функции не было достигнуто, то исследование повторяется вокруг той же базисной точки B_1 , но с уменьшенной длиной шага (обычно шаг уменьшается в десять раз).

4. Если $B_2 \neq B_1$, то производится поиск по образцу.

Шаг В. При поиске по образцу используется информация, полученная в процессе исследования, и минимизация функции совершается поиском в направлении, заданном образцом:

1. Разумно двигаться из базисной точки B_2 в направлении $(B_2 - B_1)$, поскольку поиск в этом направлении уже привел к уменьшению значения функции. Поэтому вычислим функцию в точке $P = B_1 + 2(B_2 - B_1)$.

2. Провести исследующий поиск вокруг точки P .

3. Если наименьшее значение на шаге $(B - 2)$ меньше значения в базисной точке B_2 , то получают новую базисную точку B_3 и переходят к шагу $(B - 1)$. В противном случае не проводят поиск по образцу из точки B_2 , а проводят исследующий поиск в точке B_2 .

Шаг Г. Завершить этот процесс, когда длины шагов будут уменьшены до заданного значения (точность определения точки минимума).

Строгое доказательство сходимости метода Хука-Дживса было дано только в случае строгой выпуклости и непрерывной дифференцируемости целевой функции.

Достоинством алгоритма является то, что его операции очень просты и даже в случае непредвиденных обстоятельств не приводят к запрещенным арифметическим операциям (типа деления на ноль). Еще одним достоинством является то, что алгоритм требует очень мало (порядка n) памяти компьютера. Поиск по образцу ускоряет продвижение алгоритма по дну оврага, а шаги исследующего поиска позволяют приблизительно определить направление градиента.

Недостатком алгоритма является ограничение исследующего поиска только направлениями осей координат, что может привести к преждевременной остановке алгоритма сколь угодно далеко от точки минимума, если направление оврага окажется не параллельным осям координат. Еще один недостаток – большое число вычислений целевой функции.

Были разработаны многие модификации метода поиска по образцу, но все они не являются столь популярными, как оригинальный алгоритм.

Метод вращающихся координат Розенброка

Идея Розенброка – устранить ограничения покоординатного поиска и метода Хука-Дживса на возможные изменения переменных, которые в этих методах могли изменяться только в направлениях, параллельных осям координат. В методе Розенброка шаги поиска могут производиться вдоль направлений координатных осей, которые могут вращаться в пространстве. Одна из осей устанавливается в направлении, которое представляется наиболее удачным. Для этой цели накапливается опыт успешных и неудачных перемещений в процессе оптимизации примерно так же, как и в методе Хука-Дживса. Остальные оси выбираются ортогонально первой.

Первоначально оси выбираются совпадающими с осями координат пространства оптимизации. Затем вдоль каждой из осей делается шаг определенной длины. Если шаг был успешен, то длина шага увеличивается, и попытка повторяется до тех пор, пока не последует неудача. Если первоначальный шаг был неудачен, то длина шага уменьшается, и попытка повторяется до тех пор, пока не будет получено улучшение функции. Затем координатные оси поворачиваются с помощью процедуры *ортогонализации*. Такой процесс повторяется до тех пор, пока не будет выполнено условие останова. Например, улучшение значения функции станет малым или перемещение на нескольких шагах подряд будет невелико.

Так же, как и поиск по образцу, метод Розенброка не требует информации о производных, но в процессе вращения осей координат отслеживает положение градиента функции. Это позволяет ему двигаться вдоль дна оврага и преодолевать повороты. Метод также не использует одномерную оптимизацию вдоль выбранных осей. Это делает процедуру весьма устойчивой (*робастной*).

Однако имеется дополнительный недостаток – процедура ортогонализации очень дорого стоит, и в смысле количества операций, производимых с системой координат, и в смысле необходимой памяти, которой требуется уже порядка n^2 , а не n , как в методе Хука-Дживса. Это значит, что даже в случае относительно «дешевой» целевой функции вычислительные затраты на ортогонализацию оказываются значительными в сравнении с общими затратами. Кроме того, размерность решаемых задач ограничивается требуемым объемом памяти.

Существует несколько модификаций алгоритма Розенброка, главная из которых состоит в отказе от постоянной длины шага и замене на процедуру масштабирования длины шага в зависимости от успешности поиска в данном направлении.

Метод ортогонализации Дэвиса, Свана и Кампея

Данный метод представляет собой комбинацию идеи Розенброка о вращающихся координатах и одномерных методов оптимизации.

Начиная из некоторой стартовой точки, выполняется одномерная оптимизация поочередно в направлениях, совпадающих с осями координат. После этого выполняется одномерная оптимизация вдоль направления суммарного перемещения на всех шагах. Затем выполняется ортогонализация.

Если хотя бы один из одномерных поисков был неудачным, то новое множество направлений не будет более отражать полное пространство оптимизации (система координатных осей становится линейно-зависимой).

Поэтому в процесс ортогонализации включаются только те из старых направлений, перемещение по которым было больше некоторого установленного минимума. Остальные координатные вектора остаются неизменными. Если суммарное перемещение на одной итерации оказалось меньше, чем длина шага одномерного поиска, этот шаг уменьшается в 10 раз.

После ортогонализации одно из новых направлений (первое) совпадает с направлением суммарного перемещения на данной итерации. Это интерпретируется как первый одномерный поиск на новой итерации. Остается сделать еще $(n - 1)$ шагов вдоль оставшихся координатных осей и один – вдоль направления суммарного перемещения. Итерации продолжаются до тех пор, пока длина вектора суммар-

ного перемещения на одной итерации не окажется меньше, чем заранее определенный минимум.

Численные эксперименты показали, что данный метод превосходит методы Розенброка и Хука-Дживса в случае, когда целевая функция достаточно гладкая и имеет не очень много переменных. В случае, когда число переменных велико, дорогостоящие процессы ортогонализации и одномерной оптимизации сводят на нет все преимущества алгоритма. Отмечены также случаи зигзагообразной траектории алгоритма по дну оврага, что, естественно, существенно снижает уверенность в том, что алгоритм работоспособен в овражных ситуациях.

Метод деформируемого многогранника Нелдера-Мида

В 1962 г. Спендли, Хекст и Химсворт предложили метод поиска по регулярному многограннику, разработанный ими в связи со статистическим планированием эксперимента. Этот метод получил название симплексного, так как использует так называемые *симплексы* при построении траектории поиска. Несмотря на совпадения в названиях, этот метод не имеет никакого отношения к симплекс-методу линейного программирования. Обсудим основные идеи этого метода.

Определение. Регулярным многогранником (симплексом) в n -мерном пространстве называется множество из $(n + 1)$ равноудаленной точки. На плоскости симплексом является равносторонний треугольник, в пространстве – правильный тетраэдр.

Из аналитической геометрии известно, что координаты вершин регулярного симплекса определяются следующей матрицей размерности $n \times (n + 1)$:

$$D = \begin{bmatrix} 0 & d_1 & d_2 & d_2 & \dots & d_2 \\ 0 & d_2 & d_1 & d_2 & \dots & d_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & d_2 & d_2 & d_2 & \dots & d_1 \end{bmatrix},$$

$$d_1 = \frac{t}{n\sqrt{2}}(\sqrt{n+1} + n - 1),$$

$$d_2 = \frac{t}{n\sqrt{2}}(\sqrt{n+1} - 1).$$

Здесь столбцы есть координаты вершин, а строки дают номера координат, t – расстояние между вершинами.

При $n = 2$ и $t = 1$ получим треугольник с координатами

$$D = \begin{bmatrix} 0 & 0,956 & 0,259 \\ 0 & 0,259 & 0,956 \end{bmatrix}.$$

При оптимизации этим методом целевая функция вычисляется в каждой из вершин регулярного симплекса. Из вершины, где целевая функция максимальна, проводится проектирующая прямая через центр тяжести симплекса, затем вершина с максимальным значением функции отбрасывается и строится новый отраженный симплекс из оставшихся прежних точек и одной новой точки, расположенной на проектирующей прямой на надлежащем расстоянии от центра тяжести.

Продолжение этой процедуры, в которой каждый раз отбрасывается точка с максимальным значением целевой функции, а также использование правил уменьшения размера многогранника позволяют осуществить поиск, в котором величина шага на любом шаге фиксирована, а направление поиска можно менять (производная не используется).

Определенные трудности, встречающиеся при использовании этого метода, а именно: отсутствие ускорения поиска в «хорошем» направлении и трудности при поиске на искривленных «оврагах» и «хребтах», привели к необходимости улучшений.

В 1964 г. Нелдер и Мид предложили модификацию, в которой симплекс может изменять свою форму. Так как в этом случае симплекс не будет уже регулярным, метод назвали поиском по деформируемому многограннику.

Обсудим этот метод. Минимизируется функция независимых переменных с использованием $(n + 1)$ вершин многогранника. Вершина, в которой значение целевой функции наибольшее, проектируется через центр тяжести *остальных вершин*. Улучшенные значения целевой функции находятся последовательной заменой точки с наибольшим значением на более «хорошие» точки, пока не будет найден минимум. При этом многогранник может вытягиваться или сжиматься по выбранному направлению.

Пусть $X^i = (x_1^i, x_2^i, \dots, x_n^i)^T$, $i = 1, \dots, n + 1$, является i -й вершиной на некотором $(k$ -м) этапе поиска. Пусть также

$$f(X^h) = \max \{f(X^1), f(X^2), \dots, f(X^{n+1})\},$$

$$f(X^l) = \min \{f(X^1), f(X^2), \dots, f(X^{n+1})\},$$

т. е. в точке X^h целевая функция принимает наибольшее, а в точке X^l наименьшее значение среди точек многогранника.

Так как многогранник состоит из $(n + 1)$ точек, то центр тяжести всех вершин, кроме X^h , обозначим через X^{n+2} . Координаты центра тяжести определяются по формуле

$$x_j^{n+1} = \frac{1}{n} \left[\left(\sum_{j=1}^{n+1} x_j^i \right) - x_j^h \right], \quad j = 1, \dots, n.$$

Начальный многогранник выбирается обычно в виде регулярного симплекса (но это не обязательно) с точкой № 1 в качестве начала координат. Начало координат может быть также помещено в центре тяжести многогранника.

Опишем операции, из которых состоит алгоритм:

1. Отражение – проектирование X^h через центр тяжести оставшихся вершин в соответствии с соотношением

$$X^{n+3} = X^{n+2} + \alpha \cdot (X^{n+2} - X^h),$$

где $\alpha \geq 1$ – коэффициент отражения.

2. Растяжение.

Если $f(X^{n+3}) \leq f(X^l)$, то вектор $f(X^{n+3}) \leq f(X^l)$, растягивается в соответствии с соотношением

$$X^{n+4} = X^{n+2} + \gamma \cdot (X^{n+3} - X^{n+2}),$$

где $\gamma > 1$ – коэффициент растяжения.

Если $f(X^{n+4}) < f(X^l)$, то X^h заменяется на X^{n+1} и процедура продолжается снова с операции 1.

В противном случае X^h заменяется на X^{n+3} и также осуществляется переход к операции 1.

3. Сжатие.

Если $f(X^{n+3}) > f(X^i)$ для всех $i \neq h$, то вектор $(X^h - X^{n+2})$ сжимается в соответствии с формулой

$$X^{n+5} = X^{n+2} + \beta \cdot (X^h - X^{n+2}),$$

где $0 < \beta < 1$ – коэффициент сжатия.

Затем X^h заменяется на X^{n+5} и осуществляется переход к операции 1 для продолжения поиска.

4. Редукция.

Если $f(X^{n+3}) > f(X^h)$, то все векторы $(X^i - X^l)$ уменьшаются в два раза с отсчетом от X^l в соответствии с формулой

$$X^i = X^l + 0,5(X^i - X^l), \quad i = 1, \dots, n + 1.$$

Затем – возврат к операции 1 для продолжения поиска на следующем шаге.

Если ни одна из операций (растяжение, сжатие, редукция) не выполняется, то происходит следующее отражение.

Критерий останова (окончания поиска), предложенный Нелдером и Мидом, следующий:

$$\left[\frac{1}{n+1} \sum_{i=1}^{n+1} [f(X^i) - f(X^{n+2})]^2 \right]^{\frac{1}{2}} \leq \varepsilon,$$

где ε – произвольное малое число;

X^{n+2} – центр тяжести многогранника на данном шаге.

Деформируемый многогранник в противоположность жесткому симплексу адаптируется к топографии целевой функции, вытягиваясь вдоль длинных наклонных поверхностей, изменяя направление в изогнутых впадинах и сжимаясь в окрестности минимума, что достигается выбором коэффициентов сжатия, растяжения и отражения.

Обсудим выбор коэффициентов α , β , γ . После того, как деформируемый многогранник подходящим образом промасштабирован, его размеры должны поддерживаться неизменными, пока изменения в топологии задачи не требуют применения многогранника другой формы. Это можно реализовать только при $\alpha = 1$. Было показано, что при $\alpha < 1$ требуется большее количество вычислений целевой функции. Если $\alpha \gg 1$, то это затрудняет адаптацию многогранника при изменениях направления поиска и замедляет сходимость в окрестности минимума (мешает уменьшать размеры многогранника). Таким образом, $\alpha = 1$ выбирается как компромисс.

Чтобы выяснить, какое влияние на поиск имеет выбор β и γ , были проведены решения тестовых задач. Оказалось, что влияние β на эффективность поиска более заметно, чем влияние γ . При $0 < \beta < 0.4$ существует возможность, что из-за сжатия многогранника будет иметь место преждевременное окончание процесса. При $\beta > 0.6$ может потребоваться избыточное число шагов и больше машинного времени для достижения окончательного значения. Для практических нужд были рекомендованы следующие значения коэффициентов:

$$\begin{aligned}\alpha &= 1, \\ 0.4 &\leq \beta \leq 0.6, \\ 2.8 &\leq \gamma \leq 3.0.\end{aligned}$$

Подчеркнем чисто рекомендательный характер этих значений.

Как и у всякого другого метода, получившего широкое признание, у метода Нелдера-Мида имеется много модификаций. Наиболее существенной является модификация, при которой после нормальной остановки алгоритма выполняется попытка построить новый стартовый симплекс. Для этой цели выполняются маленькие пробные шаги в каждом направлении осей координат. Если хотя бы один из этих шагов оказывается успешным, то поиск выполняется снова, но симплекс имеет стороны значительно меньшей длины. Этот рестарт рекомендуется самой процедурой поиска, так как (в особенности для большого числа переменных) симплекс обычно имеет тенденцию терять объем (размерность симплекса становится меньше размерности пространства), т. е. симплекс стягивается в точку без достижения реального минимума.

Другой модификацией является рекомендация подстраивать коэффициенты растяжения и сжатия в зависимости от того, имел ли место успех при выполнении этой операции или была неудача.

Для задач с небольшим числом переменных метод деформируемого многогранника известен как надежный и робастный метод, который, однако, является относительно дорогостоящим. Для работы алгоритма необходимо хранить $(n + 1)$ векторов, а операция отображения требует порядка n^2 операций. Нелдер и Мид эмпирически обосновали утверждение, что с ростом размерности n количество обращений к целевой функции растет примерно как $n^{2.11}$, однако это обоснование проведено на размерности меньше десяти и вряд ли будет в силе для много большего числа переменных.

Методы прямого случайного поиска (СП)

Прямой случайный поиск

После того, как задана начальная точка, строится случайная траектория для последовательности шагов, каждый из которых проводится в направлении от некоторой точки, где функция минимальна на данном этапе, до следующей точки, соответствующей еще более низкому значению функции.

Начиная из X^0 , следующая случайная точка получается по правилу

$$X^{k+1} = X^k + \lambda^k \left[\beta \frac{Z^k}{\|Z^k\|} + (1 - \beta)R^k \right],$$

где λ^k – величина шага (скаляр), который увеличивается после успешного шага и уменьшается после неудачного;

Z^k – вектор «предыстории», указывающий среднее направление поиска на предыдущих шагах: $Z^{k+1} = \gamma \times Z^k + (1 - \gamma) \times (X^{k+1} - X^k)S^k$;

β – коэффициент предыстории ($0 < \beta < 1$), заменяемый в процессе поиска;

R^k – единичный случайный вектор, компоненты которого распределены нормально, реализуемый датчиком псевдослучайных чисел;

γ – постоянный весовой множитель;

S^k – вектор масштабных множителей для «подходящего» масштабирования пространства.

Вектор $X^{k+1} + 1$ будет принят или отвергнут в зависимости от того, выполняется ли соотношение $f(X^{k+1}) < f(X^k)$. После того как X^{k+1} принят (отвергнут), λ^k увеличивают (или уменьшают) с помощью множителя, зависящего от того, был ли поиск трудным или легким.

Тестирование алгоритма прямого СП показывает, что в общем случае он менее эффективен, чем регулярные алгоритмы.

Случайный поиск на гиперсфере

Начальная точка X^0 выбирается в качестве центра n -мерной гиперсферы. На этой гиперсфере случайным образом выбираются несколько точек, вычисляется значение функции в них и отмечается

лучшая среди них \hat{x}^0 . Затем проводится регулярный одномерный поиск вдоль прямой, соединяющей X^0 и \hat{x}^0 , и определяется точка с минимальным значением функции.

После этого процедура повторяется из нового центра гиперсферы – точки X^1 . Радиус гиперсферы можно изменять. Поиск может быть ускорен, если точки на гиперсфере выбираются случайно, но с ограничением, чтобы они не отклонялись более, чем на некоторый выбранный угол друг от друга и от предыдущего направления поиска. Другая модификация – использование эллипсоида вместо сферы.

Существует множество алгоритмов СП, отличающихся друг от друга соотношением случайности и регулярности, способом выбора случайного направления, поведением при успехах и неудачах и т. п.

Завершая раздел, посвященный методам прямого поиска, подчеркнем их достоинства и недостатки.

Достоинства: простота программирования, простота подготовительного этапа, универсальность, устойчивость к ошибкам счета.

Недостатки: весьма сомнительные или отсутствующие гарантии сходимости к решению, большое количество требующих настройки параметров, от которых зависит успех решения, экспоненциальный (обычно) рост объема вычислений в зависимости от размерности.

Общая рекомендация об использовании прямого поиска, выработанная многолетним опытом, может быть сформулирована следующим образом: используйте прямой поиск только в крайнем случае, когда твердо убедитесь, что никакой другой метод не может быть применен, иначе вам придется дорого заплатить за вашу ошибку в выборе метода, так как это приведет к чрезмерным затратам при отсутствии каких-либо гарантий, что будет найдено действительное решение.

Оправданием существованию и применению методов прямого поиска служит тот факт, что при решении реальных задач слишком часто никакой другой метод и не может быть применен.

Градиентные методы

Градиентные методы безусловной оптимизации используют только первые производные целевой функции и являются методами линейной аппроксимации на каждом шаге, т. е. целевая функция на

каждом шаге заменяется касательной гиперплоскостью к ее графику в текущей точке.

На k -м этапе градиентных методов переход из точки X^k в точку X^{k+1} описывается соотношением

$$X^{k+1} = X^k + \lambda^k \cdot S^k, \quad (1)$$

где λ^k – величина шага;

\bar{S}_k – вектор в направлении $(X^{k+1} - X^k)$.

Методы наискорейшего спуска

Впервые такой метод рассмотрел и применил еще О. Коши в XVIII в. Идея его проста: градиент целевой функции $f(X)$ в любой точке есть вектор в направлении наибольшего возрастания значения функции. Следовательно, антиградиент будет направлен в сторону наибольшего убывания функции и является направлением наискорейшего спуска. Антиградиент (и градиент) ортогонален поверхности уровня $f(X)$ в точке X . Если в (1) ввести направление

$$\bar{S}_k = -\frac{\nabla f(X^k)}{\|\nabla f(X^k)\|},$$

то это будет направление наискорейшего спуска в точке X^k .

Получаем формулу перехода из X^k в X^{k+1} :

$$X^{k+1} = X^k - \lambda^k \frac{\nabla f(X^k)}{\|\nabla f(X^k)\|} = X^k - \mu \nabla f(X^k).$$

Антиградиент дает только направление спуска, но не величину шага. В общем случае один шаг не дает точку минимума, поэтому процедура спуска должна применяться несколько раз. В точке минимума все компоненты градиента равны нулю.

Все градиентные методы используют изложенную идею и отличаются друг от друга техническими деталями: вычисление производных по аналитической формуле или конечно-разностной аппроксимации; величина шага может быть постоянной, меняться по каким-либо правилам или выбираться после применения методов одномерной оптимизации в направлении антиградиента и т. д. и т. п.

Останавливаться подробно мы не будем, так как метод наискорейшего спуска не рекомендуется обычно в качестве серьезной оптимизационной процедуры.

Одним из недостатков этого метода является то, что он сходится к любой стационарной точке, в том числе и седловой, которая не может быть решением.

Но самое главное – очень медленная сходимость наискорейшего спуска в общем случае. Дело в том, что спуск является «наискорейшим» в локальном смысле. Если гиперпространство поиска сильно вытянуто («овраг»), то антиградиент направлен почти ортогонально дну «оврага», т. е. наилучшему направлению достижения минимума. В этом смысле прямой перевод английского термина «steepest descent», т. е. «спуск по наиболее крутому склону», более соответствует положению дел, чем термин «наискорейший», принятый в русскоязычной специальной литературе. Одним из выходов в этой ситуации является использование информации, даваемой вторыми частными производными. Другой выход – изменение масштабов переменных.

Метод сопряженного градиента Флетчера-Ривса

В методе сопряженного градиента строится последовательность направлений поиска \bar{S} , являющихся линейными комбинациями $-\nabla f(X^k)$ текущего направления наискорейшего спуска, и $\bar{S}^0, \bar{S}^1, \dots, \bar{S}^{k-1}$ предыдущих направлений поиска, т. е.

$$\bar{S}^{k+1} = -\nabla f(X^k) + \sum_{i=1}^k \alpha_i \bar{S}^i, k = 1, 2, \dots,$$

причем коэффициенты выбираются так, чтобы сделать направления поиска сопряженными. Доказано, что

$$\alpha_1 = \alpha_2 = \dots = \alpha_{k-1} = 0, \quad \alpha_k = \frac{\|\nabla f(X^k)\|^2}{\|\nabla f(X^{k-1})\|^2},$$

и это очень ценный результат, позволяющий строить быстрый и эффективный алгоритм оптимизации.

Алгоритм Флетчера-Ривса

1. В X^0 вычисляется $\bar{S}^0 = -\nabla f(X^0)$.
2. На k -м шаге с помощью одномерного поиска в направлении \bar{S}_k находится минимум $f(X)$, который и определяет точку X^{k+1} .
3. Вычисляются $f(X^{k+1})$ и $\nabla f(X^{k+1})$.
4. Направление \bar{S}^{k+1} определяется из соотношения

$$\bar{S}^{k+1} = -\nabla f(X^{k+1}) + \bar{S}^k \frac{\nabla^T f(X^{k+1})\nabla f(X^{k+1})}{\nabla^T f(X^k)\nabla f(X^k)}.$$

5. После $(n + 1)$ -й итерации (т. е. при $k = n$) производится рестарт: полагается $X^0 = X^{n+1}$ и осуществляется переход к шагу 1.
6. Алгоритм останавливается, когда $\|S^k\| < \varepsilon$, где ε – произвольная константа.

Преимуществом алгоритма Флетчера-Ривса является то, что он не требует обращения матрицы и экономит память ЭВМ, так как ему не нужны матрицы, используемые в ньютоновских методах, но в то же время почти столь же эффективен, как квазиньютоновские алгоритмы. Так как направления поиска взаимно сопряжены, то квадратичная функция будет минимизирована не более, чем за n шагов. В общем случае используется рестарт, который позволяет получать результат.

Алгоритм Флетчера-Ривса чувствителен к точности одномерного поиска, поэтому при его использовании необходимо устранять любые ошибки округления, которые могут возникнуть. Кроме того, алгоритм может отказать в ситуациях, где гессиан становится плохо обусловленным. Гарантии сходимости всегда и везде у алгоритма нет, хотя практика показывает, что почти всегда алгоритм дает результат.

Ньютоновские методы

Направление поиска, соответствующее наискорейшему спуску, связано с линейной аппроксимацией целевой функции. Методы, использующие вторые производные, возникли из квадратичной аппроксимации целевой функции, т. е. при разложении функции в ряд Тейлора отбрасываются члены третьего и более высоких порядков.

$$f(x) \approx f(X^k) + \nabla^T f(X^k)(X - X^k) + \frac{1}{2}(X - X^k)^T G(X^k)(X - X^k),$$

где $G(X^k)$ – матрица Гессе.

Минимум правой части (если он существует) достигается там же, где и минимум квадратичной формы. Запишем формулу для определения направления поиска \bar{S} :

$$f(X^k + \bar{S}) = f(X^k) + \nabla^T f(X^k)\bar{S} - \frac{1}{2}\bar{S}^T G(X^k)\bar{S}.$$

Минимум достигается при $G(X^k)\bar{S} = -\nabla f(X^k)$, ($\bar{S} = -G^{-1}(X^k)\nabla f(X^k)$).

Алгоритм оптимизации, в котором направление поиска определяется из этого соотношения, называется методом Ньютона, а направление \bar{S} – ньютоновским направлением.

В задачах поиска минимума произвольной квадратичной функции с положительной матрицей вторых производных метод Ньютона дает решение за одну итерацию независимо от выбора начальной точки.

Классификация ньютоновских методов

Собственно метод Ньютона состоит в однократном применении ньютоновского направления для оптимизации квадратичной функции. Если же функция не является квадратичной, то верна следующая теорема.

Теорема. *Если матрица Гессе нелинейной функции f общего вида в точке минимума X^* положительно определена, начальная точка X^0 выбрана достаточно близко к X^* и длины шагов подобраны верно, то метод Ньютона сходится к X^* с квадратичной скоростью.*

Метод Ньютона считается эталонным, с ним сравнивают все разрабатываемые оптимизационные процедуры. Однако метод Ньютона работоспособен только при положительно определенной и хорошо обусловленной матрице Гессе (определитель ее должен быть существенно больше нуля, точнее отношение наибольшего и наименьшего собственных чисел должно быть близко к единице). Для устранения этого недостатка используют *модифицированные методы* Ньютона, использующие ньютоновские направления по мере

возможности и уклоняющиеся от них только тогда, когда это необходимо.

Общий принцип модификаций метода Ньютона состоит в следующем: на каждой итерации сначала строится некоторая «связанная» с $G(X^k)$ положительно определенная матрица \bar{G}_k , а затем \bar{S} вычисляется по формуле $\bar{G}_k \bar{S} = -\nabla f(X^k)$. Так как \bar{G}_k положительно определена, то \bar{S} – обязательно будет направлением спуска. Процедуру построения \bar{G}_k организуют так, чтобы она совпадала с матрицей Гессе, если она является положительно определенной. Эти процедуры строятся на основе некоторых матричных разложений.

Другая группа методов, практически не уступающих по быстродействию методу Ньютона, основана на аппроксимации матрицы Гессе с помощью конечных разностей, так как не обязательно для оптимизации использовать точные значения производных. Эти методы полезны, когда аналитическое вычисление производных затруднительно или просто невозможно. Такие методы называются *дискретными методами Ньютона*.

Залогом эффективности методов ньютоновского типа является учет информации о кривизне минимизируемой функции, содержащейся в матрице Гессе и позволяющей строить локально точные квадратичные модели целевой функции. Но ведь возможно информацию о кривизне функции собирать и накапливать на основе наблюдения за изменением градиента во время итераций спуска. Соответствующие методы, опирающиеся на возможность аппроксимации кривизны нелинейной функции без явного формирования ее матрицы Гессе, называют *квазиньютоновскими методами*.

Отметим, что при построении оптимизационной процедуры ньютоновского типа (в том числе и квазиньютоновской) необходимо учитывать возможность появления седловой точки. В этом случае вектор наилучшего направления поиска будет все время направлен к седловой точке, вместо того, чтобы уходить от нее в направлении «вниз».

Метод Ньютона-Рафсона

Данный метод состоит в многократном использовании ньютоновского направления при оптимизации функций, не являющихся квадратичными.

Основная итерационная формула многомерной оптимизации $X^{k+1} = X^k + \lambda_{k+1} \cdot \bar{S}^{k+1}$ используется в этом методе при выборе направления оптимизации из соотношения $\bar{S}^{k+1} = -G^{-1}(X^{k+1}) \nabla f(X^{k+1})$, $\lambda_{k+1} = 1$.

Реальная длина шага скрыта в ненормализованном ньютоновском направлении \bar{S}^{k+1} . Так как этот метод не требует значения целевой функции в текущей точке, то его иногда называют непрямой или аналитическим методом оптимизации. Его способность определять минимум квадратичной функции за одно вычисление выглядит, на первый взгляд, исключительно привлекательно. Однако это «одно вычисление» требует значительных затрат. Прежде всего, необходимо вычислить n частных производных первого порядка и $n(n+1)/2$ – второго. Кроме того, матрица Гессе должна быть инвертирована. Это требует уже порядка n^3 вычислительных операций. С теми же самыми затратами методы сопряженных направлений или методы сопряженного градиента могут сделать порядка n шагов, т. е. достичь практически того же результата. Таким образом, итерация метода Ньютона-Рафсона не дает преимуществ в случае квадратичной функции.

Если же функция не квадратична, то:

- начальное направление уже, вообще говоря, не указывает действительную точку минимума, а значит, итерации должны повторяться неоднократно;

- шаг единичной длины может привести в точку с худшим значением целевой функции, а поиск может выдать неправильное направление, если, например, гессиан не является положительно определенным;

- гессиан может стать плохо обусловленным, что сделает невозможным его инвертирование, т. е. определение направления для следующей итерации.

Сама по себе стратегия не различает, к какой именно стационарной точке (минимума, максимума, седловой) приближается поиск, а вычисления значений целевой функции, по которым можно было бы отследить, не возрастает ли функция, не делаются. Значит, все зависит от того, в зоне притяжения какой стационарной точки оказывается стартовая точка поиска. Стратегия Ньютона-Рафсона редко используется сама по себе без модификации того или иного рода.

Методы Пирсона

Пирсон предложил несколько методов с аппроксимацией обратного гессиана без явного вычисления вторых производных, т. е. путем наблюдений за изменениями направления антиградиента. При этом получаются сопряженные направления. Эти алгоритмы отличаются только деталями. Приведем те из них, которые получили наиболее широкое распространение в прикладных областях.

Алгоритм Пирсона № 2

В этом алгоритме обратный гессиан аппроксимируется матрицей H^k , вычисляемой на каждом шаге по формуле

$$H^{k+1} = H^k + \frac{((X^{k+1} - X^k) - H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))) \cdot (X^{k+1} - X^k)^T}{(X^{k+1} - X^k)^T \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))}.$$

В качестве начальной матрицы H^0 выбирается произвольная положительно определенная симметрическая матрица.

Данный алгоритм Пирсона часто приводит к ситуациям, когда матрица H^k становится плохо обусловленной, а именно – она начинает осцилировать, колеблясь между положительно определенной и не положительно определенной, при этом определитель матрицы близок к нулю. Для избежания этой ситуации необходимо через каждые n шагов перезадавать матрицу, приравнивая ее к H^0 .

Алгоритм Пирсона № 3

В этом алгоритме матрица H^{k+1} определяется из формулы

$$H^{k+1} = H^k + \frac{[X^{k+1} - X^k - H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))] \cdot (H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k)))^T}{(\nabla f(X^{k+1}) - \nabla f(X^k))^T \cdot H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))}.$$

Траектория спуска, порождаемая алгоритмом, аналогична поведению алгоритма Дэвидона-Флетчера-Пауэлла, но шаги немного короче. Пирсон также предложил разновидность этого алгоритма с циклическим перезадаванием матрицы.

Проективный алгоритм Ньютона-Рафсона

Пирсон предложил идею алгоритма, в котором матрица рассчитывается из соотношения

$$H^{k+1} = H^k + \frac{(H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))) (H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k)))^T}{(\nabla f(X^{k+1}) - \nabla f(X^k))^T \cdot H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))},$$

$H^0 = R^0$, где матрица R^0 такая же, как и начальные матрицы в предыдущих алгоритмах.

Когда k кратно числу независимых переменных n , матрица H^k заменяется на матрицу R^{k+1} , вычисляемую как сумма

$$R^k + \frac{(X^{k+1} - X^k - R^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))) (H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k)))^T}{(\nabla f(X^{k+1}) - \nabla f(X^k))^T \cdot H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))}.$$

Величина $H^k \cdot (\nabla f(X^{k+1}) - \nabla f(X^k))$ является проекцией вектора приращения градиента $(\nabla f(X^{k+1}) - \nabla f(X^k))$, ортогональной ко всем векторам приращения градиента на предыдущих шагах. После каждых n шагов R^k является аппроксимацией обратного гессиана $H^{-1}(X^k)$, так что в сущности осуществляется (приближенно) поиск Ньютона.

Метод Дэвидона-Флетчера-Пауэла

Этот метод имеет и другие названия – метод переменной метрики, квазиньютоновский метод, так как он использует оба эти подхода.

Метод Дэвидона-Флетчера-Пауэла (ДФП) основан на использовании ньютоновских направлений, но не требует вычисления обратного гессиана на каждом шаге. Направление поиска на шаге k является направлением $-H_i \nabla f(X^i)$, где H_i – положительно определенная симметричная матрица, которая обновляется на каждом шаге и в пределе становится равной обратному гессиану. В качестве начальной матрицы H обычно выбирают единичную. Итерационная процедура ДФП может быть представлена следующим образом:

1. На шаге k имеются точка X^k и положительно определенная матрица H_k .

2. В качестве нового направления поиска выбирается

$$\bar{d}_k = -H_k \nabla f(X^k).$$

3. Одномерным поиском (обычно кубической интерполяцией) вдоль направления $X^k + \lambda \bar{d}_k$ определяется λ_k , минимизирующий функцию $f(X^k + \lambda \bar{d}_k)$.

4. Полагается $V_k = \lambda_k \cdot \bar{d}_k$.

5. Полагается $X^{k+1} = X^k + V_k$.

6. Определяется $f(X^{k+1})$ и $\nabla f(X^{k+1})$. Если $|V_k|$ или $|\nabla f(X^{k+1})|$ достаточно малы, процедура завершается.

7. Полагается $U_k = \nabla f(X^{k+1}) - \nabla f(X^k)$.

8. Матрица H_k обновляется по формуле

$$H_{k+1} = H_k + A_k + B_k,$$

$$A_k = \frac{V_k V_k^T}{V_k^T U_k}, \quad B_k = -\frac{H_k U_k U_k^T H_k}{U_k^T H_k U_k}.$$

9. Увеличить k на единицу и вернуться на шаг 2.

Метод эффективен на практике, если ошибка вычислений градиента $\nabla f(X^k)$ невелика и матрица H_k не становится плохо обусловленной.

Матрица A_k обеспечивает сходимость H_k к G^{-1} , матрица B_k обеспечивает положительную определенность H_{k+1} на всех этапах и в пределе исключает H_0 .

$$H_{k+1} = E + \sum_{i=1}^k A_i + \sum_{i=0}^k B_i.$$

В случае квадратичной функции $\sum_{i=1}^{n-1} A_i = G^{-1}$, $\sum_{i=0}^{n-1} B_i = -E$, т. е. алгоритм ДФП использует сопряженные направления.

Таким образом, метод ДФП использует как идеи ньютоновского подхода, так и свойства сопряженных направлений, и при минимизации квадратичной функции сходится не более чем за n итераций. Если оптимизируемая функция имеет вид, близкий к квадратичной функции, то метод ДФП эффективен за счет хорошей аппроксимации G^{-1} (метод Ньютона). Если же целевая функция имеет общий вид,

то метод ДФП эффективен за счет использования сопряженных направлений.

На практике оказалось, что метод ДФП может давать отрицательные шаги или окончиться в нестационарной точке. Это возможно, когда H_{k+1} становится плохо обусловленной. Этого можно избежать путем увеличения числа получаемых значимых цифр или перезадаванием матрицы H_{k+1} в виде специальной диагональной матрицы H , где

$$H_{ii}^{k+1} = \frac{X_i^{k+1} - X_i^k}{\nabla_i f(X^k)}.$$

Ошибки округления и особенно неточность линейного поиска может послужить причиной потери устойчивости метода ДФП и даже привести к ошибочным шагам, когда значение целевой функции на некоторой итерации возрастает, вместо того, чтобы уменьшаться.

По классификации ньютоновских методов данный алгоритм является квазиньютоновским, хотя он использует в явном виде только первые частные производные, а значит, может рассматриваться и как градиентный метод.

Практическая проверка эффективности алгоритма показала, что он столь же эффективен, как и метод Флетчера-Ривса.

Необходимые и достаточные условия безусловного экстремума

Рассмотрим функцию одной переменной $f(x)$.

Определение 1. Функция $f(x)$ имеет локальный минимум (максимум) в точке x_0 , если существует некоторая положительная величина δ , такая, что если $|x - x_0| < \delta$, то $f(x) \geq f(x_0)$, ($f(x) \leq f(x_0)$), т. е. существует окрестность точки x_0 , такая, что для всех значений x в этой окрестности $f(x)$ больше (меньше) $f(x_0)$.

Определение 2. Функция $f(x)$ имеет глобальный минимум (максимум) в точке x^* , если для всех x справедливо неравенство $f(x) \geq f(x^*)$ ($f(x) \leq f(x^*)$).

Классический подход к задаче нахождения значений x_0 , x^* состоит в поиске уравнений, которым они должны удовлетворять уравнению

$$f'(x) = 0 \quad (2)$$

удовлетворяют точки, в которых функция имеет минимум или максимум, а также точки перегиба функции. Следовательно, уравнение (2) является только необходимым условием минимума (максимума), но не является достаточным условием. Если $f''(x_0) > 0$, то в точке x_0 функция будет иметь минимум, если $f''(x_0) < 0$, то максимум. Если вторая производная равна нулю, ситуация остается неопределенной. Разрешить эту ситуацию позволяет следующая теорема.

Теорема. Если функция $f(x)$ и ее производные непрерывны, то точка x_0 является точкой экстремума (максимума или минимума) тогда и только тогда, когда порядок n ее первой, не обращающейся в нуль в точке x_0 производной, есть четное число. При этом, если $f^{(n)}(x_0) < 0$, то x_0 – точка максимума, если $f^{(n)}(x_0) > 0$, то x_0 – точка минимума.

Рассмотрим функцию n действительных переменных $f(x_1, x_2, \dots, x_n) = f(X)$.

Определение 3. Градиентом $\nabla f(x)$ непрерывно дифференцируемой функции $f(X)$ называется вектор-столбец, элементами которого являются частные производные первого порядка, вычисленные в данной точке.

Определение 4. Матрицей Гессе $G(X)$ дважды непрерывно дифференцируемой функции $f(X)$ называется матрица частных производных второго порядка, вычисленных в данной точке. Матрица Гессе является симметрической матрицей.

Для того чтобы матрица $G(X)$ была положительно определенной, необходимо и достаточно, чтобы все главные миноры матрицы были положительны. Для того чтобы матрица $G(X)$ была отрицательно определена, необходимо и достаточно, чтобы знаки главных миноров чередовались, начиная со знака минус.

Определение 5. Функция $f(X)$ имеет локальный минимум в точке X_0 , если существует окрестность точки X_0 , такая, что $f(X) > f(X_0)$ во всех точках этой окрестности, т. е. существует такое положительное значение $\delta > 0$, что для всех $\|X - X_0\| < \delta$ справедливо $f(X) > f(X_0)$.

Определение 6. Если $f(X) > f(X_0)$ для всех X , то X_0 – точка глобального минимума.

Необходимыми и достаточными условиями максимума функции $f(X)$ являются $\nabla f(X_0) = 0$, $G(X_0)$ – отрицательно определена.

Необходимыми и достаточными условиями минимума функции $f(X)$ являются $\nabla f(X_0) = 0$, $G(X_0)$ – положительно определена.

Пример 1. Исследовать на экстремум функцию $f = (x - 1)^6$.

Запишем необходимое условие экстремума первого порядка:

$$\frac{df(x)}{dx} = 6(x - 1)^5 = 0.$$

Отсюда стационарная точка $x_0 = 1$. Первая не равная нулю производная имеет порядок $n = 6$, $f^{(6)}(x) = 6! > 0$. Так как n четное, то функция в точке x_0 имеет минимум.

Пример 2. Исследовать на экстремум функцию

$$f(x) = 5x^6 - 36x^5 + 82,5x^4 - 60x^3 + 36.$$

Запишем необходимое условие экстремума первого порядка:

$$\frac{df(x)}{dx} = 30x^5 - 180x^4 + 330x^3 - 180x^2 = 30x^2(x - 1)(x - 2)(x - 3) = 0.$$

Отсюда получаем стационарные точки $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$. Проверим выполнение достаточных условий экстремума.

$$\frac{d^2 f(x)}{dx^2} = 150x^4 - 720x^3 + 990x^2 - 360x,$$

$$f''(x_1) = 0, f''(x_2) = 60 > 0, f''(x_3) = -120 < 0, f''(x_4) = 540 > 0.$$

Поэтому в точках x_2 , x_4 функция имеет минимум, а в точке x_3 – максимум. В точке x_1 достаточные условия не выполняются, поэтому вычислим третью производную:

$$f^{(3)} = 600x^3 - 2160x^2 + 1980x - 360,$$
$$f^{(3)}(x_1) = -360.$$

Так как эта производная отлична от нуля и имеет нечетный порядок, то в точке x_1 нет экстремума.

Пример 3. Исследовать на экстремум функцию

$$f = -x^2 - y^2 - z^2 - x + xy + 2z.$$

Найдем частные производные 1-го порядка:

$$\frac{\partial f}{\partial x} = -2x - 1 + y, \quad \frac{\partial f}{\partial y} = -2y + x, \quad \frac{\partial f}{\partial z} = -2z + 2.$$

Согласно необходимым условиям экстремума функции получаем систему уравнений:

$$\begin{cases} -2x - 1 + y = 0, \\ -2y + x = 0, \\ -2z + 2 = 0. \end{cases}$$

В результате решения системы получаем стационарную точку $\left(-\frac{2}{3}, -\frac{1}{3}, 1\right)$. Вычислим частные производные 2-го порядка:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial z^2} = -2, \quad \frac{\partial^2 f}{\partial x \partial y} = 1, \quad \frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial y \partial z} = 0.$$

Матрица Гессе в этом случае имеет вид

$$\begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Так как $\Delta_1 = -2$, $\Delta_2 = 3$, $\Delta_3 = -6$, т. е. знаки главных миноров чередуются, начиная с отрицательного, то в найденной стационарной точке функция имеет максимум $f\left(-\frac{2}{3}, -\frac{1}{3}, 1\right) = \frac{4}{3}$.

Пример 4. Исследовать на экстремум функцию

$$f(x, y) = x^3 + 3xy^2 - 39x - 36y + 26.$$

Найдем частные производные первого порядка:

$$\frac{\partial f}{\partial x} = 3x^2 + 3y^2 - 39, \quad \frac{\partial f}{\partial y} = 6xy - 36.$$

Согласно необходимым условиям экстремума функции получаем систему уравнений

$$\begin{cases} x^2 + y^2 = 13, \\ xy = 6. \end{cases}$$

Решив эту систему, найдем все стационарные точки – (3; 2), (–3; –2), (2; 3), (–2; –3). Вычислим частные производные второго порядка:

$$\frac{\partial^2 f}{\partial x^2} = 6x, \quad \frac{\partial^2 f}{\partial x \partial y} = 6y, \quad \frac{\partial^2 f}{\partial y^2} = 6x.$$

Матрица Гессе в данном случае имеет вид

$$\begin{pmatrix} 6x & 6y \\ 6y & 6x \end{pmatrix}.$$

Ее главные миноры Δ_1 и Δ_2 равны:

$$\Delta_1 = 6x, \quad \Delta_2 = \begin{vmatrix} 6x & 6y \\ 6y & 6x \end{vmatrix} = 36(x^2 - y^2).$$

В точке (3; 2) они положительны, следовательно, в этой точке функция имеет минимум $f(3; 2) = -100$. В точке (–3; –2) минор 1-го порядка отрицателен, 2-го порядка – положителен. Следовательно, в этой точке функция имеет максимум $f(-3; -2) = 152$. В точках (2; 3) и (–2; –3) минор 2-го порядка отрицателен, поэтому в этих стационарных точках экстремума нет.

Пример 5. Исследовать на экстремум функцию

$$f = 3x^3 + y^2 + z^2 + 6xy - 2z + 1.$$

Найдем частные производные 1-го порядка:

$$\frac{\partial f}{\partial x} = 9x^2 + 6y, \quad \frac{\partial f}{\partial y} = 2y + 6x, \quad \frac{\partial f}{\partial z} = 2z - 2.$$

Согласно необходимым условиям экстремума функции получаем систему уравнений

$$\begin{cases} 3x^2 + 2y = 0, \\ y + 3x = 0, \\ z - 1 = 0. \end{cases}$$

Найдем стационарные точки $(2; -6; 1)$, $(0; 0; 1)$. Вычислим частные производные 2-го порядка:

$$\frac{\partial^2 f}{\partial x^2} = 18x, \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial z^2} = 2, \quad \frac{\partial^2 f}{\partial x \partial y} = 6, \quad \frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial y \partial z} = 0.$$

Матрица Гессе в данном случае имеет вид

$$\begin{pmatrix} 18x & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

В точке $(2; -6; 1)$ ее главные миноры

$$\Delta_1 = 18x, \quad \Delta_2 = 36(x - 1), \quad \Delta_3 = 72(x - 1)$$

положительны. Следовательно, в этой точке функция имеет минимум

$$f(2; -6; 1) = -12.$$

Для исследования функции в точке $(0; 0; 1)$ нельзя использовать критерий Сильвестра, так как $\Delta_1 = 0$. Легко видеть, что в этой точке экстремума нет. В самом деле, $f(0; 0; 1) = 0$, а в сколь угодно малой окрестности точки $(0; 0; 1)$ функция принимает как положительные, так и отрицательные значения. Например, $f(\varepsilon; 0; 1) > 0$, если $\varepsilon > 0$, и $f(\varepsilon; 0; 1) < 0$, если $\varepsilon < 0$.

Задание

Исследовать функцию $z(x, y)$ на экстремум:

1. $z(x, y) = x^2 + xy + y^2 - 12x - 3y$.
2. $z(x, y) = 3 - x^2 + xy - y^2 + 2x - y$.
3. $z(x, y) = -x^2 - xy + y^2 + 3x + 6y$.
4. $z(x, y) = 3(x^2 + y^2) - x^3 + 4y$.
5. $z(x, y) = 2x^3 + xy^2 + 5x^2 + y^2$.
6. $z(x, y) = 3x^3 + y^3 - x - 3y^2 - 1$.
7. $z(x, y) = x^2y^2 - 2xy^2 - 6x^2y + 12xy$.

8. $z(x, y) = x^4 + y^4 - 2x^2$.
9. $z(x, y) = x^4 + y^4 - 2(x - y)^2$.
10. $z(x, y) = 2x^4 + y^4 - x^2 - 2y^2$.
11. $z(x, y) = xy^2(12 - x - y), x > 0, y > 0$.
12. $z(x, y) = ((x + y)/xy) - xy$.
13. $z(x, y) = x^3 + y^3 + 9xy$.
14. $z(x, y) = x^3 + 3xy^2 - 18x^2 - 18xy - 18y^2 + 57x + 138y + 290$.
15. $z(x, y) = x^2 + y^2 - 2x + 4y + 1$.
16. $z(x, y) = x^2 + xy + y^2 - 13x - 11y + 7$.
17. $z(x, y) = x^3 + y^3 - 6xy$.
18. $z(x, y) = x^3 - 7x^2 + xy - y^2 + 9x + 3y + 12$.
19. $z(x, y) = x^2 - xy + y^2 + 3x - 2y + 1$.
20. $z(x, y) = x^3 + y^3 - 3xy$.

Необходимые и достаточные условия условного экстремума

Условный экстремум при ограничениях типа равенств

Рассмотрим функцию $f(x) = f(x_1, x_2, \dots, x_n)$, на переменные которой наложены ограничения вида

$$\begin{aligned} g_1(x_1, x_2, \dots, x_n) &= 0, \\ g_2(x_1, x_2, \dots, x_n) &= 0, \\ g_m(x_1, x_2, \dots, x_n) &= 0 \end{aligned} \tag{3}$$

или в векторной форме $f(x), g_i(x) = 0; i = 1, 2, \dots, m; x = (x_1, x_2, \dots, x_n)$.

Уравнения (3) называются уравнениями связи.

Точку x^0 называют точкой условного максимума (минимума) функции $f(x)$ относительно уравнений связи (3), если существует такая окрестность точки x^0 , что для всех точек этой окрестности, удовлетворяющих уравнениям связи, верно неравенство $f(x) < f(x^0)$ ($f(x) > f(x^0)$).

Будем считать функции $f(x)$ и $g_i(x), i = 1, 2, \dots, m$, дважды непрерывно дифференцируемыми.

Определение. Функция $L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x)$ называется функцией Лагранжа, числа $\lambda_1, \lambda_2, \dots, \lambda_m$ называются множителями Лагранжа.

Из математического анализа известно, что точка условного минимума функции $f(x)$ совпадает с седловой точкой функции Лагранжа, при этом седловая точка должна обеспечивать минимум по переменным x_i , $i = 1, \dots, n$ и максимум по переменным λ_j , $j = 1, \dots, m$. Сформулируем необходимые и достаточные условия существования локального экстремума.

Необходимые условия для того, чтобы точка x^0 являлась точкой условного экстремума функции $f(x)$ при уравнении связи $g_i(x) = 0$, $i = 1, 2, \dots, m$.

Необходимо, чтобы x^0 удовлетворяла системе

$$\begin{cases} \frac{\partial L(x^0, \lambda)}{\partial x_k} = 0, & k = \overline{1, n}, \\ g_i(x^0) = 0, & i = \overline{1, m}. \end{cases} \quad (4)$$

Условие (2) можно записать в виде

$$\begin{cases} \frac{\partial L(x^0, \lambda)}{\partial x_k} = \frac{\partial f(x^0)}{\partial x_k} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(x^0)}{\partial x_k} = 0, & k = 1, \dots, n, \\ \frac{\partial L(x^0, \lambda)}{\partial \lambda_i} = g_i(x^0) = 0, & i = 1, \dots, m. \end{cases}$$

Достаточные условия

Пусть в точке x^0 выполняются необходимые условия существования экстремума функции $f(x)$ при ограничениях (3). Тогда, если при выполнении условий

$$dg_i(x^0) = \sum_{k=1}^n \frac{\partial g_i(x^0)}{\partial x^k} dx_k = 0 \quad (5)$$

второй дифференциал $d^2L(x^0, \lambda)$ функции Лагранжа является положительно (отрицательно) определенной квадратичной формой, то функция $f(x)$ в точке x^0 имеет условный минимум (максимум).

Если при условиях (5) второй дифференциал $d^2L(x^0, \lambda)$ является неопределенной квадратичной формой, то в точке x^0 экстремума нет.

Пример 1. Найти условный экстремум функции $f(x, y) = 6 - 5x - 4y$ относительно уравнения связи $g(x, y) = x^2 - y^2 - 9 = 0$.

Запишем функцию Лагранжа:

$$L(x, y, \lambda) = 6 - 5x - 4y + \lambda(x^2 - y^2 - 9).$$

Согласно необходимым условиям, получим систему

$$\begin{cases} \frac{\partial L}{\partial x} = -5 + 2\lambda x = 0, \\ \frac{\partial L}{\partial y} = -4 - 2\lambda y = 0, \\ x^2 - y^2 - 9 = 0, \end{cases}$$

из которой найдем стационарные точки $(-5; 4; -1/2)$, $(5; -4; 1/2)$. Таким образом, функция $f(x, y)$ может иметь условный экстремум только в двух точках: $(-5; 4)$ и $(5; -4)$.

Вычислим второй дифференциал функции Лагранжа:

$$\frac{\partial^2 L}{\partial x^2} = 2\lambda, \quad \frac{\partial^2 L}{\partial x \partial y} = 0, \quad \frac{\partial^2 L}{\partial y^2} = -2\lambda.$$

$$d^2 L = \frac{\partial^2 L}{\partial x^2} dx^2 + 2 \frac{\partial^2 L}{\partial x \partial y} dx dy + \frac{\partial^2 L}{\partial y^2} dy^2 = 2\lambda dx^2 - 2\lambda dy^2 = 2\lambda(dx^2 - dy^2).$$

Найдем первый дифференциал функции $g(x, y) = x^2 - y^2 - 9$:

$$dg(x, y) = 2x dx - 2y dy = 0, \quad x dx - y dy = 0.$$

В точке $(-5; 4)$ дифференциалы dx и dy связаны отношением

$$-5dx - 4dy = 0.$$

При выполнении этого условия второй дифференциал функции Лагранжа является положительно определенной квадратичной формой $d^2 L = \frac{9}{16} dx^2$ и, следовательно, в точке $(-5; 4)$ функция имеет условный минимум $f_{\min} = f(-5; 4) = 15$.

В точке $(5; -4)$ $d^2L = -\frac{9}{16}dx^2$ и, следовательно, в точке $(5; -4)$

функция имеет условный максимум $f_{\max} = f(5; -4) = -3$.

Пример 2. Найти условный экстремум функции $f(x_1, x_2) = 6 - 4x_1 - 3x_2$, относительно уравнений связи $g(x, y) = x_1^2 + x_2^2 - 1 = 0$.

Запишем функцию Лагранжа:

$$L(x_1, x_2, \lambda) = 6 - 4x_1 - 3x_2 + \lambda(x_1^2 + x_2^2 - 1).$$

Согласно необходимым условиям получим систему

$$\begin{cases} \frac{\partial L}{\partial x_1} = -4 + 2\lambda x_1 = 0, \\ \frac{\partial L}{\partial x_2} = -3 + 2\lambda x_2 = 0, \\ x_1^2 + x_2^2 - 1 = 0, \end{cases}$$

из которой найдем стационарные точки

$$\left(\frac{4}{5}, \frac{3}{5}\right), \lambda_1 = \frac{5}{2} \quad \text{и} \quad \left(-\frac{4}{5}, -\frac{3}{5}\right), \lambda_2 = -\frac{5}{2}.$$

Вычислим второй дифференциал функции Лагранжа:

$$\frac{\partial^2 L}{\partial x_1^2} = 2\lambda, \quad \frac{\partial^2 L}{\partial x_2^2} = 2\lambda, \quad \frac{\partial^2 L}{\partial x_1 \partial x_2} = 0.$$

$$d^2L(x_1, x_2, \lambda) = 2\lambda dx_1^2 + 2\lambda dx_2^2 = 2\lambda(dx_1^2 + dx_2^2).$$

При $\lambda_1 = \frac{5}{2}$, $d^2L > 0$ при любых dx, dy . Следовательно в точке

$\left(\frac{4}{5}, \frac{3}{5}\right)$ функция имеет условный минимум $f_{\min} = f\left(\frac{4}{5}, \frac{3}{5}\right) = 11$.

При $\lambda_2 = -\frac{5}{2}$, $d^2L < 0$ при любых dx, dy . Следовательно в точке

$\left(-\frac{4}{5}, -\frac{3}{5}\right)$ функция имеет условный локальный максимум

$$f_{\max} = f\left(-\frac{4}{5}, -\frac{3}{5}\right) = 1.$$

Задание

Найти условный экстремум функции относительно заданного уравнения связи:

1. $f(x, y) = 5 - 3x - 4y$, $x^2 + y^2 = 25$, $\min f(3, 4) = -20$,
 $\max f(-3, -4) = 30$.

2. $f(x, y) = 1 - 4x - 8y$, $x^2 - 8y^2 = 8$, $\min f(-4, 1) = 9$,
 $\max f(4, -1) = -7$.

3. $f(x, y) = x^2 + xy + y^2$, $x^2 + y^2 = 1$, $\min f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) = 1/2$,

$\max f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) = 3/2$.

4. $f(x, y) = 2x^2 + 12xy + y^2$, $x^2 + 4y^2 = 25$, $\min f(\pm 3, \pm 2) = -50$,
 $\max f(\pm 4, \pm 3/2) = 425/4$.

5. $f(x, y) = xy$, $x + y - 1 = 0$, $\max f(1/2, 1/2) = 1/4$.

6. $f(x, y) = x^2 + y^2$, $3x + 2y - 6 = 0$, $\min f(18/13, 12/13) = 36/13$.

7. $f(x, y) = x^2 - y^2$, $2x - y - 3 = 0$, $\max f(2, 1) = 3$.

8. $f(x, y) = xy^2$; $x + 2y - 1 = 0$, $\min f(1, 0) = 0$, $\max f(1/3, 1/3) = 1/27$.

9. $f(x, y) = \cos^2 x + \cos^2 y$, $x - y - \frac{\pi}{4} = 0$, $\min f\left(\frac{5\pi}{8}, \frac{3\pi}{8}\right) = 1 - \frac{1}{\sqrt{2}}$,

$\max f\left(\frac{\pi}{8}, -\frac{\pi}{8}\right) = 1 + \frac{1}{\sqrt{2}}$.

10. $f(x, y) = x^2 + y^2 - xy + x + y - 4$, $x + y + 3 = 0$,
 $\min f(-3/2, -3/2) = -19/4$.

11. $f(x, y) = 1/x + 1/y$, $x + y = 2$, $\min f(1, 1) = 2$.

12. $f(x, y) = \frac{x - y - 4}{\sqrt{2}}$, $x^2 + y^2 = 2$, $\min f\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = -1 - 2\sqrt{2}$,

$\max f\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = 1 - 2\sqrt{2}$.

13. $f(x, y) = 2x + y$, $x^2 + y^2 = 2$, $\min f\left(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right) = -\sqrt{5}$,

$\max f\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) = \sqrt{5}$.

14. $f(x, y) = x^2 + y^2$, $x + y = 2$, $\min f(1, 1) = 2$.

15. $f(x, y) = xy$, $x^2 + y^2 = 2a^2$, $\min f(\pm a, \pm a) = -a^2$, $\max f(\pm a, \pm a) = a^2$, ($a = 1$).

$$16. f(x, y) = 1/x + 1/y, \frac{1}{x^2} + \frac{1}{y^2} = \frac{1}{a^2}, \min f(-a\sqrt{2}, -a\sqrt{2}) = -\frac{\sqrt{2}}{a},$$

$$\max f(a\sqrt{2}, a\sqrt{2}) = \frac{\sqrt{2}}{a}, (a = 1).$$

$$17. f(x, y) = x^2 + y^2, 3x + 2y = 11, \min f(3, 1).$$

$$18. f(x, y) = xy, x^2 + y^2 = 2a^2, (a = 2) = 8.$$

$$19. f(x, y) = 1/x + 1/y, \frac{1}{x^2} + \frac{1}{y^2} = \frac{1}{4}, (a = 2).$$

$$20. f(x, y) = 2x^2 + 12xy + y^2, x^2 + y^2 = 25.$$

$$21. f(x, y) = 2x^2 + 3y^2, x + y = 1.$$

$$22. f(x, y) = 5x^2 + y^2, x + y = 1.$$

$$23. f(x, y) = x^2 + y^2, \frac{x}{2} + \frac{y}{3} = 1.$$

$$24. f(x, y) = x^2 + y^2, \frac{x}{3} + \frac{y}{4} = 1, \min f(4/7, 3/7).$$

$$25. f(x, y) = 1/x + 1/y, 2x + 3y = 1.$$

Условный экстремум при ограничениях типа неравенств

Рассмотрим задачу математического программирования: найти минимальное значение функции $f(x_1, \dots, x_n)$ при наличии ограничений $g_i(x_1, x_2, \dots, x_n) \leq b_i, i = 1, 2, \dots, m$.

Метод Лагранжа можно распространить и на случай, если ограничения заданы в виде неравенств. Ограничения в виде неравенств могут быть преобразованы в ограничения в виде равенств путем добавления к каждому из них неотрицательной ослабляющей переменной u_i^2 :

$$g_i(x) + u_i^2 - b_i = 0.$$

Таким образом, задача сводится к минимизации функции при наличии ограничений в виде равенства. Сформируем функцию Лагранжа

$$L(x, \lambda, u) = f(x) + \sum_{i=1}^m \lambda_i (g_i(x) + u_i^2 - b_i).$$

Необходимыми условиями, которые должны выполняться в стационарной точке, являются следующие:

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \quad j = 1, 2, \dots, n,$$

$$\frac{\partial L}{\partial \lambda_i} = g_i(x) + u_i^2 - b_i = 0, \quad i = 1, 2, \dots, m,$$

$$\frac{\partial L}{\partial u_i} = 2\lambda_i u_i = 0, \quad i = 1, 2, \dots, m.$$

Преобразуем последнее уравнение системы. Умножим его на $u_i/2$ и получим

$$\lambda_i u_i^2 = 0, \quad \lambda_i (b_i - g_i(x)) = 0.$$

Полученное уравнение означает, что либо $\lambda_i = 0$, либо $b_i - g_i(x) = 0$.

Если $\lambda_i \neq 0$, то $g_i(x) = b_i$ и ограничения являются активными и представляют собой ограничения в виде равенства. С другой стороны, если ограничения являются ограничениями в виде строгого неравенства $g_i(x) < b_i$, то соответствующий множитель Лагранжа $\lambda_i = 0$. Приведем необходимые условия минимума функции $f(x) = f(x_1, x_2, \dots, x_n)$ при наличии ограничений $g_i(x) \leq b_i, i = 1, 2, \dots, m$:

$$\frac{\partial f}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \quad j = \overline{1, n};$$

$$g_i(x) \leq b_i, \quad i = \overline{1, m};$$

$$\lambda_i (g_i(x) - b_i) = 0, \quad i = 1, 2, \dots, m;$$

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

Знак λ_i меняется на противоположный, если рассматривается максимум.

Пример 1. Найти условный минимум в задаче

$$f(x) = x_1^2 + x_2^2,$$

$$x_1 + x_2 - 2 \leq 0.$$

Составим функцию Лагранжа:

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

Выпишем необходимые условия существования условного минимума:

$$\begin{cases} 2x_1 + \lambda = 0; \\ 2x_2 + \lambda = 0; \\ x_1 + x_2 - 2 \leq 0; \\ \lambda(x_1 + x_2 - 2) = 0; \\ \lambda \geq 0. \end{cases}$$

Решим систему для двух случаев.

Если $\lambda = 0$, то решение системы в этом случае будет $x_1^* = 0, x_2^* = 0$, что является и решением неравенства $x_1 + x_2 - 2 \leq 0$.

В стационарной точке $(0, 0)$ проверим достаточные условия:

$$d^2L = 2dx_1^2 + 2dx_2^2 > 0,$$

значит в точке $(0, 0)$ функция имеет условный минимум $f(0, 0) = 0$, который совпал с безусловным минимумом этой функции.

Если $\lambda \neq 0$, то в этом случае $x_1 + x_2 - 2 = 0$. Решаем систему:

$$\begin{cases} 2x_1 + \lambda = 0; \\ 2x_2 + \lambda = 0; \\ x_1 + x_2 - 2 = 0. \end{cases}$$

Решение системы $x_1 = 1, x_2 = 1, \lambda = -2$. Так как $\lambda < 0$, то в этой точке не выполняются необходимые условия существования условного минимума.

Найти условный экстремум:

$$1. f = 3x_1^2 + 4x_1x_2 + 5x_2^2, x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \geq 4.$$

$$\text{Ответ: } x_1 = 3, x_2 = 1, f_{min} = 44.$$

$$2. f = x_1 + x_2, x_1^2 + x_2^2 - 1 \leq 0.$$

Ответ: $\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), f_{\min} = -\sqrt{2},$
 $\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right), f_{\max} = \sqrt{2}.$

Содержание отчета

1. Титульный лист.
 2. Цель и задачи работы.
 3. Результаты выполнения работы.
 4. Ответы на контрольные вопросы.
- Защитить отчет у преподавателя.

Контрольные вопросы

1. Перечислите основные методы оптимизации функций многих переменных.
2. При каких условиях функция $f(x)$ имеет локальный минимум (максимум) в точке x ?
3. При каких условиях функция $f(x)$ имеет глобальный минимум (максимум) в точке x ?
4. Назовите основные методы классической безусловной оптимизации.
5. Основное содержание методов прямого поиска.
6. Методы прямого случайного поиска.
7. Перечислите ньютоновские методы оптимизации.
8. Что такое градиент непрерывно дифференцируемой функции?
9. Что такое матрица Гессе (гессиан)?
10. Кратко изложите суть метода Лагранжа.

Практическая работа № 3

МЕТОДЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ В ОБЛАСТИ НАХОЖДЕНИЯ ОПТИМАЛЬНЫХ РЕШЕНИЙ ОПРЕДЕЛЕННЫХ ПРАКТИЧЕСКИХ ЗАДАЧ

Цель работы: получить практические навыки решения задач линейного программирования при нахождении рациональных значений целевой функции с использованием установленного критерия оптимизации.

Построение математических моделей задач линейного программирования

Линейное программирование – это раздел прикладной математики, посвященный методам нахождения наибольших или наименьших значений линейной функции многих переменных, т. е. функций вида

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n, \quad (1)$$

причем переменные x_j ($j = 1, 2, \dots, n$) должны удовлетворять дополнительным условиям, имеющим вид линейных уравнений

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i \quad (2)$$

или линейных неравенств вида

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i \quad (3)$$

или

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i, \quad (4)$$

где a_{ij}, b_i, c_j ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) – действительные числа.

Обычно в задачах линейного программирования на переменные налагаются еще условия неотрицательности:

$$x_j \geq 0 \quad (j = 1, 2, \dots, n). \quad (5)$$

Линейная функция z называется целевой функцией, или функцией цели, а дополнительные условия (1)–(5) называются ограничениями.

Линейное программирование широко используется в экономике, поскольку многие экономические задачи в математической формулировке сводятся к соотношениям (1)–(5), или, иначе говоря, соотношения (1)–(5) являются математической моделью экономической задачи.

Построение экономико-математической модели данной задачи выполняется в следующем порядке:

– вводятся переменные величины задачи x_j ($j = 1, 2, \dots, n$), т. е. такие величины, заданием числовых значений которых однозначно определяется один из вариантов исследуемого процесса;

– исходя из условий задачи, записываются ограничения, которым должны удовлетворять введенные переменные, при этом нужно следить, чтобы ни одно из условий не было упущено;

– составляется целевая функция, т. е. линейная функция переменных задачи, которая в математической форме выражает критерий выбора лучшего варианта.

Заметим, что переменные величины задачи могут быть введены не единственным образом, и от их выбора часто зависит сложность математической модели и, следовательно, удобство ее анализа. В простейших случаях, однако, выбор переменных естественным образом следует из условия задачи.

Рассмотрим примеры построения математических моделей линейного программирования.

Пример 1. Завод производит два вида продукции: велосипеды и мотоциклы. При этом цех по сборке велосипедов имеет мощность 100 тыс. шт. в год, цех по сборке мотоциклов – 30 тыс. шт. Одна группа механических цехов завода может производить либо детали для 120 тыс. велосипедов, либо детали для 40 тыс. мотоциклов, либо другую комбинацию деталей, ограниченную этими данными. Другая группа механических цехов может выпускать детали либо для 80 тыс. велосипедов, либо для 60 тыс. мотоциклов, либо любую допустимую их комбинацию. В результате реализации каждой тысячи велосипедов завод получает прибыль в 2 тыс. руб., а каждой тысячи мотоциклов – 3 тыс. руб.

Найти такое сочетание выпусков продукции, которое дает наибольшую сумму прибыли.

Решение. В качестве переменных задачи естественно взять количество велосипедов и мотоциклов, выпускаемых заводом в год, тыс. шт.: x_1 и x_2 .

Учитывая возможности сборочных цехов, необходимо потребовать, чтобы

$$x_1 \leq 100, \quad (6)$$

$$x_2 \leq 30. \quad (7)$$

Перейдем теперь к анализу возможностей цехов. При этом необходимо учитывать, что при выпуске обоих видов продукции должно выполняться условие пропорциональности количества продукции данного вида доле производственной мощности, занятой ее выпуском. Если предусматривается производство 1000 велосипедов (единицы продукции первого вида, т. е. $x_1 = 1$), то доля занятых производственных мощностей механических цехов первой группы составит $1/120$ суммы всех мощностей, принимаемой в данном случае за единицу; на выпуск же x_1 тыс. велосипедов потребуется занять $(1/120) x_1$ всей мощности. Аналогично для производства x_2 тыс. мотоциклов необходимо выделить $(1/40) x_2$ всей мощности. Так что для реализации плана (x_1, x_2) потребуется предусмотреть $((1/120) x_1 + (1/40) x_2)$ мощности механических цехов первой группы. Но в производственном процессе может быть использована не более чем вся наличная производственная мощность рассматриваемых цехов, следовательно

$$(1/120)x_1 + (1/40)x_2 \leq 1. \quad (8)$$

Точно так же получаем ограничения по производственной мощности механических цехов второй группы:

$$(1/80)x_1 + (1/60)x_2 \leq 1. \quad (9)$$

По смыслу задачи

$$x_1 \geq 0, x_2 \geq 0. \quad (10)$$

Любой план (x_1, x_2) , удовлетворяющий ограничениям (6)–(10), будет допустимым и даст предприятию прибыль, тыс. руб.

$$z = 2x_1 + 3x_2. \quad (11)$$

Соотношения (6)–(11) образуют математическую модель задачи:

$$z = 2x_1 + 3x_2 \rightarrow \max ,$$

$$\begin{cases} x_1 \leq 100, \\ x_2 \leq 30, \\ (1/120)x_1 + (1/40)x_2 \leq 1, \\ (1/80)x_1 + (1/60)x_2 \leq 1, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Итак, математически задача отыскания оптимального плана производства велосипедов и мотоциклов сводится к определению таких x_1^* и x_2^* , удовлетворяющих линейным ограничениям (6)–(10), которые доставляют максимум линейной функции (11).

Пример 2. Для строительства домов на 100 строительных площадках выбраны 5 типов проектов. По каждому из проектов известны длительность закладки фундаментов и строительства остальной части здания, а также жилая площадь дома (табл. 1). Параллельно можно вести закладку 10 фундаментов и строительство 15 зданий.

Таблица 1 – Исходные данные для решения задачи

Вид работы	Длительность выполнения для типового проекта, дней				
	1	2	3	4	5
Закладка фундамента	20	30	35	30	40
Остальные работы	40	20	60	35	25
Жилая площадь, м ²	3000	2000	5000	4000	6000

Составить план строительства, максимизирующий ввод жилой площади в течение года (300 раб. дней) при условии, что домов типа 2 должно быть построено не менее 10.

Решение. Обозначим через x_1, x_2, x_3, x_4, x_5 количества домов каждого типа, планируемых к строительству. По условию всего должно быть построено 100 домов. В принятых обозначениях этот факт можно выразить так:

$$x_1 + x_2 + x_3 + x_4 + x_5 = 100. \quad (12)$$

Поскольку одновременно можно вести закладку не более 10 фундаментов, то годовой фонд времени по этому виду работ ограничен

величиной $300 \cdot 10 = 3\,000$ раб. дней. Для реализации плана только на закладку фундаментов потребуется $(20x_1 + 30x_2 + 35x_3 + 30x_4 + 40x_5)$ раб. дней. Это количество не может превышать имеющегося фонда времени, предусмотренного для данного вида работ, поэтому должно выполняться неравенство

$$20x_1 + 30x_2 + 35x_3 + 30x_4 + 40x_5 \leq 3000. \quad (13)$$

Фонд времени на строительство остальной части зданий составляет $300 \cdot 15 = 4\,500$ раб. дней. На этот вид работ фактически будет потрачено $(40x_1 + 20x_2 + 60x_3 + 35x_4 + 25x_5)$ раб. дней. Ясно, что это количество не может превышать имеющегося резерва, т. е.

$$40x_1 + 20x_2 + 60x_3 + 35x_4 + 25x_5 \leq 4500. \quad (14)$$

И, наконец, учитывая последнее условие задачи, приходим к неравенству

$$x_2 \geq 10. \quad (15)$$

Остается присоединить естественные условия неотрицательности:

$$(x_j \geq 0 = 1, 2, \dots, 5). \quad (16)$$

В принятых обозначениях целевая функция задачи – вводимая в течение года жилая площадь – имеет вид

$$z = 3x_1 + 2x_2 + 5x_3 + 4x_4 + 6x_5 \rightarrow \max, \quad (17)$$

где жилая площадь исчисляется в тыс. кв. м.

Соотношения (12)–(17) служат математической моделью данной задачи.

Пример 3. При составлении суточного рациона кормления скота можно использовать свежее сено (не более 50 кг) и силос (не более 85 кг). Рацион должен содержать не более 30 кормовых единиц, 1 кг белка, 100 г кальция и 80 г фосфора. В таблице 2 приведены данные о содержании указанных компонентов в 1 кг каждого корма и себестоимости этих кормов.

Таблица 2 – Данные о содержании питательных компонентов в корме

Корм	Компонент, г/кг				Себестоимость
	Кол-во корм. ед.	Белок	Кальций	Фосфор	
Сено свежее	0,5	40	1,25	2	1,2
Силос	0,5	10	2,5	1	0,8

Определить оптимальный рацион, исходя из условий минимума его себестоимости.

Решение. Обозначим через x_1, x_2 соответственно количество сена и силоса, которые предполагается включить в рацион. Из условий задачи сразу следуют два ограничения:

$$x_1 \leq 50, \quad (18)$$

$$x_2 \leq 85. \quad (19)$$

Количество кормовых единиц, содержащихся в рационе (x_1, x_2) , можно выразить суммой

$$(0,5x_1 + 0,5x_2) \geq 30. \quad (20)$$

Ограничения по содержанию в рационе белка, кальция, фосфора примут следующий вид:

$$40x_1 + 10x_2 \geq 1000, \quad (21)$$

$$1,25x_1 + 2,5x_2 \geq 100, \quad (22)$$

$$2x_1 + x_2 \geq 80. \quad (23)$$

Естественно также, что

$$x_1 \geq 0, x_2 \geq 0. \quad (24)$$

В принятых обозначениях себестоимость рациона выражается следующей функцией:

$$z = 1,2x_1 + 0,8x_2, \quad (25)$$

которую необходимо минимизировать.

Математической моделью задачи являются соотношения (18)–(25).

Пример 4. Полосы листового проката длиной 200 см необходимо разрезать на заготовки трех типов: А, Б и В соответственно 57, 82 и 101 см для производства 50 изделий. На каждое изделие требуется по 4 заготовки типов А и Б и 5 заготовок типа В. Можно указать пять способов раскроя одной полосы. Количество заготовок, нарезаемых из одной полосы при каждом способе раскроя, приведено в таблице 3.

Таблица 3 – Количество заготовок

Способ	Количество заготовок типа		
	А	Б	В
1	3	–	–
2	2	1	–
3	1	–	1
4	–	2	–
5	–	–	1

Определить, какое количество полос проката нужно разрезать каждым способом для изготовления 50 изделий, чтобы отходы от раскроя были наименьшими.

Решение. Обозначим через x_j ($j = 1, 2, \dots, 5$) количество полос, раскраиваемых j -м способом.

Для производства 50 изделий необходимо $4 \cdot 50 = 200$ заготовок типа А, 200 – типа Б и $5 \cdot 50 = 250$ типа В. Если использовать все способы раскроя, то общее количество заготовок типа А при условии, что способом 1 раскроено x_1 полос, способом 2 – x_2 полос и т. д., можно выразить суммой

$$3x_1 + 2x_2 + 1x_3 + 0x_4 + 0x_5.$$

По условию эта сумма должна равняться 200:

$$3x_1 + 2x_2 + x_3 = 200. \quad (26)$$

Аналогично получаются условия выполнения задачи по другим типам заготовок:

$$x_2 + 2x_4 + x_5 = 200, \quad (27)$$

$$x_3 + x_5 = 250. \quad (28)$$

По смыслу задачи

$$x_j \geq 0 \quad (j = 1, 2, \dots, 5). \quad (29)$$

Чтобы составить целевую функцию, выражающую суммарную величину отходов, подсчитаем сначала величины отходов при раскрое одной полосы по каждому из способов. При первом способе отходы от каждой полосы составляют $200 - 57 \cdot 3 = 29$ (см), при втором способе $200 - (57 \cdot 2 + 82) = 4$ (см), при третьем, четвертом и пятом способах соответственно 42, 36 и 17 см.

Суммарную величину отходов (минимизируемую целевую функцию) можем теперь записать в виде

$$z = 29x_1 + 4x_2 + 42x_3 + 36x_4 + 17x_5. \quad (30)$$

Математическая модель задачи представлена соотношениями (26)–(30).

Пример 5. Найти оптимальное сочетание посевов пшеницы и кукурузы на участках различного плодородия площадью 100 и 200 га. Данные об урожайности приведены в таблице 4. По плану должно быть собрано не менее 1500 ц пшеницы и 4500 ц кукурузы. Цена 1 ц пшеницы – 400 руб., кукурузы – 250 руб. Критерий оптимальности – максимум валовой продукции в денежном выражении.

Таблица 4 – Данные об урожайности культур

Культура	Урожайность участка, ц/га	
	Участок 1	Участок 2
Пшеница	20	15
Кукуруза	35	30

Решение. Обозначим через x_1 площадь, отводимую под посев пшеницы на участке 1, через x_2 – на участке 2, через x_3 и x_4 – площади, отводимые под посев кукурузы соответственно на участках 1 и 2. Величины площадей выражаются неотрицательными числами, т. е.

$$x_j \geq 0 \quad (j = 1, 2, \dots, 4). \quad (31)$$

Так как на участке 1 планируется x_1 га засеять пшеницей и x_3 га – кукурузой, то должно выполняться равенство

$$x_1 + x_3 = 100. \quad (32)$$

Для участка 2 аналогичное условие запишется так:

$$x_2 + x_4 = 200. \quad (33)$$

С участка 1 предполагается собрать $20x_1$, а с участка 2 – $15x_2$ ц пшеницы. Всего же необходимо собрать не менее 1500 ц. Это требование можно выразить неравенством

$$20x_1 + 15x_2 \geq 1500. \quad (34)$$

Аналогичное требование к валовому сбору кукурузы приводит к неравенству

$$35x_3 + 30x_4 \geq 4500. \quad (35)$$

Стоимость пшеницы, которую предполагается собрать с обоих участков, составит $6(20x_1 + 15x_2)$ руб., стоимость кукурузы – $4(35x_3 + 30x_4)$ руб., а общая стоимость валовой продукции (максимизируемая функция цели) выразится суммой

$$z = 120x_1 + 90x_2 + 140x_3 + 120x_4. \quad (36)$$

Соотношения (31)–(36) служат математической моделью задачи.

Построить математические модели следующих задач линейного программирования.

1. В сплав может входить не менее 4 % никеля и не более 80 % железа. Для составления сплава используются три вида сырья, содержащего никель, железо и прочие вещества. Стоимость различных видов сырья и процентное содержание в нем соответствующих компонентов сплава представлены в таблице 5.

Таблица 5 – Исходные данные для решения задачи

Компоненты	Содержание компонентов для видов сырья, %		
	1	2	3
Железо	70	90	85
Никель	5	2	7
Прочие	25	8	8
Стоимость, коп/кг	6	4	5

Определить состав шихты таким образом, чтобы стоимость 1 кг сплава была минимальной.

2. Металлургический цех выпускает три вида продукции: А, Б, В. Прибыль от тонны произведенной продукции каждого вида составляет соответственно 35, 25 и 40 руб. Цех располагает необходимым оборудованием, каждый тип которого имеет свой фонд рабочего времени и производительность (табл. 6).

Таблица 6 – Фонд рабочего времени и производительность оборудования

Тип оборудования	Фонд времени, ч	Производительность по видам продукции, т/ч		
		А	Б	В
Печь обжига	2766	3,5	2,8	–
Травильный агрегат	624	0,083	0,083	0,104
Прокатный стан	416	0,067	0,1	0,083
Отделочный стан № 1	250	1	–	–
Отделочный стан № 2	1250	–	1	–
Отделочный стан № 3	1250	–	–	1

Составить план выпуска продукции, обеспечивающий максимум прибыли.

3. Предприятию задан план производства по времени и номенклатуре: требуется за 6 единиц времени выпустить 30 единиц продукции P_1 и 96 единиц продукции P_2 . Каждый из видов продукции может производиться машинами А и Б, значения мощностей которых и затраты, вызванные изготовлением каждого из видов продукции на той или иной машине, заданы в таблице 7.

Таблица 7 – Исходные данные для решения задачи

Машина	Мощность машины для вида продукции		Затраты на производство продукции	
	P_1	P_2	P_1	P_2
А	6	24	4	47
Б	13	13	13	26

Требуется составить оптимальный план работы машин, а именно: найти, сколько времени каждая из машин А и Б должна быть занята изготовлением каждого из видов продукции P_1 и P_2 , чтобы затраты на производство были минимальными при выполнении плана производства как по времени, так и по номенклатуре.

4. Совхоз отвел три земельных массива размерами в 5 000, 8 000 и 9 000 га под посеvy ржи, пшеницы и кукурузы соответственно средняя урожайность по массивам указана в таблице 8.

Таблица 8 – Урожайность культур

Культура	Средняя урожайность массива, ц/га		
	1	2	3
Рожь	12	14	15
Пшеница	14	15	22
Кукуруза	30	35	22

За один 1 ц ржи совхоз получает 2 руб. прибыли, за 1 ц пшеницы – 2,5 руб., за 1 ц кукурузы – 1,4 руб.

Сколько гектаров и на каких массивах совхоз должен отвести под каждую культуру, чтобы получить максимальную прибыль, если по плану он обязан сдать не менее 1 900 т ржи, 15 800 т пшеницы и 30 000 т кукурузы?

5. Три типа самолетов следует распределить между двумя авиалиниями. В таблице 9 заданы количество самолетов каждого типа, месячный объем перевозок каждым самолетом на каждой авиалинии и соответствующие расходы.

Таблица 9 – Исходные данные для решения задачи

Тип самолета	Число самолетов	Месячный объем перевозок одним самолетом по авиалиниям		Эксплуатационные расходы на один самолет по авиалиниям	
		1	2	1	2
		1	50	15	10
2	20	30	25	70	28
3	30	25	50	40	70

Требуется распределить самолеты по авиалиниям так, чтобы при минимальных суммарных эксплуатационных расходах перевезти по каждой из них соответственно не менее 300 и 200 единиц груза.

Задача минимизации (максимизации) функции

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

может быть заменена задачей максимизации (минимизации) функции

$$\bar{z} = -z = -c_1x_1 - \dots - c_nx_n,$$

поскольку $\min(z) = -\min(-z)$.

В первоначальной формулировке задача линейного программирования записывается, как правило, в общей форме, т. е. ограничения являются смешанными, они содержат как равенства, так и неравенства обоих видов.

Пример 1. Записать в канонической форме следующую задачу:

$$\begin{aligned} z = 3x_1 - 2x_2 - 5x_4 + x_5 &\rightarrow \max, \\ \begin{cases} 2x_1 + x_3 - x_4 + x_5 \leq 2, \\ x_1 - x_3 + 2x_4 + x_5 \leq 3, \\ 2x_2 + x_3 - x_4 + 2x_5 \leq 6, \\ x_1 + x_4 - 5x_5 \geq 8, \end{cases} \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{aligned}$$

Решение. Для перехода к канонической форме необходимо перейти от ограничений-неравенств к ограничениям-равенствам. Система ограничений состоит из четырех неравенств, поэтому надо ввести четыре дополнительные переменные. К первым трем неравенствам прибавляем по одной неотрицательной переменной, а из последнего – вычитаем. Система ограничений принимает вид

$$\begin{cases} 2x_1 + x_3 - x_4 + x_5 + x_6 = 2, \\ x_1 - x_3 + 2x_4 + x_5 + x_7 = 3, \\ 2x_2 + x_3 - x_4 + 2x_5 + x_8 = 6, \\ x_1 + x_4 - 5x_5 - x_9 = 8, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 9). \end{cases}$$

Целевая функция остается прежней.

Пример 2. Записать в канонической форме задачу

$$z = -x_1 + 2x_2 - x_3 + x_4 \rightarrow \min,$$

$$\begin{cases} 2x_1 - x_2 - x_3 + x_4 \leq 6, \\ x_1 + 2x_2 + x_3 - x_4 \geq 8, \\ 3x_1 - x_2 + 2x_3 + 2x_4 \leq 10, \\ -x_1 + 3x_2 + 5x_3 - 3x_4 = 15, \\ x_j \geq 0 \ (j = 1, 2, \dots, 4). \end{cases}$$

Решение:

$$z = -x_1 + 2x_2 - x_3 + x_4 \rightarrow \min,$$

$$\begin{cases} 2x_1 - x_2 - x_3 + x_4 + x_5 = 6, \\ x_1 + 2x_2 + x_3 - x_4 - x_6 = 8, \\ 3x_1 - x_2 + 2x_3 + 2x_4 + x_7 = 10, \\ -x_1 + 3x_2 + 5x_3 - 3x_4 = 15, \\ x_j \geq 0 \ (j = 1, 2, \dots, 7). \end{cases}$$

Пример 3. В следующей задаче перейти от канонической формы к стандартной:

$$z = x_1 + x_2 - 2x_3 + x_4 \rightarrow \max, \tag{45}$$

$$\begin{cases} x_1 + x_2 + x_3 - 0,5x_4 = 6, \\ 2x_1 - 3x_2 + 2x_3 = 4, \\ x_j \geq 0 \ (j = 1, 2, 3, 4). \end{cases} \tag{46}$$

Решение. Поскольку задача на максимум, то в стандартной форме ограничения должны быть неравенствами вида « \leq ». Воспользуемся первым способом перехода к неравенствам; задача принимает вид

$$z = x_1 + x_2 - 2x_3 + x_4 \rightarrow \max,$$

$$\begin{cases} x_1 + x_2 + x_3 - 0,5x_4 \leq 6, \\ -x_1 - x_2 - x_3 + 0,5x_4 \leq -6, \\ 2x_1 - 3x_2 + 2x_3 \leq 4, \\ -2x_1 + 3x_2 - 2x_3 \leq -4, \\ x_j \geq 0 \ (j = 1, 2, 3, 4). \end{cases}$$

Количество переменных осталось прежним, а число ограничивающих условий возросло вдвое.

Применим теперь второй способ перехода к неравенствам. Здесь требуется разрешить систему уравнений (46) относительно каких-либо двух переменных, т. е. преобразовать систему так, чтобы одна из переменных была только в первом уравнении, а другая – только во втором. Одна такая переменная уже есть, это x_4 в первом уравнении. Исключим теперь одну из остальных переменных, скажем x_3 , из первого уравнения, для чего умножим первое уравнение на 2 и вычтем из него второе, получаем

$$\begin{cases} 5x_2 - x_4 = 8, \\ 2x_1 - 3x_2 + 2x_3 = 4 \end{cases}$$

или

$$\begin{cases} x_4 = 5x_2 - 8, \\ x_3 = 2 - x_1 + \frac{3}{2}x_2. \end{cases} \quad (47)$$

С помощью этих равенств исключаем переменные x_3 и x_4 из целевой функции:

$$z = x_1 + x_2 - 4 + 2x_1 - 3x_2 + 5x_2 - 8 = 3x_1 + 3x_2 - 12.$$

Учитывая, что $x_3 \geq 0$ и $x_4 \geq 0$, из равенств (47) получаем следующие неравенства:

$$\begin{cases} 5x_2 - 8 \geq 0, \\ 2 - x_1 + \frac{3}{2}x_2 \geq 0. \end{cases}$$

Переносим свободные члены в этих неравенствах в правую часть и записывая неравенства со знаком « \leq », окончательно получаем стандартную форму данной задачи:

$$z = 3x_1 + 3x_2 - 12 \rightarrow \max,$$

$$\begin{cases} -5x_2 \leq -8, \\ x_1 - \frac{3}{2}x_2 \leq 2, \end{cases}$$

$$x_1, x_2 \geq 0.$$

Теперь задача содержит две переменные и два ограничивающих условия.

Перейти к канонической форме.

$$1. z = -2x_1 + x_2 + 5x_3 \rightarrow \max, \begin{cases} 4x_1 + 2x_2 + 5x_3 \leq 12, \\ 6x_1 - 3x_2 + 4x_3 = 18, \\ 3x_1 + 3x_2 - 2x_3 \geq 16, \\ x_j \geq 0 \ (j = 1, 2, 3). \end{cases}$$

$$2. z = 2x_1 - 5x_2 - 3x_3 \rightarrow \max, \begin{cases} -x_1 + x_2 + x_3 \geq 4, \\ 2x_1 - x_2 + x_3 \leq 1\zeta, \\ 3x_1 + x_2 + x_3 \geq 1\xi, \\ x_j \geq 0 \ (j = 1, 2, 3). \end{cases}$$

Перейти к стандартной форме.

$$3. z = x_1 - x_2 + 3x_4 + x_5 \rightarrow \min, \begin{cases} 2x_1 - x_2 - 3x_3 + x_4 - x_5 = 4, \\ x_1 + 2x_2 + 3x_3 + 3x_4 + x_5 = 15, \\ 2x_1 - 2x_2 - x_3 + 2x_5 = 8, \\ x_j \geq 0 \ (j = 1, 2, \dots, 5). \end{cases}$$

$$4. z = x_1 - 2x_2 + 2x_3 + x_4 + 2x_5 \rightarrow \max, \begin{cases} x_1 - x_2 + 2x_3 - 3x_4 + x_5 = -2, \\ 2x_1 + 3x_2 + x_4 - x_5 = 6, \\ -x_1 + 2x_2 - x_3 + 2x_4 = 3, \\ x_j \geq 0 \ (j = 1, 2, \dots, 5). \end{cases}$$

Графический метод решения задачи линейного программирования

Графическим методом можно решать задачу с двумя переменными с ограничениями в виде неравенств. Пусть требуется решить задачу

$$z = c_1x_1 + c_2x_2 \rightarrow \max, \tag{48}$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1, \\ a_{21}x_1 + a_{22}x_2 \leq b_2, \\ \dots\dots\dots, \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m, \end{cases} \tag{49}$$

$$x_1 \geq 0, x_2 \geq 0, \tag{50}$$

т. е. найти такие значения x_1 и x_2 , или, иначе говоря, такой вектор $X = (x_1, x_2)$, который удовлетворяет системе ограничений (49), (50) и максимизирует линейную функцию (48).

Каждое неравенство (49)

$$a_{i1}x_1 + a_{i2}x_2 \leq b_i$$

определяет в плоскости Ox_1x_2 полуплоскость с граничной прямой (рис. 2, а):

$$a_{i1}x_1 + a_{i2}x_2 = b_i,$$

т. е. точки (x_1, x_2) , лежащие в этой полуплоскости, удовлетворяют данному неравенству.

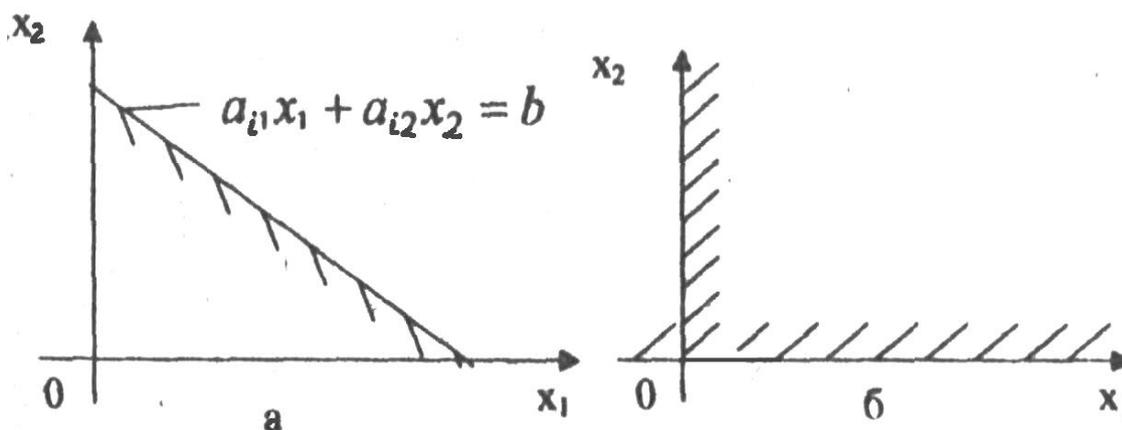


Рисунок 2 – Плоскость решения задачи

Неравенства (50) определяют соответственно правую и верхнюю полуплоскости (рис. 2, б).

Множество точек пересечения всех этих полуплоскостей образует многоугольник решений (область допустимых значений). Каждая точка этого многоугольника есть решение системы ограничений или план задачи линейного программирования. Многоугольник решений может быть и неограниченной многоугольной областью или вырождаться в прямую (луч, отрезок) или точку (рис. 3).

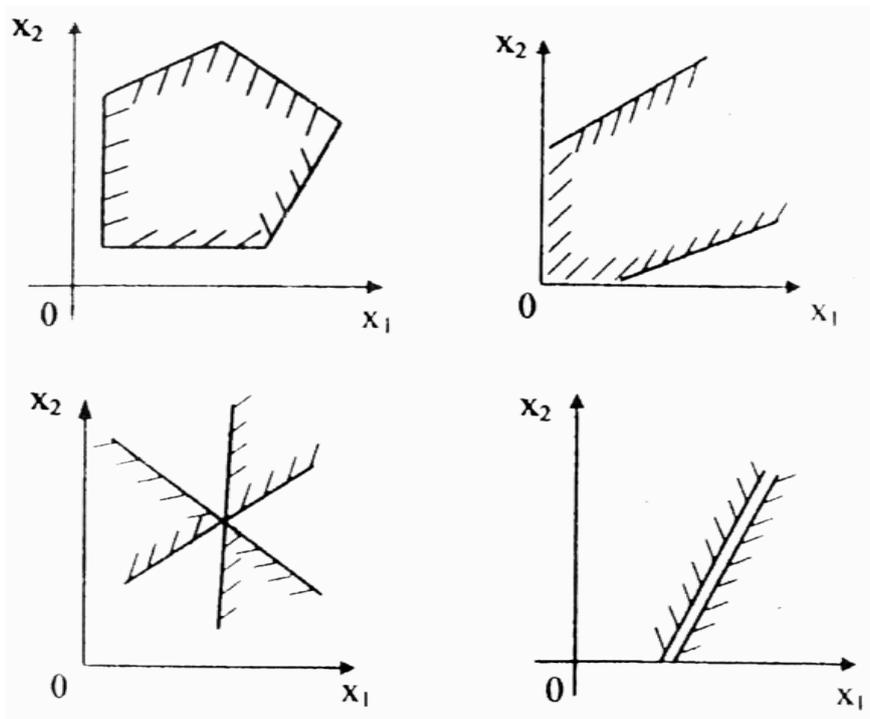


Рисунок 3 – Область решения задачи

Если система ограничений несовместна, то область допустимых значений пустая, т. е. полуплоскости, пересекаясь, не образуют общей для всех области.

Графическое решение задачи начинают с построения многоугольника решений. После построения многоугольника нужно найти такую его точку (x_1^*, x_2^*) в которой целевая функция z достигает максимума.

При фиксированном значении $z = z_0$ равенство (48) представляет собой уравнение прямой $c_1x_1 + c_2x_2 = z_0$, так называемой линии уровня, для всех точек которой функция принимает одно и то же значение z_0 . Если перемещать эту прямую в направлении вектора $N = (c_1, c_2)$, являющегося градиентом функции $z_0 = c_1x_1 + c_2x_2$, то соответствующее значение функции будет возрастать. В случае ограниченного многоугольника решений линия уровня при ее перемещении дважды становится опорной по отношению к многоугольнику. (Опорной прямой для многоугольника называется такая прямая, которая имеет хотя бы одну общую точку с многоугольником, а все остальные точки многоугольника лежат по одну сторону этой прямой). Одна из этих опорных прямых соответствует минимуму функции z , а другая – максимуму z в области допустимых решений (рис. 4). Соот-

ветствующая точка (x_1^*, x_2^*) многоугольника решений или вектор $X^* = (x_1^*, x_2^*)$ называется оптимальным планом, это и есть искомое решение задачи линейного программирования.

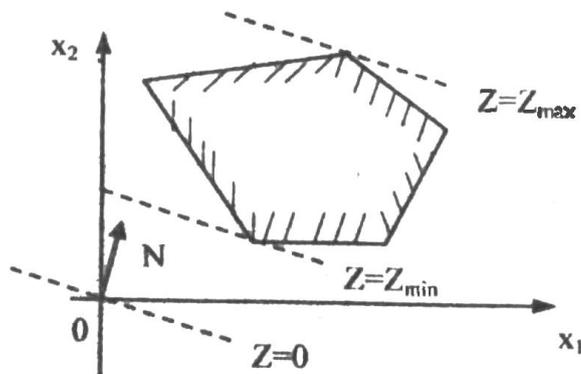


Рисунок 4 – Многоугольник решений

Очевидно, что опорная прямая обязательно проходит хотя бы через одну угловую точку многоугольника решений, поэтому оптимальному плану соответствует угловая точка. Линия уровня, соответствующая оптимальному плану, может совпасть со стороной многоугольника решений (см. рис. 4). В этом случае оптимальное решение не единственное – имеет место так называемый альтернативный оптимум.

Пусть векторы $X_1 = (x_1^{(1)}, x_2^{(1)})$ и $X_2 = (x_1^{(2)}, x_2^{(2)})$ – оптимальные планы, соответствующие угловым точкам многоугольника решений, тогда любой вектор

$$X = \lambda X_1 + (1 - \lambda) X_2, \quad (0 \leq \lambda \leq 1),$$

являющийся выпуклой линейной комбинацией векторов X_1 и X_2 , также будет оптимальным планом, соответствующим некоторой точке отрезка АВ.

Если многоугольная область допустимых решений не ограничена, то задача линейного программирования может иметь или не иметь решение (оптимальный план). Например, в случае, представленном на рисунке 5, задача на отыскание максимума не имеет решения, так как целевая функция не ограничена в области допустимых решений, а задача на отыскание минимума имеет решение.

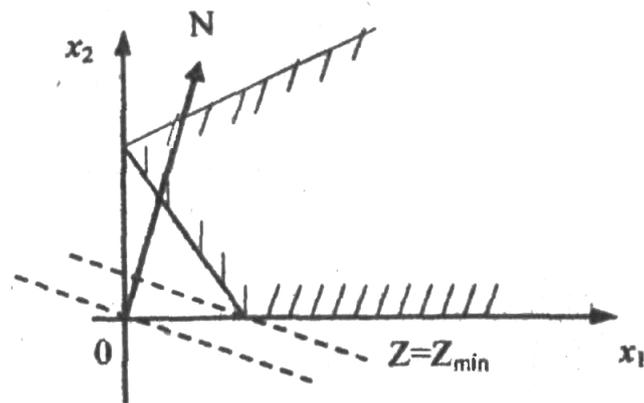


Рисунок 5 – Линия уровня

Пример 1. Графическим методом решить задачу

$$z = -2x_1 + 5x_2 \rightarrow \min, \quad (51)$$

$$\begin{cases} 7x_1 + 2x_2 \geq 14, \\ 5x_1 + 6x_2 \leq 30, \\ 3x_1 + 8x_2 \geq 24, \end{cases} \quad (52)$$

$$x_1 \geq 0, x_2 \geq 0. \quad (53)$$

Решение. Строим вначале многоугольник решений. Для этого надо в плоскости Ox_1x_2 построить все полуплоскости, определяемые неравенствами (52), (53), и выделить их общую область. Построим полуплоскость, определяемую первым неравенством (52). Сначала строим границу

$$7x_1 + 2x_2 = 14. \quad (54)$$

Полагая в этом уравнении $x_1 = 0$, находим $x_2 = 7$, затем, полагая $x_2 = 0$, находим $x_1 = 2$. По точкам $(0; 7)$ и $(2; 0)$ строим прямую (54) (рис. 6). Затем определяем, по какую сторону от этой прямой лежит область, задаваемая неравенством

$$7x_1 + 2x_2 \geq 14.$$

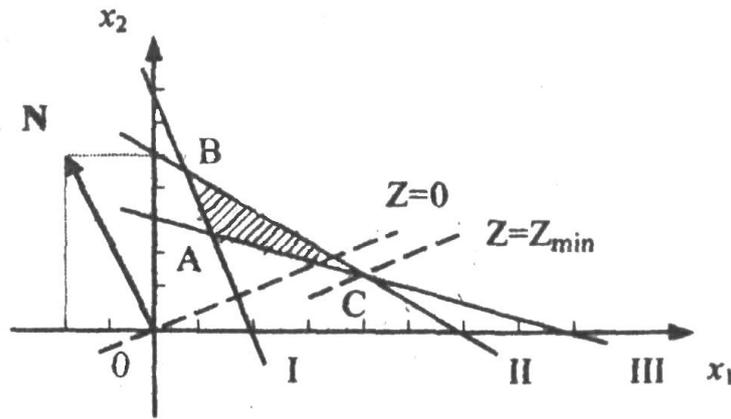


Рисунок 6 – Графическое решение (пример 1)

Это удобно сделать, подставив в левую часть неравенства точку $(0; 0)$. Получаем $0 \geq 14$, т. е. эта точка (начало координат) не удовлетворяет неравенству и, следовательно, определяемая этим неравенством область лежит справа от граничной прямой 1. Аналогично строим остальные полуплоскости. Их общая область образует треугольник ABC – многоугольник решений.

Чтобы построить линии уровня, строим из начала координат вектор $N = (-2; 5)$ и перпендикулярно ему проводим нулевую линию уровня. Поскольку решаем задачу на минимум, то смещаем линию уровня противоположно вектору N ; она становится опорной по отношению к треугольнику ABC в точке C. Координаты точки C и есть компоненты оптимального плана; они могут быть приближенно определены из графика или найдены точно путем решения системы уравнений прямых 1 и 2.

$$\begin{cases} 5x_1 + 6x_2 = 30, \\ 3x_1 + 8x_2 = 24. \end{cases}$$

Решая эту систему, находим $x_1 = \frac{48}{11}$, $x_2 = \frac{15}{11}$, т. е. оптимальный план

$$X^* = \left(\frac{48}{11}, \frac{15}{11} \right).$$

Этому плану отвечает оптимальное значение целевой функции

$$z_{\min} = -2\left(\frac{48}{11}\right) + 5\left(\frac{15}{11}\right) = -\frac{21}{11}.$$

Пример 2. Решить задачу

$$z = 3x_1 - 2x_2 \rightarrow \max,$$

$$\begin{cases} 7x_1 + 2x_2 \geq 14, \\ -x_1 + 2x_2 \geq 2, \\ 7x_1 + 10x_2 \leq 28, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Решение. На графике (рис. 7) многоугольник решений выродился в точку, найдем ее координаты. Из первых двух уравнений границ

$$\begin{cases} 7x_1 + 2x_2 = 14, \\ -x_1 + 2x_2 = 2 \end{cases}$$

находим $x_1 = \frac{3}{2}$, $x_2 = \frac{7}{4}$, подставим эти значения в третье уравнение:

$$7\frac{3}{2} + 10\frac{7}{4} = 28, \quad 28 = 28.$$

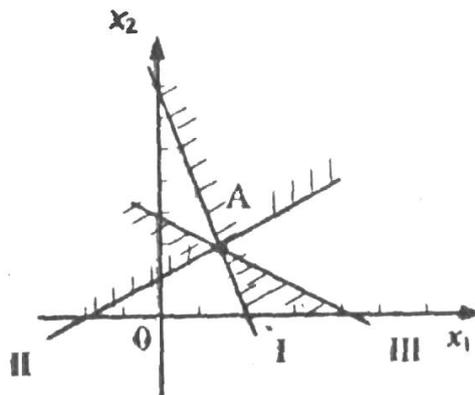


Рисунок 7 – Графическое решение (пример 2)

Уравнение обратилось в тождество, следовательно, множество допустимых решений действительно состоит из одной точки, т. е.

имеется единственный план задачи $X = (\frac{3}{2}, \frac{7}{4})$, которому отвечает значение целевой функции

$$z = 3(\frac{3}{2}) - 2(\frac{7}{4}) = 1.$$

Пример 3. Решить задачу:

$$\begin{aligned} z &= 2x_1 - 10x_2 \rightarrow \min, \\ \begin{cases} x_1 - x_2 \geq 0, \\ x_1 - 5x_2 \geq -5, \\ x_1 \geq 0, x_2 \geq 0. \end{cases} \end{aligned}$$

Решение. Линия уровня параллельна граничной прямой 2 (рис. 8), это следует из пропорциональности коэффициентов при переменных в целевой функции и во втором неравенстве системы ограничений. Поэтому минимальное значений достигается не только в угловой точке А, но и во всех точках прямой 2.

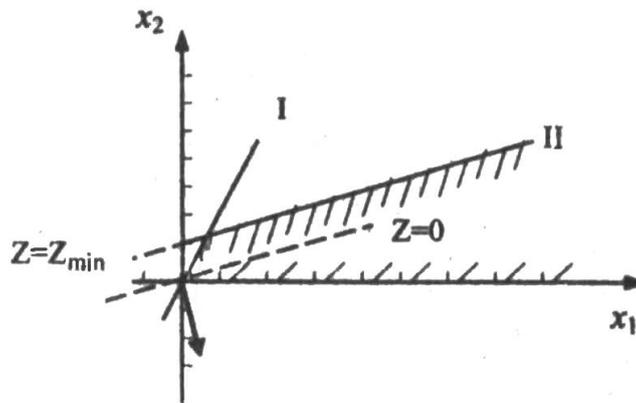


Рисунок 8 – Графическое решение (пример 3)

Точке А соответствует оптимальный план

$$X^* = (\frac{5}{4}, \frac{5}{4}).$$

Оптимальным будет также любой план:

$$X^* = (x, \frac{x}{5} + 1), x \geq \frac{5}{4}.$$

Каждому такому плану отвечает минимальное значение целевой функции

$$z_{\min} = 2x - 10\left(\frac{x}{5} + 1\right) = -10.$$

Пример 4. Решить задачу:

$$z = x_1 - 10x_2 \rightarrow \min ,$$

$$\begin{cases} x_1 - \frac{1}{2}x_2 \geq 0, \\ x_1 - 5x_2 \geq -5, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Линию уровня можно сколь угодно долго смещать противоположно вектору N (рис. 9), она всегда при этом будет пересекать область допустимых решений. Это значит, что целевая функция не ограничена снизу на множестве планов, т. е. задача не имеет оптимального плана (не имеет решения).

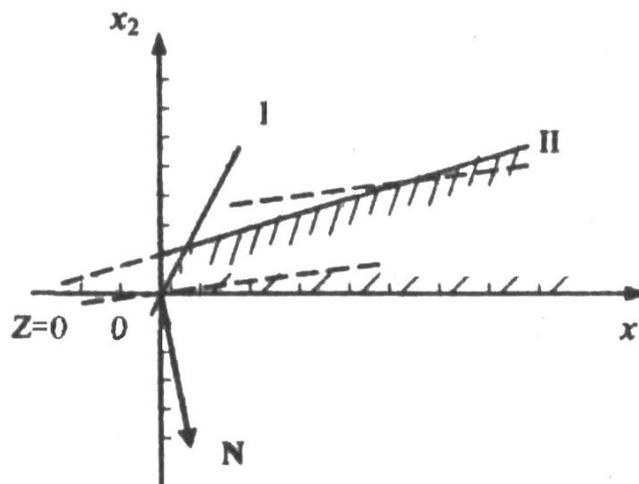


Рисунок 9 – Графическое решение (пример 4)

Пример 5. Решить задачу:

$$z = 2x_1 - 5x_2 \rightarrow \max,$$

$$\begin{cases} 4x_1 + 3x_2 \leq 12, \\ 3x_1 + 4x_2 \geq 12, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Решение

Определяемые ограничениями-неравенствами четыре полуплоскости не образуют общей области, значит система ограничений не совместна и задача не имеет решения (рис. 10).

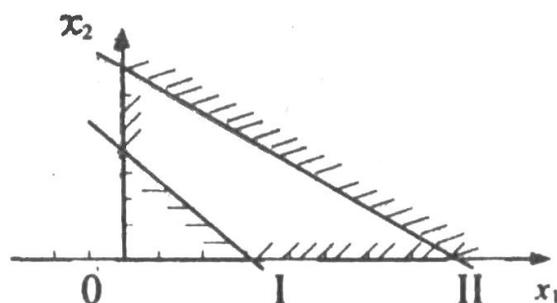


Рисунок 10 – Графическое решение (пример 5)

Графическим методом можно решить задачу линейного программирования с числом переменных и более двух, если ее можно свести к задаче в стандартной форме, содержащей не более двух переменных. Это всегда можно сделать, если система ограничений канонической формы задачи содержит m линейно независимых уравнений с n неизвестными и $n - m \leq 2$.

Пример 6. Записать в канонической форме задачу:

$$z = 2x_1 + x_2 + 6x_3 - 12x_4 - 9x_5 \rightarrow \max, \quad (55)$$

$$\begin{cases} x_1 + x_2 + 7x_3 - 3x_4 - 7x_5 = 13, \\ x_1 + 2x_2 + 13x_3 + 2x_4 - 14x_5 = 20, \\ x_1 + 3x_2 + 20x_3 + 6x_4 - 23x_5 = 19, \end{cases} \quad (56)$$

$$x_j \geq 0 \quad (j = 1, 2, \dots, 5). \quad (57)$$

Решение. Задача содержит $n = 5$ и $m = 3$ ограничений-равенств, поэтому есть возможность свести ее к стандартной задаче с двумя переменными (если уравнения-ограничения линейно зависимы). С этой целью разрешим систему (56) относительно первых трех переменных методом Жордана-Гаусса. Преобразования запишем в матричном виде:

$$\begin{pmatrix} 1 & 1 & 7 & -3 & -7 & | & 13 \\ 1 & 2 & 13 & 2 & -14 & | & 20 \\ 1 & 3 & 20 & 6 & -23 & | & 19 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 7 & -3 & -7 & | & 13 \\ 0 & 1 & 6 & 5 & -7 & | & 7 \\ 0 & 2 & 13 & 9 & -16 & | & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & -8 & 0 & | & 6 \\ 0 & 1 & 6 & 5 & -7 & | & 7 \\ 0 & 0 & 1 & -1 & -2 & | & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & -7 & 2 & | & 14 \\ 0 & 1 & 0 & 11 & 5 & | & 55 \\ 0 & 0 & 1 & -1 & -2 & | & -8 \end{pmatrix}.$$

Итак, преобразовали систему к виду

$$\begin{cases} x_1 - 7x_4 + 2x_5 = 14, \\ x_2 + 11x_4 + 5x_5 = 55, \\ x_3 - x_4 - 2x_5 = -8 \end{cases}$$

или

$$\begin{cases} x_1 = 14 + 7x_4 - 2x_5 \geq 0, \\ x_2 = 55 - 11x_4 - 5x_5 \geq 0, \\ x_3 = -8 + x_4 + 2x_5 \geq 0. \end{cases} \quad (58)$$

Исключим из целевой функции переменные x_1, x_2, x_3 , подставив в (55) их выражения из (58):

$$z = 18 + 14x_4 - 4x_5 + 55 - 11x_4 - 5x_5 - 48 + 6x_4 + 12x_5 - 12x_4 - 9x_5 = -35 - 3x_4 - 6x_5.$$

Таким образом, исходную задачу преобразовали к задаче

$$z = 35 - 3x_4 - 6x_5 \rightarrow \max, \quad (59)$$

$$\begin{cases} -7x_4 + 2x_5 \leq 14 \\ 11x_4 + 5x_5 \leq 55, \end{cases} \quad (60)$$

$$\begin{cases} -x_4 - 2x_5 \leq -8 \\ x_4 \geq 0, x_5 \geq 0, \end{cases} \quad (61)$$

содержащей только две переменные.

Задачу (59)–(61) решаем графическим методом (рис. 11).

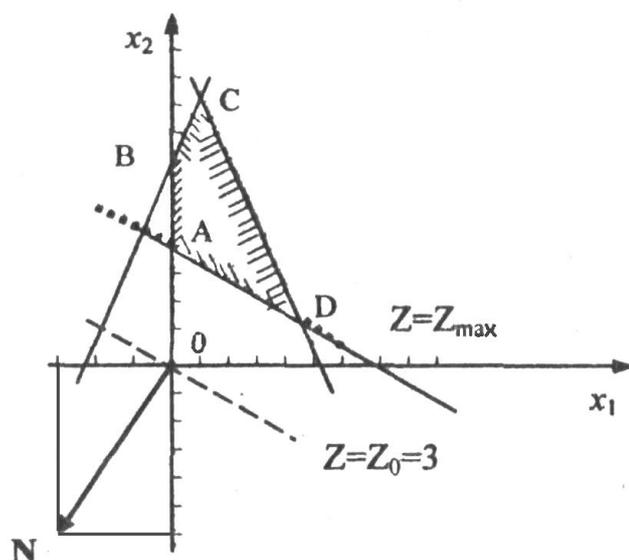


Рисунок 11 – Графическое решение (пример б)

Линия уровня параллельна стороне AD многоугольника решений, максимуму целевой функции соответствует опорная прямая, проходящая через точки A и D, т. е. задача (59)–(61) имеет множество решений (оптимальных планов). Возьмем план, отвечающий точке A:

$$X^* = (0; 4).$$

Компоненты плана исходной задачи (55)–(57) x_1, x_2, x_3 найдем, пользуясь равенствами (58)

$$x_1 = 6, x_2 = 35, x_3 = 0.$$

Оптимальный план исходной задачи

$$X^* = (6, 35, 0, 0, 0, 4).$$

Оптимальное значение целевой функции $z_{\max} = 11$.

Графическим методом решить следующие задачи:

$$1. z = 2x_1 + 3x_2 \rightarrow \max, \begin{cases} 3x_1 + 3x_2 \leq 6, \\ x_1 + x_2 \geq 1, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Ответ: $X^* = (0; 2), z_{\max} = 6$.

$$2. z = x_1 + 3x_2 \rightarrow \max, \begin{cases} x_1 - x_2 \leq 1, \\ 2x_1 + x_2 \geq 2, \\ x_1 - x_2 \geq 0, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Ответ: решения нет, $z \rightarrow \infty$.

$$3. z = 2x_1 + 3x_2 \rightarrow \max, \begin{cases} 3x_1 + 2x_2 \geq 6, \\ x_1 + 4x_2 \geq 4, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Ответ: $X^* = (8/5; 3/5), z_{\min} = 5$.

$$4. z = x_1 + 3x_2 \rightarrow \max, \begin{cases} -x_1 - x_2 \geq -3, \\ 2x_1 - 3x_2 \leq 6, \\ x_1 + x_2 \geq 4, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Ответ: решения нет, система ограничений несовместна.

$$5. z = x_1 + x_2 \rightarrow \max, \begin{cases} x_1 + 2x_2 \leq 10, \\ x_1 + 2x_2 \geq 2, \\ 2x_1 + x_2 \leq 10, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Ответ: $X^* = (10/3; 10/3), z_{\max} = 20/3$.

$$6. z = -4x_1 + 3x_2 + x_4 - x_5 \rightarrow \max, \begin{cases} 2x_1 - x_2 - x_3 = -1, \\ x_1 - 3x_2 - x_4 = -13, \\ 4x_1 + x_2 + x_5 = 26, \\ x_1 - 3x_2 + x_6 = 0, \\ x_1 + x_2 - x_7 = 4, \\ x_j \geq 0 (j=1, 2, \dots, 7). \end{cases}$$

Ответ: $X^* = (5; 6; 5; 0; 0; 13; 7), z_{\max} = -2$.

$$7. z = x_1 + x_2 \rightarrow \max, \begin{cases} 4x_1 - 3x_2 - x_3 + x_4 + x_5 = 6, \\ x_1 + 4x_2 + x_3 + x_5 = 15, \\ 2x_1 - 4x_2 - x_3 + x_4 = -3, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 5). \end{cases}$$

Ответ. Решение не единственное:

$$X_1^* = (3; 0; 9; 0; 3), X_2^* = (4; 1; 7; 0; 0), X^* = \lambda X_1^* + (1-\lambda)X_2^*, (0 \leq \lambda \leq 1), z_{\max} = 3.$$

Симплексный метод

Симплексный метод является основным методом решения задач линейного программирования. Он применяется к задаче, записанной в канонической форме:

$$z = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \rightarrow \min(\max), \quad (62)$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\ \dots, \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m, \end{cases} \quad (63)$$

$$x_j \geq 0 \quad (j = 1, 2, \dots, n), \quad (64)$$

где $b_i \geq 0$ ($i = 1, 2, \dots, m$).

Система ограничений (63) может быть представлена в векторной форме:

$$x_1 A_1 + x_2 A_2 + \dots + x_n A_n = A_0, \quad (65)$$

где $A_1, A_2, \dots, A_n, A_0$ – векторы:

$$A_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}, \dots, A_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix}, A_0 = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}.$$

Равенство (65) можно рассматривать как разложение вектора A_0 по векторам A_1, A_2, \dots, A_n . В этом случае переменные x_1, x_2, \dots, x_n , или

компоненты вектора $X = (x_1, x_2, \dots, x_n)$, суть коэффициенты разложения вектора A_0 .

Напомним, что вектор $X = (x_1, x_2, \dots, x_n)$ называется планом, или допустимым решением задачи линейного программирования, если он удовлетворяет условиям (63) и (64). План называется опорным, если положительным (ненулевым) компонентам вектора X соответствуют линейно независимые векторы A_j в разложении (65). Опорный план называется невырожденным, если он содержит m ненулевых компонент, в противном случае план называется вырожденным. План, доставляющий наименьшее (наибольшее) значение линейной функции (62), называется оптимальным планом, или решением задачи линейного программирования.

Множество планов данной задачи образует многогранник в m -мерном пространстве, называемый многогранником решений. В теории линейного программирования доказывается, что если целевая функция ограничена на многограннике решений, то существует такая угловая точка (вершина) многогранника, в которой целевая функция достигает своего наименьшего (наибольшего) значения, и что каждой угловой точке соответствует опорный план. Следовательно, для отыскания оптимального плана достаточно исследовать только опорные планы.

Симплексный метод позволяет переходить от одного опорного плана к другому, более близкому к оптимальному. Геометрически симплекс-метод можно трактовать, как перемещение по ребрам многогранника решений от вершины к вершине в сторону оптимального плана. Симплексный метод можно применить, если известен какой-нибудь исходный опорный план. Такой опорный план можно непосредственно записать, если в разложении (65) имеется m линейно независимых единичных векторов A_j . Пусть, например,

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}, \dots, A_n = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix},$$

т. е. система ограничений (63) имеет вид

$$z_j - c_j \leq 0 \quad (j = 1, 2, \dots, n), \quad (69)$$

то план является оптимальным. При решении задачи на максимум условие оптимальности – неотрицательность всех оценок:

$$z_j - c_j \geq 0 \quad (j = 1, 2, \dots, n). \quad (70)$$

Если условие оптимальности нарушено хотя бы для одной из оценок, то необходимо в соответствующем столбце просмотреть величины a_{ij} : если все $a_{ij} \leq 0$ хотя бы для одного из таких столбцов, то задача линейного программирования не имеет решений (целевая функция не ограничена в области допустимых решений).

Если же в столбце, где нарушено условие оптимальности, есть положительные величины a_{ij} , то нужно перейти к другому опорному плану. Перейти к новому опорному плану – значит разложить вектор A_0 по другому базису. Новый базис можно получить, выведя из старого базиса один из векторов и введя в него один из небазисных векторов. В принципе, можно вводить в базис любой из векторов, однако для целенаправленного и наиболее быстрого движения к оптимуму применяется следующий алгоритм выбора вектора, вводимого в базис:

1. Для каждого столбца с нарушением условия оптимальности вычисляется величина

$$\theta_{0j} = \min_i \frac{b_j}{a_{ij}},$$

где минимум берется по тем i , для которых $a_{ij} \geq 0$. Соответствующее значение a_{ij} каким-либо образом отмечается в таблице.

2. Вычисляются величины $\theta_{0j}(z_j - c_j)$ и определяется

$$\max_j [\theta_{0j}(z_j - c_j)] \quad (71)$$

при решении задачи на минимум и

$$\min_j [\theta_{0j}(z_j - c_j)] \quad (72)$$

при решении задачи на максимум. Элемент a_{ij} , отмеченный на предыдущем шаге и стоящий в столбце, соответствующем условию (71) или (72), выделяется в качестве ведущего элемента; строка и столбец, содержащие ведущий элемент, называются направляющими. Если имеется несколько одинаковых максимальных значений $\theta_{0j}(z_j - c_j)$ при решении задачи на минимум, то в соответствующих столбцах ведущим элементом выбирается отмеченный элемент в столбце, которому соответствует наименьшее c_j . В аналогичном случае при решении задачи на максимум ведущим принимается отмеченный элемент в столбце, которому соответствует наибольшее c_j .

3. С учетом выбранного ведущего элемента выполняется преобразование по методу полного исключения над всеми элементами столбцов $A_0, A_1, A_2, \dots, A_n$, включая строку $(m+1)$ -ю, и заполняется новая симплексная таблица (табл. 11). При этом вектор A_i направляющей строки выводится из базиса, а вектор A_j направляющего столбца становится базисным.

Правильность вычислений проверяется пересчетом элементов $(m+1)$ -й строки по формулам (67), (68).

В столбце A_θ стоят компоненты нового опорного плана

$$X_1 = (b'_1, b'_2, \dots, 0, \dots, b'_m, 0, \dots, b'_k, 0, \dots, 0).$$

Если для нового опорного плана выполнены условия оптимальности, то решение задачи на этом заканчивается. Если условия оптимальности не выполнены, то, как и на предыдущем шаге, выясняем, не оказалась ли задача неразрешимой, и если нет, то переходим к следующему опорному плану, повторяя шаги 1–3.

Как известно из теории, после конечного числа таких этапов (итераций) мы либо приходим к оптимальному плану, либо убеждаемся в том, что задача не имеет решения. (Мы пренебрегаем возможностью так называемого заикливания ввиду того, что оно практически не встречается в реальных задачах.) Если в симплексной таблице с оптимальным опорным планом нулевые оценки соответствуют только базисным векторам, то оптимальный план единственный. Если ненулевые оценки есть у небазисных векторов, то решение не единственное.

Таблица 10 – Первый опорный план

I	Базис	C_{σ}	A_0	C_1	C_2	...	C_r	...	C_m	C_{m+1}	...	C_k	...	C_n
				A_1	A_2	...	A_r	...	A_m	A_{m+1}	...	A_k	...	A_n
1	A_1	C_1	b_1	1	0	...	0	...	0	a_{1m+1}	...	a_{1k}	...	a_{1n}
2	A_2	C_2	b_2	0	1	...	0	...	0	a_{2m+1}	...	a_{2k}	...	a_{2n}
...
r	A_r	C_r	b_r	0	0	...	1	...	0	a_{rm+1}	a_{rn}
...
m	A_m	C_m	b_m	0	0	...	0	...	1	a_{mm+1}	...	a_{mk}	...	a_{mn}
m+1	$z_j - c_j$		z_0	0	0	...	0	...	0	Δ_{m+1}	...	Δ_k	...	Δ_n

Таблица 11 – Второй опорный план

I	Базис	C_{σ}	A_0	C_1	C_2	...	C_r	...	C_m	C_{m+1}	...	C_k	...	C_n
				A_1	A_2	...	A_r	...	A_m	A_{m+1}	...	A_k	...	A_n
1	A_1	C_1	b'_1	1	0	...	a'_{1r}	...	0	a'_{1m+1}	...	0	...	a'_{1n}
2	A_2	C_2	b'_2	0	1	...	a'_{2r}	...	0	a'_{2m+1}	...	0	...	a'_{2n}
...
r	A_k	C_k	b'_r	0	0	...	a'_{rr}	...	0	a'_{rm+1}	...	1	...	a'_{rn}
...
m	A_m	C_m	b'_m	0	0	...	a'_{mr}	...	1	a'_{mm+1}	...	0	...	a'_{mn}
m+1	$z_j - c_j$		z'_0	0	0	...	Δ'_r	...	0	Δ'_{m+1}	...	0	...	Δ'_n

Пример 1. Симплексным методом решить следующую задачу линейного программирования:

$$z = x_1 + x_2 + x_3 \rightarrow \min,$$

$$\begin{cases} x_1 - x_4 - 2x_6 = 5, \\ x_2 + 2x_4 - 3x_5 + x_6 = 3, \\ x_3 + 2x_4 - 5x_5 + 6x_6 = 5, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases}$$

Решение. Задача записана в канонической форме, система ограничений имеет положительные правые части и содержит три единичных линейно независимых вектора A_1, A_2, A_3 , поэтому можем заполнить первую симплексную таблицу (табл. 12).

Таблица 12 – Первый опорный план (пример 1)

I	Базис	$C_{\bar{b}}$	A_0	1	1	1	0	0	0
				A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	5	1	0	0	-1	0	-2
2	A_2	1	3	0	1	0	2	-3	1
3	A_3	1	5	0	0	1	2	-5	6
$m+1$	$z_j - c_j$		13	0	0	0	3	-8	5

Вносим в нее исходные данные задачи и вычисляем элементы $(m+1)$ -й строки.

Значение z_0 вычисляется следующим образом: элементы столбца $C_{\bar{b}}$ умножаются на соответствующие элементы столбца A_0 (компоненты вектора A_0), и эти произведения суммируются:

$$z_0 = 1 \cdot 5 + 1 \cdot 3 + 1 \cdot 5 = 13.$$

Значения оценок $\Delta_j = z_j - c_j$ вычисляются аналогично, но из суммы произведений вычитается соответствующее значение c_j , записанное сверху в шапке таблицы, например:

$$\Delta_1 = 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 0 = 1$$

и т. д.

$$\Delta_6 = 1 \cdot (-2) + 1 \cdot 1 + 1 \cdot 6 - 0 = 5.$$

Среди оценок есть положительные ($\Delta_4 = 3$ и $\Delta_6 = 5$), значит первоначальный опорный план не оптимальный и необходимо перейти к новому опорному плану. Такой переход возможен, так как в столбцах A_4 и A_6 , где нарушено условие оптимальности, есть положительные элементы a_{ij} . Находим $\theta_{04} = \min(\frac{3}{2}, \frac{5}{2}) = \frac{3}{2}$ и отмечаем элемент $a_{24} = 2$, дающий этот минимум, а также находим $\theta_{06} = \min(\frac{3}{1}, \frac{5}{6}) = \frac{5}{6}$ и отмечаем элемент $a_{36} = 6$.

Затем вычисляем

$$\theta_{04}(z_4 - c_4) = \frac{3}{2} \cdot 3 = \frac{9}{2}$$

и

$$\theta_{06}(z_6 - c_6) = \frac{5}{6} \cdot 5 = \frac{25}{6}.$$

Поскольку решается задача на минимум и $\frac{9}{2} \geq \frac{25}{6}$, то в базис надо ввести вектор A_4 , а так как ведущий элемент стоит во второй строке (взят в рамочку в таблице 13), то из базиса выводится вектор A_2 .

Таблица 13 – Второй опорный план (пример 1)

I	Базис	C_b		1	1	1	0	0	0
			A_0	A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	13/2	1	1/2	0	0	-3/2	-3/2
2	A_4	0	3/2	0	1/2	0	1	-3/2	1/2
3	A_3	1	2	0	-1	1	0	-2	5
$m + 1$	$z_j - c_j$		17/2	0	-3/2	0	0	-7/2	7/2

Далее расчеты по заполнению строк новой таблицы производим в соответствии с методом треугольника. Для вычисления какого-

нибудь из элементов столбцов таблицы предварительно находим три числа:

1) число, расположенное в исходной таблице на месте искомого элемента новой симплекс-таблицы;

2) число, расположенное в исходной симплекс-таблице на пересечении строки, в которой находится искомый элемент новой симплекс-таблицы, и столбца, соответствующего вектору, вводимому в базис;

3) число, расположенное в новой симплекс-таблице на пересечении столбца, в котором стоит искомый элемент, и строки вновь вводимого в базис вектора. Эта строка получается из строки исходной, первой таблицы делением ее элементов на разрешающий элемент.

Окончательно, для определения искомого элемента новой симплекс-таблицы, из первого числа вычитают произведение второго и третьего.

Заполнив вторую симплексную таблицу, пересчитываем элементы $(m + 1)$ -й строки по формулам (67), (68):

$$z'_0 = 1 \cdot \frac{13}{2} + 0 \cdot \frac{3}{2} + 1 \cdot 2 = \frac{17}{2},$$

$$\Delta'_1 = 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 0 - 1 = 0,$$

$$\Delta'_6 = 1 \cdot \left(-\frac{3}{2}\right) + 0 \cdot \frac{1}{2} + 1 \cdot 5 - 0 = \frac{7}{2}.$$

Убедившись в правильности расчетов, приступаем к анализу второй таблицы: план не оптимальный – есть положительная оценка $\Delta'_6 = \frac{7}{2}$, и так как в столбце A_6 есть положительные элементы, то можем перейти к следующему опорному плану.

Находим $\theta_{06} = \min \left\{ \frac{(3/2)}{(1/2)}, \frac{2}{5} \right\}$, выделяем ведущий элемент $a'_{36} = 5$

и, выполняя преобразования по методу полного исключения, заполняем третью симплексную таблицу (табл. 14) и проверяем ее.

Таблица 14 – Третья симплекс-таблица

I	Базис	$C_{\bar{b}}$	A_0	1	1	1	0	0	0			
				A_1	A_2	A_3	A_4	A_5	A_6			
1	A_1	1	71/10	1	1/5	3/10	0	-21/10	0			
2	A_4	0	13/10	0	3/5	-1/10	1	-13/10	0			
3	A_6	0	2/5	0	-1/5	1/5	0	-2/5	1	3/2	1/2	7/2
$m+1$	$z_j - c_j$		71/10	0	-4/5	-7/10	0	-21/10	0			

Пересчетом элементов последней строки. Все оценки в этой таблице неположительны, следовательно, записанный в ней опорный план является оптимальным, минимизирующим целевую функцию. Значения базисных переменных x_1^*, x_4^*, x_6^* записаны в таблице в столбце A_0 . Свободные переменные x_2^*, x_3^*, x_5^* равны нулю, т. е.

$$X^* = \left(\frac{71}{10}, 0, 0, \frac{13}{10}, 0, \frac{2}{5} \right).$$

Минимальное значение целевой функции записано в $(m+1)$ -й строке столбца A_0 :

$$z_{\min} = \frac{71}{10}.$$

Заметим, что последовательные симплексные таблицы записываются вместе под одной общей шапкой.

Пример 2. Симплексным методом решить задачу

$$\begin{aligned} z &= 5x_1 + 3x_2 \rightarrow \max, \\ \begin{cases} 3x_1 + 5x_2 \leq 15, \\ 5x_1 + 2x_2 \leq 10, \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

и дать геометрическую интерпретацию решения.

Решение. Приведем к канонической форме

$$z = 5x_1 + 3x_2 \rightarrow \max,$$

$$\begin{cases} 3x_1 + 5x_2 + x_3 = 15, \\ 5x_1 + 2x_2 + x_4 = 10, \\ x_j \geq 0 \quad (j = 1, 2, 3, 4). \end{cases}$$

При этом в системе ограничений появляются два линейно независимых единичных вектора A_3 и A_4 , следовательно, известен первоначальный опорный план и можно заполнить симплексную таблицу. Дальнейший ход решения представлен в таблице 15.

Таблица 15 – Четвертая симплекс-таблица

I	Базис	$C_{\bar{b}}$	A_0	5	3	0	0
				A_1	A_2	A_3	A_4
1	A_3	0	15	3	5	1	0
2	A_4	0	10	5	2	0	1
$m+1$	$z_j - c_j$		0	-5	-3	0	0
1	A_3	0	9	0	19/5	1	-3/5
2	A_4	5	2	1	2/5	0	1/5
$m+1$	$z_j - c_j$		10	0	-1	0	1
1	A_2	3	45/19	0	1	5/19	-3/19
2	A_1	5	20/19	1	0	-2/19	5/19
$m+1$	$z_j - c_j$		235/19	0	0	5/19	16/19

Первоначальный опорный план

$$X_0 = (0, 0, 15, 10)$$

оказался неоптимальным; полученный после первой итерации опорный план

$$X_1 = (2, 0, 9, 0)$$

также неоптimalен; после второй итерации получен оптимальный план

$$X^* = \left(\frac{20}{19}, \frac{45}{19}, 0, 0 \right),$$

для которого

$$z_{\max} = \frac{235}{19}.$$

Исходная стандартная задача может быть решена графически (рис. 12).

Последовательным итерациям в симплекс-методе соответствует на графике переход от угловой точки $O(0, 0)$ многоугольника решений к угловым точкам $P(2, 0)$ и $Q(20/19, 45/19)$. В последней точке прямая $z = 5x_1 + x_2 = \text{const}$ становится опорной, соответствующей максимуму целевой функции.

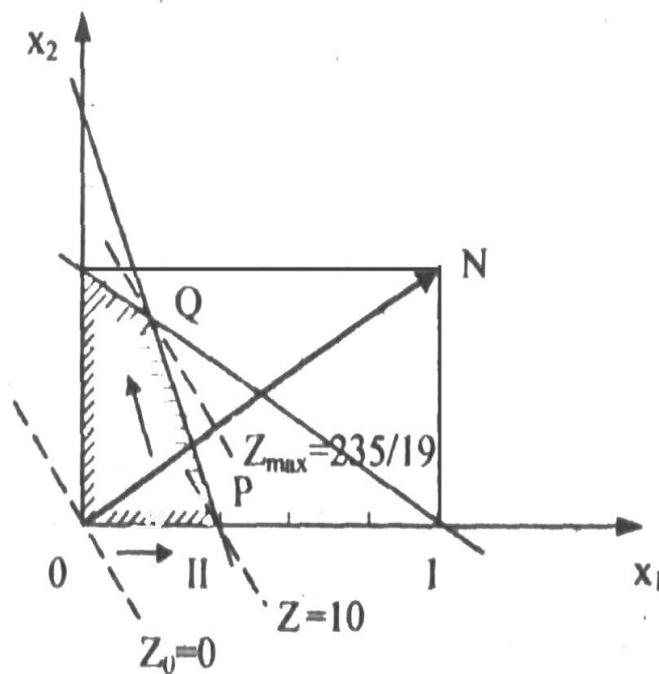


Рисунок 12 – Графическое решение задачи (пример 2)

Пример 3. Для изготовления различных изделий A , B и C предприятие использует три различных вида сырья. Нормы расхода сырья на производство одного изделия каждого вида, цена одного изделия, а также общее количество сырья каждого вида приведены в таблице 16.

Таблица 16 – Нормы затрат и общее количество сырья

Показатель	Норма затрат сырья на одно изделие			Общее количество сырья
	<i>A</i>	<i>B</i>	<i>C</i>	
Вид сырья: I	18	15	12	360
II	6	4	8	192
III	5	3	3	180
Цена одного изделия	9	10	16	–

Составить план производства изделий, при котором общая стоимость произведенной продукции будет максимальной.

Искомый выпуск изделий *A* обозначим через x_1 , изделий *B* – через x_2 , изделий *C* – через x_3 . Поскольку имеются ограничения на выделенный предприятию фонд сырья каждого вида, переменные x_1 , x_2 и x_3 должны удовлетворять системе неравенств

$$\begin{cases} 18x_1 + 15x_2 + 12x_3 \leq 360, \\ 6x_1 + 4x_2 + 8x_3 \leq 192, \\ 5x_1 + 3x_2 + 3x_3 \leq 180. \end{cases}$$

Общая стоимость произведенной предприятием продукции составляет

$$F = 9x_1 + 10x_2 + 16x_3.$$

По своему экономическому смыслу переменные x_1 , x_2 и x_3 могут принимать лишь неотрицательные значения:

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.$$

Запишем задачу в форме канонической. Для этого перейдем от ограничений-неравенств к ограничениям-равенствам. Введем три дополнительные переменные, в результате ограничения запишутся в виде системы уравнений

$$\begin{cases} 18x_1 + 15x_2 + 12x_3 + x_4 = 360, \\ 6x_1 + 4x_2 + 8x_3 + x_5 = 192, \\ 5x_1 + 3x_2 + 3x_3 + x_6 = 180. \end{cases}$$

Эти дополнительные переменные по экономическому смыслу означают не используемое при данном плане производства количество сырья того или иного типа, например P_4 – это неиспользуемое сырье I вида.

Преобразованную систему уравнений запишем в векторной форме:

$$x_1P_1 + x_2P_2 + x_3P_3 + x_4P_4 + x_5P_5 + x_6P_6 = P_0,$$

где

$$P_1 = \begin{pmatrix} 18 \\ 6 \\ 5 \end{pmatrix}, P_2 = \begin{pmatrix} 15 \\ 4 \\ 3 \end{pmatrix}, P_3 = \begin{pmatrix} 12 \\ 8 \\ 3 \end{pmatrix}, P_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, P_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, P_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, P_0 = \begin{pmatrix} 360 \\ 192 \\ 180 \end{pmatrix}.$$

Поскольку среди векторов P_1, \dots, P_6 имеются три единичных вектора, то можно непосредственно записать опорный план $X = (0, 0, 0, 360, 192, 180)$, определяемый системой трехмерных единичных векторов P_4, P_5, P_6 , которые образуют базис векторного пространства.

Составим симплекс-таблицу для итерации 1 (табл. 17) и проверим исходный план на оптимальность:

$$F_0 = (C_0, P_0) = 0; z_1 = (C_0, P_1) = 0; z_2 = (C_0, P_2) = 0; z_3 = (C_0, P_3) = 0;$$

$$z_1 - c_1 = 0 - 9 = -9; z_2 - c_2 = 0 - 10 = -10; z_3 - c_3 = 0 - 16 = -16.$$

$$\text{Для векторов базиса } \Delta_j = z_j - c_j = 0.$$

Таблица 17 – Симплекс-таблица итерации 1

I	Базис	C_0	P_0	9	10	16	0	0	0
				P_1	P_2	P_3	P_4	P_5	P_6
1	P_4	0	360	18	15	12	1	0	0
2	P_5	0	192	6	4	8	0	1	0
3	P_6	0	180	5	3	3	0	0	1
4	$z_j - c_j$		0	-9	-10	-16	0	0	0

Опорный план X не является оптимальным, поскольку в четвертой строке симплекс-таблицы имеются три отрицательных числа. Это свидетельствует о возможности увеличения общей стоимости производимой продукции и показывает, на сколько она увеличится при введении в план единицы того или иного вида.

Так, число (-9) означает, что при включении в план производства одного изделия A обеспечивается увеличение выпуска продукции на 9 единиц. Поэтому с экономической точки зрения наиболее целесообразным является включение в план производства изделий C (соответствующие вектору P_3). Это же необходимо сделать и на основании формального признака симплекс-метода, поскольку максимальное по абсолютной величине отрицательное число $z_i - c_i$ стоит в четвертой строке таблицы и соответствует вектору P_3 , который следует ввести в базис. Для определения вектора, подлежащего исключению, находим $\min b/a$ при $a > 0$, т. е. $\min(360/12; 192/8; 180/3) = 192/8 = 24$. Таким образом, из базиса следует исключить вектор P_5 . Ограничивающим фактором для производства изделий C является имеющийся объем сырья II вида. С учетом этого предприятие может выпустить 24 изделия C .

Итак, столбец P_3 и строка P_5 являются в симплекс-таблице направляющими. Составим таблицу для итерации 2 (табл. 18).

Таблица 18 – Симплекс-таблица итерации 2

I	Базис	C_b	P_0	9	10	16	0	0	0
				P_1	P_2	P_3	P_4	P_5	P_6
1	P_4	0	72	9	9	0	1	$-3/2$	0
2	P_3	16	24	$3/4$	$1/2$	1	0	$1/8$	0
3	P_6	0	108	$11/4$	$3/2$	0	0	$-3/8$	1
4	$z_i - c_i$		384	3	-2	0	0	2	0

Сначала заполним строку для вновь введенного в базис вектора P_3 (вторую строку в таблице 18). Элементы этой строки получаются из элементов второй строки таблицы 17 делением их на разрешающий элемент (на 8). При этом в столбце C_b записываем коэффициент $c_3 = 16$, соответствующий вводимому в базис вектору P_3 . Для опреде-

ления остальных элементов таблицы 18 применяем правило треугольника.

Вычислим, например, число, являющееся первым элементом вектора P_0 в таблице 18. Для его вычисления находим три числа, расположенные:

1) в таблице 17 на пересечении столбца вектора P_0 и первой строки (360);

2) в таблице 17 на пересечении столбца вектора P_3 и первой строки (12);

3) в таблице 18 на пересечении столбца вектора P_0 и второй строки (24).

Вычитая из первого числа произведение двух других, находим искомый элемент: $360 - 12 \cdot 24 = 72$. Аналогично находим элементы столбцов P_1, P_2, P_3 .

По окончании расчета всех элементов в таблице 18 получаем новый опорный план $X = (0,24; 0,72; 0,108)^T$. При данном плане производства изготавливается 24 изделия C и остаются неиспользованными 72 единицы сырья вида I и 108 – сырья вида III. Стоимость всей производимой при этом плане продукции равна 384.

Найденный на итерации 2 план также не является оптимальным, поскольку в четвертой строке в таблице 18 на пересечении со столбцом P_2 стоит отрицательное число (-2). Далее в базис следует ввести вектор P_2 . Найдем вектор, подлежащий исключению из базиса. Для этого находим $\min(72/9, 48/1, 360/3) = 72/9 = 8$. Следовательно, исключению из базиса подлежит вектор P_4 . Число 9 является разрешающим элементом, а столбец P_2 и строка P_1 являются направляющими. Составим таблицу 19 для итерации 3.

Таблица 19 – Симплекс-таблица итерации 3

I	Базис	C_b	P_0	9	10	16	0	0	0
				P_1	P_2	P_3	P_4	P_5	P_6
1	P_2	10	8	1	1	0	1/9	-1/6	0
2	P_3	16	20	1/4	0	1	-1/18	5/24	0
3	P_6	0	96	5/4	0	0	-1/6	-1/8	1
4	$z_i - c_i$		400	5	0	0	2/9	5/3	0

В таблице 19 сначала заполняем элементы первой строки (вектор P_2). Элементы этой строки получаем из элементов первой строки таблицы 18 делением на разрешающий элемент (на 9). При этом в столбце C_6 записываем коэффициент $c_2 = 10$, соответствующий вводимому в базис вектору P_2 .

Затем по правилу треугольника вычисляем элементы остальных столбцов. В результате получаем новый опорный план $X = (0; 8; 20; 0; 0; 96)^T$ и коэффициенты разложения векторов P_i ($i = 1, \dots, 6$) через базисные векторы P_2, P_3, P_6 , а также значения четвертой строки и столбца P_0 .

Проверяем, является ли данный опорный план оптимальным. Для этого рассмотрим четвертую строку таблицы 19. В этой строке среди чисел нет отрицательных. Следовательно, найденный план является оптимальным и $F_{max} = 400$.

С экономической точки зрения план выпуска продукции, включающий изготовление 8 изделий B и 20 изделий C , является оптимальным. При данном плане выпуска полностью используется сырье вида I и II, остаются неиспользованными 96 единиц сырья вида III, а стоимость произведенной продукции равна 400. Оптимальным планом не предусматривается изготовление изделий A .

Решить симплексным методом следующие задачи:

$$1. z = 2x_1 - x_2 + 3x_3 - 2x_4 + x_5 \rightarrow \max, \begin{cases} -x_1 + x_2 + x_3 = 1, \\ x_1 - x_2 + x_4 = 1, \\ x_1 + x_2 + x_5 = 2, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 5). \end{cases}$$

Ответ: $X^* = (1; 0; 2; 0; 1)$, $z_{max} = 9$.

$$2. z = -x_1 + 3x_2 + 2x_3 \rightarrow \min, \begin{cases} x_1 + x_2 + 2x_3 \geq 5, \\ 2x_1 - 3x_2 + x_3 \leq 3, \\ 2x_1 - 5x_2 + 6x_3 \leq 5, \\ x_j \geq 0 \quad (j = 1, 2, 3). \end{cases}$$

Ответ: $X^* = (3/2; 0; 0)$, $z_{min} = -3/2$.

$$3. z = x_1 + 5x_2 + 4x_3 - 6x_4 \rightarrow \max, \begin{cases} 2x_1 + 3x_2 - 4x_3 - 5x_4 \leq 1, \\ 5x_1 - 6x_2 + x_3 - x_4 \leq 3, \\ 4x_1 + x_2 - 2x_3 + 3x_4 \leq 2, \\ x_j \geq 0 \quad (j = 1, 2, 3, 4). \end{cases}$$

Ответ: решения нет, $z \rightarrow \infty$.

$$4. z = x_1 - x_2 + x_3 + x_4 - x_6 \rightarrow \min, \begin{cases} x_1 + x_4 + 6x_6 = 9, \\ 3x_1 + x_2 - 4x_3 + 2x_6 = 2, \\ -5x_1 + 8x_3 + x_5 - 2x_6 = 2, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases}$$

Ответ: $X^* = (0; 3/2; 5/8; 0; 0; 3/2)$, $z_{\min} = -19/8$.

$$5. z = x_4 - x_5 \rightarrow \max, \begin{cases} 2x_2 + 4x_3 - x_4 + x_5 + x_6 \leq 2, \\ -2x_2 + x_3 + x_4 - x_5 \leq 0, \\ 3x_2 + x_3 + x_4 - x_5 \leq 1, \\ x_1 + x_2 + x_3 = 1, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases}$$

Ответ: $X^* = (4/5; 1/5; 0; 2/5; 0; 2)$, $z_{\min} = 2/5$.

Метод искусственного базиса

Если система ограничений задачи линейного программирования, записанной в канонической форме, не содержит m единичных линейно независимых векторов, то не представляется возможным непосредственно указать первоначальный опорный план. В этом случае для решения задачи применяется симплексный метод, дополненный методом искусственного базиса, сущность которого состоит в следующем.

Для данной задачи линейного программирования

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min(\max),$$

Значения целевой функции z и оценок $z_j - c_j$ расширенной задачи состоят из двух слагаемых, одно из которых зависит от M , а другое не зависит. Поэтому симплексная таблица для расширенной задачи содержит на одну строку больше. В дополнительную $(m + 2)$ -ю строку помещают коэффициенты при M , а в $(m + 1)$ -ю – слагаемые, не зависящие от M .

Вначале итерационный процесс перехода от одного опорного плана к другому ведут по $(m + 2)$ -й строке, т. е. учитывая только части оценок, стоящие в этой строке. При этом, когда какой-либо искусственный вектор выводится из базиса, он исключается из дальнейшего рассмотрения.

Когда все искусственные векторы будут выведены из базиса, дальнейший итерационный процесс ведется по $(m + 1)$ -й строке, т. е. фактически решается уже исходная задача; $(m + 2)$ -я строка тогда отбрасывается.

Если же окажется, что в $(m + 2)$ -й строке условия оптимальности выполняются для всех столбцов, а из базиса исключены не все искусственные векторы – это означает, что исходная задача не имеет решения.

Заметим, что если система ограничений исходной задачи содержит несколько единичных векторов, то следует включить их в базис расширенной задачи, уменьшив соответственно число вводимых искусственных переменных и векторов и сократив тем самым объем вычислений при решении задачи.

Пример 1. Решить задачу:

$$z = -x_1 - 2x_2 + 3x_3 - 10x_4 \rightarrow \max ,$$

$$\begin{cases} x_1 + x_2 + 2x_3 - 6x_4 = 1, \\ x_1 + x_2 + 4x_3 - 8x_4 = 1, \\ 4x_1 + 2x_2 + x_3 - 4x_4 = 3, \\ x_j \geq 0 \quad (j = 1, 2, 3, 4). \end{cases}$$

Решение. Система ограничений не содержит единичных векторов, поэтому для получения первоначального опорного плана применим метод искусственного базиса. Запишем для данной задачи расширенную

$$\bar{z} = -x_1 - 2x_2 + 3x_3 - 10x_4 - Mx_5 - Mx_6 - Mx_7 \rightarrow \max ,$$

$$\begin{cases} x_1 + x_2 + 2x_3 - 6x_4 + x_5 = 1, \\ x_1 + x_2 + 4x_3 - 8x_4 + x_6 = 1, \\ 4x_1 + 2x_2 + x_3 - 4x_4 + x_7 = 3, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 7). \end{cases}$$

и применим к ней симплексный метод в соответствии с вышеизложенной методикой (табл. 20).

Таблица 20 – Таблица итерации 1

I	Базис	C_{σ}	A_0	1	-2	3	-10	-M	-M	-M
				A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	A_5	-M	1	1	$1\sqrt{}$	2	-6	1	0	0
2	A_6	-M	1	1	$1\sqrt{}$	$4\sqrt{}$	-8	0	1	0
3	A_7	-M	3	$4\sqrt{}$	2	1	-4	0	0	1
m+1	$z_j - c_j$		0	-1	2	-3	10	0	0	0
m+2			-5	-6	-4	-7	18	0	0	0
1	A_5	-M	1/4	0	$\frac{1}{2}\sqrt{}$	7/4	-5	1	0	
2	A_6	-M	1/4	0	$\frac{1}{2}\sqrt{}$	15/4	-7	0	1	
3	A_1	1	3/4	1	1/2	1/4	-1	0	0	
m+1	$z_j - c_j$		$\frac{3}{4}$	0	5/2	-11/4	9	0	0	
m+2			-1/2	0	-1	-11/2	12	0	0	
1	A_5	-M	0	0	0	-2	$2\sqrt{}$	1		
2	A_2	-2	1/2	0	1	15/2	-14	0		
3	A_1	1	1/2	1	0	-7/2	6	0		
m+1	$z_j - c_j$		-1/2	0	0	-43/2	44	0		
m+2			0	0	0	2	-2	0		
1	A_4	-10	0	0	0	-1	1			
2	A_2	-2	1/2	0	1	-13/2	0			
3	A_1	-1	1/2	1	0	5/2	0			
m+1	$z_j - c_j$		-1/2	0	0	45/2	0			

m+2		0	0	0	0	0			
-----	--	---	---	---	---	---	--	--	--

Первоначальный опорный план исходной задачи, полученный после выделения из базиса всех искусственных векторов, оказался оптимальным, т. е.

$$X^* = (0.5, \quad 0.5, \quad 0, \quad 0),$$

$$z_{\max} = -0.5.$$

Пример 2. Решить задачу:

$$z = 2x_1 + 3x_2 + 2.5x_3 \rightarrow \min,$$

$$\begin{cases} 2x_1 + x_2 + 3x_3 \geq 6, \\ 2x_1 + 4x_2 + 3x_3 \geq 16, \\ 3x_1 + 4x_2 + 2x_3 \geq 12, \\ x_j \geq 0 \quad (j = 1, 2, 3). \end{cases}$$

Решение. Прежде всего, приведем задачу к канонической форме, для чего запишем систему ограничений в виде равенств:

$$\begin{cases} 2x_1 + x_2 + 3x_3 - x_4 = 6, \\ 2x_1 + 4x_2 + 3x_3 - x_5 = 16, \\ 3x_1 + 4x_2 + 2x_3 - x_6 = 12, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases} \quad (74)$$

Дополнительные переменные x_4, x_5, x_6 входят в эти равенства со знаком «−», поэтому векторы A_4, A_5, A_6 не образуют нужного нам единичного базиса. (Если бы мы приняли векторы A_4, A_5, A_6 за базис, то получили бы, что коэффициенты разложения вектора A_0 в этом базисе, $x_4 = -6, x_5 = -16, x_6 = -12$, отрицательные и поэтому не являются компонентами плана задачи.) В подобных случаях, однако, можно так преобразовать систему уравнений-ограничений, что все дополнительные переменные, кроме одной, будут входить в уравнения со знаком «+». Делается это следующим образом: уравнение с наибольшим свободным членом в правой части (второе уравнение 75) оставляем без изменения и из него вычитаем пооче-

редно остальные уравнения, получая тем самым преобразованную систему уравнений:

$$\begin{cases} 3x_2 + x_4 - x_5 = 10, \\ 2x_1 + 4x_2 + 3x_3 - x_5 = 16, \\ -x_1 + x_3 - x_5 + x_6 = 4, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases} \quad (75)$$

Система ограничений (75) содержит два единичных вектора

$$A_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

поэтому для образования единичного базиса достаточно ввести лишь один искусственный вектор

$$A_7 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

т. е. одну искусственную переменную x_7 во второе уравнение.

$$\begin{aligned} \bar{z} &= 2x_1 + 3x_2 + 2.5x_3 + Mx_7 \rightarrow \min, \\ &\begin{cases} 3x_2 + x_4 - x_5 = 10, \\ 2x_1 + 4x_2 + 3x_3 - x_5 + x_7 = 16, \\ -x_1 + x_3 - x_5 + x_6 = 4, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 7), \end{cases} \end{aligned}$$

и решить эту систему симплексным методом (табл. 21).

Таблица 21 – Таблица итерации 2

I	Базис	C_b	A_0	2	3	2.5	0	0	0	M
				A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	A_4	0	10	0	3	0	1	-1	0	0
2	A_7	M	16	2	4	3	0	-1	0	1
3	A_6	0	4	-1	0	1	0	-1	1	0

m+1	$z_j - c_j$	0	-2	-3	-5/2	0	0	0	0
m+2		16	2	4	3	0	-1	0	0

Окончание табл. 21

I	Базис		C_b	A_0	2	3	2.5	0	0	0
					A_1	A_2	A_3	A_4	A_5	A_6
1	A_4	0	10	0	3	0	1	-1	0	
2	A_1	2	8	1	2	3/2	0	-1/2	0	
3	A_6	0	12	0	2	5/2	0	-3/2	1	
m+1	$z_j - c_j$		16	0	1	1/2	0	-1	0	
m+2			0	0	0	0	0	0	0	0
1	A_2	3	10/3	0	1	0	1/3	1/3	0	
2	A_1	2	4/3	1	0	3/2	-2/3	1/6	0	
3	A_6	0	16/3	0	0	5/2	-2/3	-5/6	1	
m+1	$z_j - c_j$		38/3	0	0	1/2	-1/3	-2/3	0	
1			A_2	3	10/3	0	1	0	1/3	-1/3
2	A_3	5/2	8/9	2/3	0	1	-4/9	1/9	0	
3	A_6	0	28/9	-5/3	0	0	4/9	-10/9	1	
m+1	$z_j - c_j$		110/9	-1/3	0	0	-1/9	-13/18	0	

Ответ: $X^* = (0, \frac{10}{3}, \frac{8}{9})$, $z_{\min} = \frac{110}{9}$.

Пример 3. Решить задачу:

$$z = x_1 - 2x_2 + 3x_3 \rightarrow \min ,$$

$$\begin{cases} -2x_1 + x_2 + 3x_3 = 2, \\ 2x_1 + 3x_2 + 4x_3 = 1, \\ x_j \geq 0 \quad (j = 1, 2, 3). \end{cases}$$

Решение. Переходим к расширенной задаче:

$$\bar{z} = x_1 - 2x_2 + 3x_3 + Mx_4 + Mx_5 \rightarrow \min ,$$

$$\begin{cases} -2x_1 + x_2 + 3x_3 + x_4 = 2 , \\ 2x_1 + 3x_2 + 4x_3 + x_5 = 1, \\ x_j \geq 0 (j = 1, 2, \dots, 5) \end{cases}$$

и применяем симплексный метод (табл. 22).

Таблица 22 – Таблица итерации 1

I	Базис	$C_{\bar{b}}$	A_0	1	-2	3	M	M
				A_1	A_2	A_3	A_4	A_5
1	A_4	M	2	-2	1	3	1	0
2	A_5	M	1	2	3	4	0	1
m+1	$z_j - c_j$		0	-1	2	-3	0	0
m+2			3	0	4	7	0	0
1	A_4	M	5/4	-7/2	-5/4	0	1	
2	A_3	3	1/4	1/2	3/4	1	0	
m+1	$z_j - c_j$		3/4	1/2	17/4	0	0	
m+2			5/4	-7/2	-5/4	0	0	

Оптимальное решение расширенной задачи достигнуто, но в базисе при этом остался искусственный вектор A_4 , значит исходная задача не имеет решения (система ограничений несовместна).

Решить симплексным методом.

$$1. z = -3x_1 + x_2 + 3x_3 - x_4 \rightarrow \min , \begin{cases} x_1 + 2x_2 - x_3 + x_4 = 0, \\ 2x_1 - 2x_2 + 3x_3 + 3x_4 = 9, \\ x_1 - x_2 + 2x_3 - x_4 = 6, \\ x_j \geq 0 (j = 1, 2, 3, 4). \end{cases}$$

Ответ: $X^* = (1; 1; 3; 0)$, $z_{\min} = 7$.

$$2. z = x_1 + x_2 - x_3 - 2x_5 \rightarrow \min , \begin{cases} -x_1 + 2x_2 - x_4 = 3, \\ x_3 - 2x_4 = 2, \\ 3x_2 - x_4 + x_5 \leq 5, \\ x_2 + x_5 \leq 3, \end{cases}$$

$$x_j \geq 0 \quad (j = 1, 2, \dots, 5).$$

Ответ: $X^* = (0; 2; 4; 1; 0)$, $z_{\min} = -2$.

$$3. \quad z = 2x_1 - x_2 + 3x_3 - 5x_5 \rightarrow \max, \quad \begin{cases} x_1 + x_2 + x_3 = 6, \\ -x_2 + x_5 = 3, \\ -x_3 + x_6 = 3, \\ x_4 + x_5 + x_6 = 15, \\ x_j \geq 0 \quad (j = 1, 2, \dots, 6). \end{cases}$$

Ответ: $X^* = (0; 0; 6; 3; 3; 9)$, $z_{\max} = 3$.

Следующие задачи решить сначала симплексным методом, а затем – графическим и дать геометрическую интерпретацию процесса решения симплекс-методом.

$$4. \quad z = 4x_1 + 3x_2 \rightarrow \max, \quad \begin{cases} x_1 - x_2 \geq -2, \\ 5x_1 + 3x_2 \leq 15, \\ x_2 \leq \frac{5}{2}, \\ x_1 - 2x_2 \leq 2, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

$$5. \quad z = x_1 \rightarrow \max, \quad \begin{cases} x_1 - 2x_2 \leq 0, \\ x_1 - x_2 \geq -1, \\ x_1 + x_2 \geq 1, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Содержание отчета

1. Титульный лист.
2. Цель и задачи работы.
3. Результаты выполнения работы.

4. Ответы на контрольные вопросы.

Защитить отчет у преподавателя.

Контрольные вопросы

1. Перечислите основные методы линейного программирования.

2. Построение математической модели оптимизационной задачи линейного программирования.

3. Что такое каноническая и стандартная формы задачи линейного программирования?

4. Перечислить основные этапы графического метода решения оптимизационной задачи.

5. Графическое решение оптимизационной задачи с использованием метода Жордана-Гаусса.

6. Симплекс-метод решения задачи линейного программирования.

7. Назовите особенности метода искусственного базиса.

ЗАКЛЮЧЕНИЕ

Представленные в учебно-методическом пособии классические, общеизвестные методы решения различных задач оптимизации направлены на нахождение наиболее рационального решения той или иной проблемы, в том числе в областях технических, научно-практических, организационно-экономических.

Основное внимание уделяется обоснованию математической постановки задачи и метода ее решения, практическому применению методов оптимизации к решению конкретных задач в области применения различных математических инструментов нахождения оптимальных или рациональных подходов к принятию управляющих воздействий и механизмов на основе математических моделей оптимизации.

В результате изучения дисциплины студент должен знать основные понятия методов оптимизации, такие как численные методы математического программирования и элементы выпуклого анализа; уметь применять эти знания на практике в решении задач и владеть навыками работы со справочной литературой.

Изучение основ различных методик оптимизации позволит будущим специалистам в области рационального использования сельскохозяйственной техники решать конкретные производственные или научно-прикладные проблемы с целью нахождения оптимумов при принятии решений, являясь ЛПР (лицами, принимающими решения).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Андреева, Е.А. Вариационное исчисление и методы оптимизации / Е.А. Андреева, В.М. Цирулева. – М.: Высш. шк., 2006.
2. Ванько, В.И. Вариационное исчисление и оптимальное управление / В.И. Ванько, О.В. Ермошина, Г.Н. Кувыркин. – М.: Изд-во МГТУ, 2006.
3. Григорьев, А.В. Практикум по линейному программированию / А.В. Григорьев, Т.В. Григорьева. – Красноярск: Изд-во КГТУ, 1998.
4. Кузнецов, Ю.Н. Математическое программирование / Ю.Н. Кузнецов, В.И. Кузубов, А.Б. Волощенко. – М.: Высш. шк., 1976. – 352 с.
5. Пантелеев, А.В. Методы оптимизации в примерах и задачах / А.В. Пантелеев, Т.А. Летова. – М.: Высш. шк., 2005.
6. Пантелев, В.А. Оптимальное управление в примерах и задачах / В.А. Пантелев, А.С. Бортакровский. – М.: Высш. шк., 2001.
7. Гончаров, В.А. Методы оптимизации: учеб. пособие / В.А. Гончаров. – М.: Высшее образование, 2009. – 192 с.

МЕТОДЫ ОПТИМИЗАЦИИ

Учебно-методическое пособие

Журавлев Сергей Юрьевич

Редактор
О.Ю. Потапова

Санитарно-эпидемиологическое заключение № 24.49.04.953.П. 000381.09.03 от 25.09.2003 г.

Подписано в печать 28.10.2014. Формат 60x84/16. Бумага тип. № 1.

Печать – ризограф. Усл. печ. л. 8,25. Тираж 110 экз. Заказ № 495

Издательство Красноярского государственного аграрного университета
660017, Красноярск, ул. Ленина, 117