

5. Методы решения задач выпуклого программирования.

К содержанию

5.1. Экстремальные задачи без ограничений

Экстремальная точка функции $f(X)$ определяет либо максимум, либо минимум этой функции. Точка $X_0 = (x_1, \dots, x_j, \dots, x_n)$ является точкой максимума, если неравенство

$$f(X_0 + h) \leq f(X_0)$$

выполняется для всех $h = (h_1, \dots, h_j, \dots, h_n)$, таких, что $|h_j|$ достаточно малы для всех j . Другими словами, X_0 является точкой максимума, если значение функции f в каждой точке достаточно малой окрестности X_0 не превышает $f(X_0)$. Аналогично X_0 является точкой минимума, если для вектора h , определенного выше, справедливо неравенство

$$f(X_0 + h) \geq f(X_0).$$

На рис. 5.1 показаны максимумы и минимумы функции одной переменной $f(x)$ на отрезке $[a, b]$. (Условие $a \leq x \leq b$ не означает, что на $f(x)$ наложены ограничения.) Точки x_1, x_2, x_3, x_4 и x_6 определяют экстремумы $f(x)$, причем x_1, x_3 , и x_6 — максимумы, а x_2 и x_4 — минимумы. Так как

$$f(x_6) = \max(f(x_1), f(x_3), f(x_6)),$$

то $f(x_6)$ представляет собой глобальный, или абсолютный, максимум, а $f(x_1)$ и $f(x_3)$ являются локальными, или относительными максимумами. Аналогично $f(x_4)$ есть локальный минимум, а $f(x_2)$ — глобальный минимум.

Максимум в точке x_1 (см. рис. 5.1) отличается от остальных локальных максимумов тем, что значение функции по крайней мере в одной точке окрестности x_1 равно $f(x_1)$. По этой причине x_1 называют точкой нестрогого миниму-

ма в отличие, например, от точки x_3 , которая определяет строгий максимум $f(x_3)$.

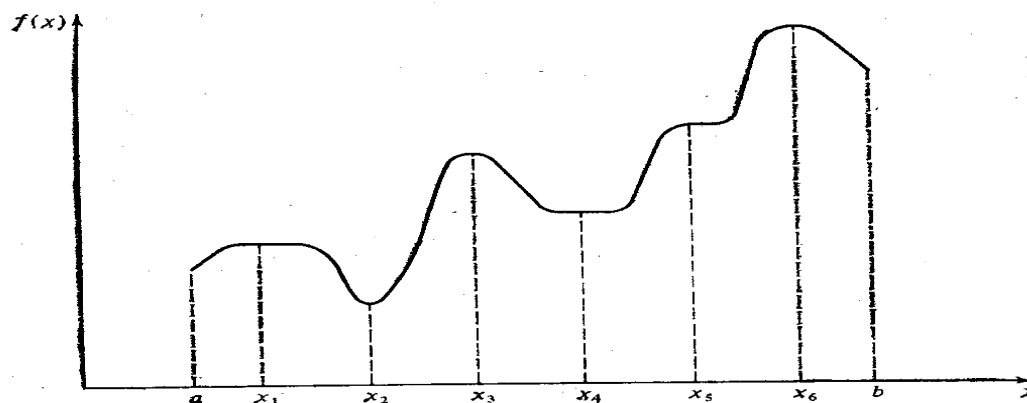


Рис. 5.1.

Под нестрогим максимумом, таким образом, понимается наличие (бесконечного количества) несовпадающих точек, которым соответствует одно и то же максимальное значение функции. В точке x_4 наблюдается нестрогий минимум, который определяется аналогичным образом. Вообще говоря, X_0 является точкой нестрогого максимума, если $f(X_0 + h) \leq f(X_0)$ и точкой строгого максимума, если $f(X_0 + h) < f(X_0)$ где h — вектор, определение которого дано выше.

С помощью рис. 5.1 нетрудно заметить, что первая производная (тангенс угла наклона касательной к графику) функции f стремится к нулю по мере приближения к экстремальным точкам. Однако это характерно не только для экстремумов. Например, тангенс угла наклона касательной к графику $f(x)$ в точке x , также равен нулю.

Так как стремление к нулю первой производной (в общем случае градиента) играет важную роль при поиске максимумов и минимумов функций (см. следующий раздел), целесообразно выделить точки, подобные x_5 отдельный класс точек перегиба (или в особых случаях седловых точек). Если точка, в ко-

торой угол наклона касательной к графику функции (градиент) равен нулю, не является точкой экстремума (максимума или минимума), то она автоматически оказывается точкой перегиба.

[К содержанию](#)

5.1.1. Необходимые и достаточные условия существования экстремумов

В данном подразделе представлены теоремы, в которых формулируются необходимые и достаточные условия существования экстремумов функции n переменных $f(X)$. При этом предполагается, что первая и вторая частные производные $f(X)$ непрерывны в каждой точке X .

Теорема 5.1.1. Если точка X , является экстремальной точкой функции $f(X)$, то $\nabla f(X_0)=0$.

Из теоремы 5.1.1 следует, что условие $\nabla f(X_0)=0$ должно выполняться для любой экстремальной точки, т. е. градиент в экстремальной точке должен быть нулевым вектором.

Для функции одной переменной (например, $f(y)$) это условие записывается следующим образом: $f'(y_0)=0$.

Как было отмечено ранее, полученное условие удовлетворяется также в точках перегиба и седловых точках функции. Следовательно, оно является необходимым, но недостаточным для идентификации экстремальных точек. В связи с этим точки, удовлетворяющие уравнению

$$\nabla f(X_0)=0$$

будем называть *стационарными*. Следующая теорема устанавливает достаточные условия для того, чтобы X , была экстремальной точкой.

Теорема 5.1.2. Стационарная точка X_0 является экстремальной, когда матрица Гессе H в точке X_0 оказывается (1) положительно определенной (тогда X_0 – точка минимума); (2) отрицательно определенной (тогда X_0 – точка максимума).

Пример 5.1.1. Дана функция

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2x_3 - x_1^2 - x_2^2 - x_3^2.$$

Из необходимого условия $\nabla f(X_0) = 0$ следует, что $\partial f / \partial x_1 = 1 - 2x_1 = 0$, $\partial f / \partial x_2 = x_3 - 2x_2 = 0$, $\partial f / \partial x_3 = 2 + x_2 - 2x_3 = 0$. Решением этой системы является вектор $X_0 = (1/2, 2/3, 4/3)$. Чтобы проверить выполнение достаточного условия, вычислим

$$H|_{X_0} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix}_{X_0} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}.$$

Главные миноры $H|_{X_0}$ равны -2 , 4 и -6 соответственно. В этом случае $H|_{X_0}$ является отрицательно определенной матрицей, откуда следует, что $X_0 = (1/2, 2/3, 4/3)$ – точка максимума.

Если $H|_{X_0}$ – неопределенная матрица, то X_0 является седловой точкой. В случае когда матрица Гессе оказывается полуопределенной, в точке X_0 может быть экстремум. Формулировка достаточного условия существования экстремума в этом случае значительно усложняется, так как необходимо рассчитывать члены высших порядков в разложении Тейлора. (См. теорему 5.1.3, которая служит иллюстрацией вышеизложенного для функции одной переменной.) Однако не всегда необходимы такие сложные процедуры, поскольку диагона-

лизация матрицы H в ряде случаев позволяет сделать вывод о наличии или отсутствии экстремума.

Достаточные условия, установленные теоремой 5.1.2, весьма просто формулируются для функций одной переменной. Пусть y_0 – стационарная точка, тогда

(1) $f''(y_0) < 0$ – достаточное условие существования максимума в точке y_0

(2) $f''(y_0) > 0$ – достаточное условие существования минимума в точке y_0 .

Эти условия непосредственно вытекают из того факта, что матрица Гессе для функции одной переменной содержит один элемент.

Если в случае одной переменной значение $f''(y_0)$ равно нулю, то необходимо исследовать производные высших порядков.

Теорема 5.1.3. Если в стационарной точке y_0 первые $(n-1)$ производных функции $f(y)$ обращаются в нуль, а $f^{(n)}(y) \neq 0$, то при $y = y_0$ функция $f(y)$ имеет

(1) точку перегиба, если n – нечетное;

(2) экстремальную точку, если n – четное. Экстремальной точке соответствует максимум при $f^{(n)}(y) < 0$ и минимум при $f^{(n)}(y) > 0$

[К содержанию](#)

5.1.2. Метод Ньютона

Использование необходимого условия $\nabla f(X_0) = 0$ для нахождения стационарных точек связано с определенными трудностями, которые возникают в процессе численного решения соответствующей системы уравнений. Метод Ньютона рассматривается здесь как итерационная процедура решения систем нелинейных уравнений. Указанный метод, вообще говоря, относится к числу градиентных методов поиска экстремумов функций при отсутствии ограничений.

Рассмотрим систему уравнений

$$f_i(X)=0, \quad i=1, 2, \dots, m.$$

Пусть X^k – заданная точка. Из разложения Тейлора следует, что

$$f_i(X) \approx f_i(X^k) + \nabla f_i(X^k)(X - X^k), \quad i = 1, 2, \dots, m.$$

Таким образом, исходную систему уравнений можно приближенно представить в следующем виде:

$$f_i(X^k) + \nabla f_i(X^k)(X - X^k) = 0, \quad i=1, 2, \dots, m$$

Запишем полученную систему в матричной форме:

$$A_k + B_k(X - X^k) = 0.$$

Предположим, что все $f_i(X)$ независимы; тогда B_k – невырожденная матрица. Из предыдущего уравнения получаем

$$X = X^k - B_k^{-1} A_k.$$

Идея рассматриваемого метода заключается в следующем. На первом шаге задается начальная точка X^0 . Если X^k известна, то с помощью полученного уравнения можно вычислить координаты новой точки X^{k+1} . Процесс вычислений завершается, когда $X^m \approx X^{m+1}$, при этом X^m – приближенное решение исходной системы.

Методу Ньютона нетрудно дать геометрическую интерпретацию. На рис. 5.2 изображен график функции одной переменной. Связь между x^k и x^{k+1} в случае функции одной переменной $f(x)$ выражается формулой

$$x^{k+1} = x^k - [f(x^k) / f'(x^k)]$$

или

$$f'(x^k) = f(x^k) / (x^k - x^{k+1})$$

Как следует из рис. 5.2, положение точки x^k определяется наклоном касательной к графику $f(x)$ в точке x^k , т. е. тангенсом угла θ , $\operatorname{tg} \theta = f'(x^k)$.

К числу недостатков изложенного метода относится то обстоятельство, что сходимость метода в существенной степени зависит от выбора начальной точки. Из рис. 5.2 видно, что при выборе точки a в качестве x^0 итерационный процесс расходится. Осуществить выбор

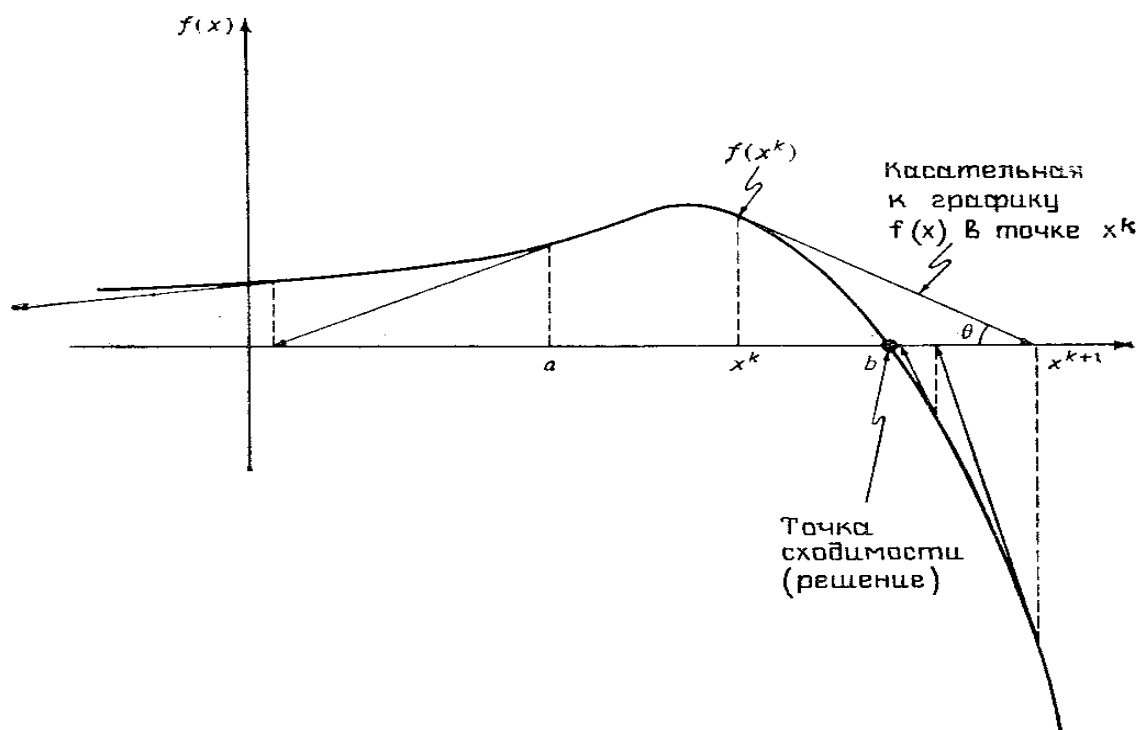


Рис. 5.2.

«хорошего» начального приближения x^0 каким-либо простым способом в общем случае невозможно.

К содержанию

5.2. Экстремальные задачи при наличии ограничений

5.2.1. Ограничения в виде равенств

В данном разделе рассматриваются два метода оптимизации при наличии ограничений в виде равенств. Один из них – метод Якоби – представляет со-

бой обобщение симплекс-метода линейного программирования. Действительно, все процедуры, связанные с реализацией симплекс-метода, можно обосновать, пользуясь методом Якоби. Другой метод, метод множителей Лагранжа, тесно связан с методом Якоби и является его логическим развитием. Указанная связь позволяет дать методу множителей Лагранжа содержательную экономическую интерпретацию.

5.2.1.1. Метод приведенного градиента (метод Якоби)

Рассмотрим следующую задачу:

$$\text{Найти} \quad \min Z=f(X)$$

при ограничениях

$$g(X)=0,$$

где $X=(x_1, x_2, \dots, x_n)$, $g=(g_1, g_2, \dots, g_m)$. Функции $f(X)$ и $g_i(X)$, $i=1, 2, \dots, m$, предполагаются дважды непрерывно дифференцируемыми.

Идея использования приведенного градиента для решения сформулированной выше задачи заключается в том, чтобы найти аналитическое выражение для первых частных производных функции $f(X)$ во всех точках, удовлетворяющих ограничению $g(X)=0$. При этом стационарными будут точки, в которых указанные частные производные обращаются в нуль. Для классификации стационарных точек можно пользоваться достаточными условиями существования экстремумов, сформулированными в разд. 5.1.

Чтобы пояснить вышеизложенное, рассмотрим функцию $f(x_1, x_2)$, которая графически представлена на рис. 5.4. Требуется минимизировать эту функцию при ограничении

$$g_1(x_1, x_2) = x_2 - b = 0,$$

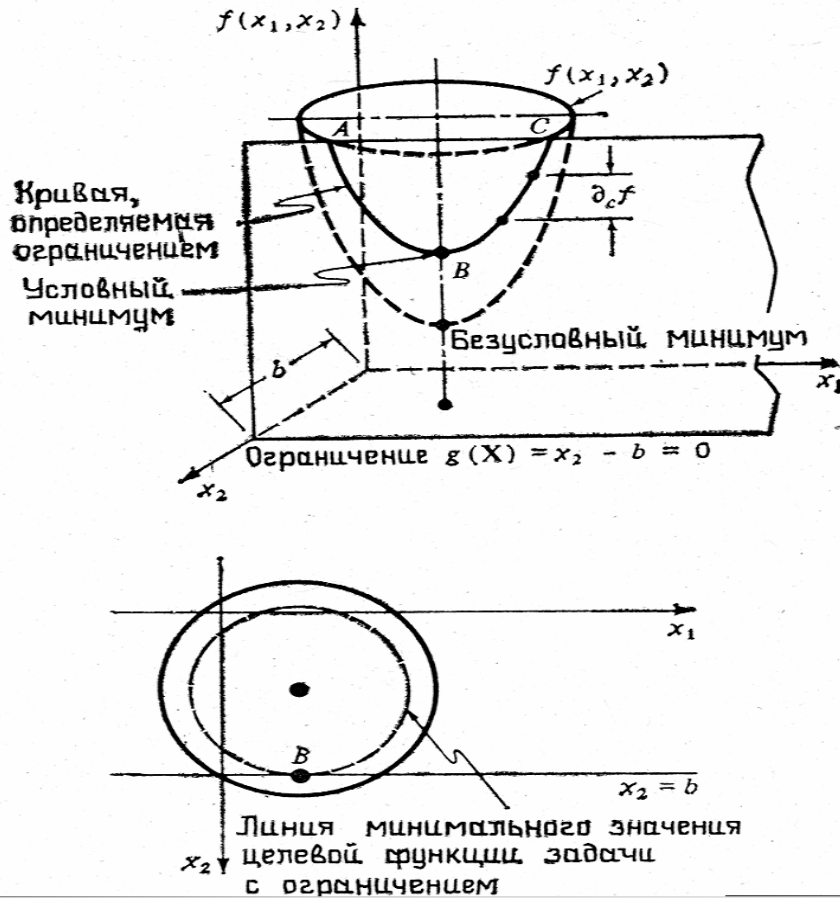


Рис. 5.4

где b – константа. Из рис. 5.4 видно, что значения $f(x_1, x_2)$ в точках, удовлетворяющих заданному ограничению, лежат на кривой, которая проходит через точки A , B и C . В соответствии с рассматриваемым методом определяются компоненты приведенного градиента функции $f(x_1, x_2)$ в каждой точке кривой ABC . Точка, в которой приведенные производные обращаются в нуль, является стационарной для исходной задачи с ограничением (точка B на рис. 5.4). Из рисунка также видно, как определяется бесконечно малое допустимое приращение $\partial_c f$ функции f .

Обратимся к задаче в общей постановке. Из теоремы Тейлора следует, что для точки $X + \Delta X$ из допустимой окрестности точки X можно записать следующие формулы:

$$f(X + \Delta X) - f(X) = \nabla f(X) \Delta X + o(\Delta X),$$

$$g(X + \Delta X) - g(X) = \nabla g(X) \Delta X + o(\Delta X).$$

При $\Delta x_j \rightarrow 0$ имеем

$$\partial f(X) = \nabla f(X) \partial X,$$

$$\partial g(X) = \nabla g(X) \partial X.$$

Так как $g(X) = 0$, то $\partial g(X) = 0$ в допустимой области. Отсюда

$$\partial f(X) - \nabla f(X) \partial X = 0, \quad \nabla g(X) \partial X = 0.$$

Полученная система содержит $(m+1)$ уравнений с $(n+1)$ неизвестными, которыми являются $\partial f(X)$ и ∂X . Неизвестную величину $\partial f(X)$ можно определить, если найден вектор ∂X . Это означает, что у нас имеется m уравнений с n неизвестными.

Если $m > n$, то по крайней мере $(m - n)$ уравнений оказываются избыточными. После устранения избыточности количество независимых уравнений в системе становится равным $m \leq n$. В случае когда $m = n$, решение тривиально: $\partial X = 0$. При этом точка X не имеет допустимой окрестности, и, следовательно, пространство решений состоит из единственной точки. Такая ситуация не представляет интереса. Оставшийся случай, когда $m < n$, необходимо рассмотреть подробно. Пусть $X = (Y, Z)$, где

$$Y = (y_1, y_2, \dots, y_m) \text{ и } Z = (z_1, z_2, \dots, z_{n-m})$$

являются соответственно *зависимыми* и *независимыми* переменными, образующими вектор X . В новых обозначениях градиенты функций f и g имеют следующий вид:

$$\nabla f(Y, Z) = (\nabla_Y f, \nabla_Z f)$$

$$\nabla g(Y, Z) = (\nabla_Y g, \nabla_Z g)$$

Введем определения двух матриц:

$$J = \nabla_Y g = \begin{pmatrix} \nabla_Y g_1 \\ \cdot \\ \cdot \\ \cdot \\ \nabla_Y g_m \end{pmatrix}, \quad C = \nabla_Z g = \begin{pmatrix} \nabla_Z g_1 \\ \cdot \\ \cdot \\ \cdot \\ \nabla_Z g_m \end{pmatrix}$$

Матрицу $J_{m \times m}$ называют матрицей Якоби, а $C_{m \times (n-m)}$ – матрицей управления. Предполагается, что матрица Якоби J невырожденная. Этот факт всегда имеет место, поскольку рассматриваемые m уравнений независимы по определению. Следовательно, компоненты вектора Y можно выбрать среди компонент X таким образом, что матрица J окажется невырожденной.

Используя данные выше определения, перепишем исходную систему уравнений с неизвестными $\partial f(X)$ и ∂X в следующем виде:

$$\partial f(Y, Z) = \nabla_Y f \partial Y + \nabla_Z f \partial Z$$

$$J \partial Y = -C \partial Z.$$

Так как матрица J – невырожденная, существует обратная матрица J^{-1} . Следовательно, $\partial Y = -J^{-1} C \partial Z$

Эти уравнения определяют ∂Y как функцию ∂Z (напомним, что Z является вектором независимых переменных). Замена ∂Y в уравнении для $\partial f(Y, Z)$ позволяет выразить ∂f через ∂Z :

$$\partial f(Y, Z) = (\nabla_Z f - \nabla_Y f J^{-1} C) \partial Z.$$

Из этого уравнения следует, что дифференцирование функции f по вектору независимых переменных Z дает формулу

$$\nabla_C f = \frac{\partial_C f(Y, Z)}{\partial_C Z} = \nabla_Z f - \nabla_Y f J^{-1} C,$$

где $\nabla_C f$ – приведенный градиент функции f . Вектор $\nabla_C f(Y, Z)$ должен обращаться в нуль в стационарных точках.

Достаточные условия наличия экстремума в стационарной точке аналогичны условиям, представленным в разд. 5.1. В этом случае элементы матрицы Гессе соответствуют компонентам вектора независимых переменных Z и являются *приведенными* вторыми производными, которые можно вычислить, воспользовавшись равенством

$$\nabla_c f = \nabla_z f - WC.$$

Вектор $\partial \nabla_c f / \partial z_i$ задает i -ю строку (приведенной) матрицы Гессе. Заметим, что W является функцией Y , а Y — функцией Z ; при этом

$$\partial Y = -J^{-1} C \partial Z$$

Таким образом, при вычислении частной производной $\nabla_c f$ по z_i следует применить к компонентам W правило дифференцирования сложной функции. Это означает, что

$$\partial w_j / \partial z_i = (\partial w_j / \partial y_j)(\partial y_j / \partial z_i).$$

Пример 5.2.1. В этом примере будет показано, как можно вычислить $\nabla_c f$ в заданной точке с помощью полученных выше формул. Последующий пример 5.2.2 иллюстрирует схему реализации метода приведенного градиента.

Рассмотрим задачу, в которой

$$f(X) = x_1^2 + 3x_2^2 + 5x_1x_3^2$$

$$g_1(X) = x_1x_3 + 2x_2 + x_2^2 - 11 = 0,$$

$$g_2(X) = x_1^2 + 2x_1x_2 + x_3^2 - 14 = 0.$$

Требуется оценить приращение функции $f (= \partial_c f)$ в допустимой окрестности допустимой точки $X^0 = (1, 2, 3)$. Пусть $Y = (x_1, x_3)$ и $Z = x_2$

Имеем

$$\nabla_Y f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_3} \right) = (2x_1 + 5x_3^2, 10x_1x_3),$$

$$\nabla_z f = \frac{\partial f}{\partial x_2} = 6x_2, \quad ,$$

$$J = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_3} \end{pmatrix} = \begin{pmatrix} x_3 & x_1 \\ 2x_1 + 2x_2 & 2x_3 \end{pmatrix} \quad C = \begin{pmatrix} \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_2 + 2 \\ 2x_1 \end{pmatrix}$$

Для того чтобы получить оценку приращения $\partial_c f$ в допустимой окрестности допустимой точки $X^0 = (1, 2, 3)$, вызванного малым изменением $\partial x_2 = 0,01$, следует вычислить

$$J^{-1}C = \begin{pmatrix} 3 & 1 \\ 6 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 6/12 & -1/12 \\ -6/12 & 3/12 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} \approx \begin{pmatrix} 2,83 \\ -2,5 \end{pmatrix}$$

Следовательно,

$$\partial_c f = (\nabla_z f - \nabla_y f J^{-1}C) \partial Z = \left(6(2) - (47, \quad 30) \begin{bmatrix} 2,83 \\ -2,5 \end{bmatrix} \right) \partial x_2 \approx -46 \partial x_2 = -0,46.$$

Если задана величина изменения ∂x_2 *независимой* переменной x_2 то допустимые значения ∂x_1 , и ∂x_3 *зависимых* переменных x_1 и x_3 определяются в соответствии с формулой

$$\partial Y = -J^{-1}C \partial Z$$

При $\partial x_2 = 0,01$ получаем

$$\begin{pmatrix} \partial x_1 \\ \partial x_3 \end{pmatrix} = -J^{-1}C \partial x_2 = \begin{pmatrix} -0,0283 \\ 0,025 \end{pmatrix}$$

Для того чтобы проверить точность полученной выше оценки $\partial_c f$, можно вычислить значения f в точках X^0 и $X^0 + \partial X$. При этом $X^0 + \partial X = (1 - 0,0283, 2 + 0,01, 3 + 0,025) = (0,9717, 2,01, 3,025)$. Отсюда следует, что $f(X^0) = 58$ и $f(X^0 + \partial X) = 57,523$ или $\partial_c f = f(X^0 + \partial X) - f(X^0) = -0,477$.

Полученное значение не совпадает с величиной $\partial_c f$, вычисленной вы-

ше. Различие между двумя результатами ($-0,477$ и $-0,46$) является следствием линейной аппроксимации фигурирующих в задаче нелинейных функций в окрестности точки X^0 . Поэтому использованную выше формулу можно применять лишь в случаях, когда отклонения от точки X^0 малы.

Пример 5.2.2. Данный пример служит иллюстрацией применения метода приведенного градиента. Рассмотрим задачу, в которой требуется

$$\text{минимизировать} \quad f(X) = x_1^2 + x_2^2 + x_3^2,$$

при ограничениях

$$g_1(X) = x_1 + x_2 + 3x_3 - 2 = 0,$$

$$g_2(X) = 5x_1 + 2x_2 + x_3 - 5 = 0.$$

Будем искать экстремальные точки целевой функции с учетом ограничений следующим образом. Положим

$$Y = (x_1, x_2) \text{ и } Z = x_3.$$

При этом

$$\nabla_Y f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (2x_1, 2x_2), \quad \nabla_Z f = \frac{\partial f}{\partial x_3} = 2x_3,$$

$$J = \begin{pmatrix} 1 & 1 \\ 5 & 2 \end{pmatrix}, \quad J^{-1} = \begin{pmatrix} -2/3 & 1/3 \\ 5/3 & -1/3 \end{pmatrix}, \quad C = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Следовательно,

$$\nabla_C f = \frac{\partial_C f}{\partial_C x_3} = 2x_3 - (2x_1, 2x_2) \begin{pmatrix} -2/3 & 1/3 \\ 5/3 & -1/3 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = (10/3)x_1 - (28/3)x_2 + 2x_3.$$

Стационарная точка (точки) определяется из условия $\nabla_C f = 0$, которое вместе с $g_1(X) = 0$ и $g_2(X) = 0$ образует систему уравнений

$$\begin{pmatrix} 10 & -28 & 6 \\ 1 & 1 & 3 \\ 5 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 5 \end{pmatrix}$$

Решением этой системы является вектор $X \approx (0,81, 0,35, 0,28)$.

Далее необходимо проверить выполнение достаточных условий Наличия экстремума в полученной стационарной точке. Так как x_3 – независимая переменная, из выражения для $\nabla_c f$ следует, что

$$\frac{\partial_c^2 f}{\partial_c x_3^2} = \frac{10}{3} \left(\frac{dx_1}{dx_3} \right) - \frac{28}{3} \left(\frac{dx_2}{dx_3} \right) + 2 = \left(\frac{10}{3}, -\frac{28}{3} \right) \begin{pmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{pmatrix} + 2$$

Метод Якоби позволяет получить формулу

$$\begin{pmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{pmatrix} = -J^{-1}C = \begin{pmatrix} 5/3 \\ -14/3 \end{pmatrix}$$

Из этих формул можно вычислить значение $\frac{\partial_c^2 f}{\partial_c x_3^2} = 460/9 > 0$, откуда X^0 –

точка минимума.

Одна из основных трудностей, возникающих при использовании метода Якоби, связана с получением J^{-1} в случае, когда количество ограничений достаточно велико. В такой ситуации можно воспользоваться правилом Крамера, которое позволяет выразить ∂f через ∂Z . Если z_j представляет собой j -ю компоненту Z , а y_i есть i -я компонента Y , то можно показать, что

$$\frac{\partial_c f}{\partial_c z_j} = \frac{\partial(f, g_1, \dots, g_m) / \partial(z_j, y_1, \dots, y_m)}{\partial(g_1, \dots, g_m) / \partial(y_1, \dots, y_m)}$$

где

$$\frac{\partial(f, g_1, \dots, g_m)}{\partial(z_j, y_1, \dots, y_m)} \equiv \begin{vmatrix} \frac{\partial f}{\partial z_j} & \frac{\partial f}{\partial y_1} & \cdot & \cdot & \cdot & \frac{\partial f}{\partial y_m} \\ \frac{\partial g_1}{\partial z_j} & \frac{\partial g_1}{\partial y_1} & \cdot & \cdot & \cdot & \frac{\partial g_1}{\partial y_m} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \frac{\partial g_m}{\partial z_j} & \frac{\partial g_m}{\partial y_1} & \cdot & \cdot & \cdot & \frac{\partial g_m}{\partial y_m} \end{vmatrix}$$

$$\frac{\partial(g_1, \dots, g_m)}{\partial(y_1, \dots, y_m)} \equiv \begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \cdot & \cdot & \cdot & \frac{\partial g_1}{\partial y_m} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \frac{\partial g_m}{\partial y_1} & \cdot & \cdot & \cdot & \frac{\partial g_m}{\partial y_m} \end{vmatrix} = |J|$$

Таким образом, необходимые условия принимают следующий вид:

$$\frac{\partial_c f}{\partial_c z_j} = 0, \quad j=1, 2, \dots, n-m.$$

Аналогично (i, j) -й элемент матрицы

$$\partial Y / \partial Z = -J^{-1} C$$

вычисляется в соответствии с формулой

$$\frac{\partial y_i}{\partial z_j} = \frac{\partial(g_1, \dots, g_m) / \partial(y_1, \dots, y_{i-1} z_j, y_{i+1}, \dots, y_m)}{\partial(g_1, \dots, g_m) \partial(y_1, \dots, y_m)},$$

которая выражает скорость изменения зависимой переменной y_i , вызываемого изменением независимой переменной z_j .

Наконец, чтобы получить достаточное условие наличия экстремума, необходимо выразить элементы вектора $W \equiv \nabla_y f J^{-1}$ через соответствующие определители. Для i -го элемента W имеем формулу

$$\omega_i = \frac{\partial(g_1, \dots, g_{i-1}, f, g_{i+1}, \dots, g_m) / \partial(y_1, \dots, y_m)}{\partial(g_1, \dots, g_m) \partial(y_1, \dots, y_m)}$$

Чтобы проиллюстрировать возможности применения описанного выше метода, установим необходимое условие наличия экстремума в задаче из примера 5.2.2. Получим

$$\frac{\partial_c f}{\partial_c x_3} = \frac{\begin{vmatrix} 2x_3 & 2x_1 & 2x_2 \\ 3 & 1 & 1 \\ 1 & 5 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 5 & 2 \end{vmatrix}} = \frac{10}{3}x_1 - \frac{28}{3}x_2 + 2x_3$$

[К содержанию](#)

5.2.1.2.. Метод множителей Лагранжа

Пусть $\lambda = \nabla_{y_0} f J^{-1} = \partial f / \partial g$, откуда $\partial f - \lambda \partial g = 0$.

Это уравнение отражает *необходимые* условия стационарности точек, так как формула для $\partial f / \partial g$ была получена с учетом предположения о том, что $\nabla_c f = 0$. Уравнение можно записать в более удобной форме, если перейти к частным производным по всем x_j , что приводит к системе

$$\frac{\partial}{\partial x_j} (f - \lambda g) = 0, j=1, 2, \dots, n.$$

Полученные уравнения вместе с ограничениями $g=0$ дают возможность определить допустимые векторы X и λ , которые удовлетворяют *необходимым* условиям стационарности.

Описанная выше процедура составляет основу *метода множителей Лагранжа*, который позволяет идентифицировать стационарные точки при решении оптимизационных задач с ограничениями в виде *равенств*. Схему этого

метода можно формально представить следующим образом.

Пусть

$$L(X, \lambda) = f(X) - \lambda g(X).$$

Функция L называется функцией Лагранжа. Параметры λ носят название множителей Лагранжа и, как следует из определения, имеют тот же смысл, что и коэффициенты чувствительности. Уравнения

$$\partial L / \partial \lambda = 0 \text{ и } \partial L / \partial X = 0$$

выражают рассмотренные выше необходимые условия наличия экстремума, которые порождаются функцией Лагранжа непосредственно. Это означает, что задача оптимизации с целевой функцией $f(X)$ при наличии ограничения $g(X)=0$ эквивалентна задаче нахождения безусловного экстремума функции Лагранжа $L(X, \lambda)$.

Достаточные условия, используемые при реализации метода множителей Лагранжа, формулируются ниже без доказательства.

Определим матрицу

$$H^B = \begin{pmatrix} 0 & P \\ P^T & Q \end{pmatrix}_{(m+n) \times (m+n)}, \text{ где } P = \begin{pmatrix} \nabla g_1(X) \\ \vdots \\ \nabla g_m(X) \end{pmatrix}_{m \times n} \text{ и } Q = \left\| \frac{\partial^2 L(X, \lambda)}{\partial x_i \partial x_j} \right\|_{n \times n} \quad \forall i, j.$$

Матрица H^B представляет собой так называемую окаймленную матрицу Гессе.

Пусть дана стационарная точка (X_0, λ_0) функции Лагранжа $L(X, \lambda)$, а окаймленная матрица Гессе H^B сформирована из значений соответствующих элементов в точке (X_0, λ_0) . Тогда X_0 является

1) точкой максимума, если, начиная с главного минора порядка $(m+1)$, последующие $(n-m)$ главных миноров матрицы H^B образуют знакопеременный

числовой ряд, знак первого члена которого определяется множителем $(-1)^{m+1}$;

2) точкой минимума, если, начиная с главного минора порядка $(m+1)$, знак последующих $(n-m)$ главных миноров матрицы H^B определяется множителем $(-1)^m$.

Сформулированные условия оказываются достаточными для идентификации экстремальной точки, но не являются необходимыми. Иными словами, стационарная точка, не удовлетворяющая этим условиям, может быть экстремальной.

Существуют другие условия для идентификации экстремальных точек, которые являются как необходимыми, так и достаточными. Однако практическое использование этих условий в ряде случаев связано со значительными вычислительными трудностями. Определим матрицу

$$\Delta = \left(\begin{array}{c|c} 0|P \\ \hline P^T|Q - \mu I \end{array} \right)$$

образованную значениями соответствующих функций в стационарной точке (X_0, λ_0) . Здесь μ – неизвестный параметр, а P и Q определены выше. Пусть $|\Delta|$ – определитель матрицы Δ ; тогда каждый из $(m-n)$ действительных корней u_i полинома $|\Delta|=0$ должен быть 1) отрицательным, если X_0 – точка максимума; 2) положительным, если X_0 – точка минимума.

Пример 5.2.3. минимизировать $f(X) = x_1^2 + x_2^2 + x_3^2$,
при ограничениях $g_1(X) = x_1 + x_2 + 3x_3 - 2 = 0$,
 $g_2(X) = 5x_1 + 2x_2 + x_3 - 5 = 0$.

В этом случае функция Лагранжа

$$L(X, \lambda) = x_1^2 + x_2^2 + x_3^2 - \lambda_1(x_1 + x_2 + 3x_3 - 2) - \lambda_2(5x_1 + 2x_2 + x_3 - 5).$$

Необходимые условия наличия экстремума записываются в виде следую-

щей системы уравнений:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 - \lambda_1 - 5\lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 2x_2 - \lambda_1 - 2\lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_3} = 2x_3 - 3\lambda_1 - \lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -(x_1 + x_2 + 3x_3 - 2) = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = -(5x_1 + 2x_2 + x_3 - 5) = 0.$$

Решением данной системы являются векторы

$$X_0 = (x_1^0, x_2^0, x_3^0) = (0,81, 0,35, 0,28),$$

$$\lambda_0 = (\lambda_1^0, \lambda_2^0) = (0,0867, 0,3067).$$

Чтобы показать, что в найденной точке имеет место минимум, рассмотрим матрицу

$$H^B = \left(\begin{array}{cc|ccc} 0 & 0 & 1 & 1 & 3 \\ \hline 0 & 0 & 5 & 2 & 1 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{array} \right)$$

Так как $n=3$ и $m=2$, то $n-m=1$. Таким образом, необходимо проверить только определитель H^B , знак которого должен определяться множителем $(-1)^2$ в точке минимума. Поскольку $\det H^B = 460 > 0$, X_0 является точкой минимума.

$$\det H^B = \begin{vmatrix} 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 5 & 2 & 1 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{vmatrix} = \begin{vmatrix} -4,5 & -1,5 & 1 & 1 & 0 \\ -1,5 & -0,5 & 5 & 2 & 0 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{vmatrix} = \begin{vmatrix} -5 & -2,5 & 1 & 0 & 0 \\ -2,5 & -1,5 & 5 & 0 & 0 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{vmatrix} =$$

$$= \begin{vmatrix} -5,5 & -5 & 0 & 0 & 0 \\ -5 & -15 & 0 & 0 & 0 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{vmatrix} = \begin{vmatrix} -3,8 & 0 & 0 & 0 & 0 \\ -5 & -15 & 0 & 0 & 0 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{vmatrix} = 460$$

Для решения систем уравнений, выражающих необходимые условия наличия экстремума, следует обратиться к соответствующим численным методам, например к методу Ньютона (разд. 5.1.2).

Пример 5.2.4. Рассмотрим следующую задачу:

минимизировать $f(X) = x_1^2 + x_2^2 + x_3^2$

при ограничении

$$4x_1 + x_2^2 + 2x_3 - 14 = 0.$$

При этом функция Лагранжа

$$L(X, \lambda) = x_1^2 + x_2^2 + x_3^2 - \lambda(4x_1 + x_2^2 + 2x_3 - 14).$$

Необходимые условия наличия экстремума имеют вид и системы:

$$\frac{\partial L}{\partial x_1} = -2x_1 - 4\lambda = 0, \quad x_1 = 2\lambda$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - 2\lambda x_2 = 0, \quad x_2(1 - \lambda) = 0$$

$$\frac{\partial L}{\partial x_3} = 2x_3 - 2\lambda = 0, \quad x_3 = \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(4x_1 + x_2^2 + 2x_3 - 14) = 0, \quad x_2^2 = 14 - 8\lambda - 2\lambda$$

подставляем $x_2 = \sqrt{14 - 10\lambda}$ во второе уравнение и получаем $\sqrt{14 - 10\lambda} (1 - \lambda) = 0$

В результате решениями являются векторы

$$(X_0, \lambda_0)_1 = (2, 2, 1, 1), \quad (1 - \lambda) = 0$$

$$(X_0, \lambda_0)_2 = (2, -2, 1, 1), \quad (1 - \lambda) = 0$$

$$(X_0, \lambda_0)_3 = (2, 8, 0, 1, 4, 1, 4) \quad \sqrt{14 - 10\lambda} = 0.$$

Чтобы использовать достаточные условия, вычислим элементы матрицы

$$H^B = \begin{pmatrix} 0 & 4 & 2x_2 & 2 \\ 4 & 2 & 0 & 0 \\ 2x_2 & 0 & 2 - 2\lambda & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}.$$

Поскольку $m=1$ и $n=3$, то стационарная точка является точкой минимума, если знак последних $(3-1)=2$ главных миноров определяется множителем $(-1)^m = -1$. Таким образом, в точке $(X_0, \lambda_0)_1 = (2, 2, 1, 1)$,

$$\begin{vmatrix} 0 & 4 & 4 \\ 4 & 2 & 0 \\ 4 & 0 & 0 \end{vmatrix} = -32 < 0 \text{ и } \begin{vmatrix} 0 & 4 & 4 & 2 \\ 4 & 2 & 0 & 0 \\ 4 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = -64 < 0.$$

В точке $(X_0, \lambda_0)_2 = (2, -2, 1, 1)$

$$\begin{vmatrix} 0 & 4 & -4 \\ 4 & 2 & 0 \\ -4 & 0 & 0 \end{vmatrix} = -32 < 0 \text{ и } \begin{vmatrix} 0 & 4 & -4 & 2 \\ 4 & 2 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = -64 < 0.$$

Наконец, в точке $(X_0, \lambda_0)_3 = (2, 8, 0, 1, 4, 1, 4)$

$$\begin{vmatrix} 4 & 4 & 0 \\ 4 & 2 & 0 \\ 0 & 0 & -0,8 \end{vmatrix} = 12,8 > 0 \text{ и } \begin{vmatrix} 0 & 4 & 0 & 2 \\ 4 & 2 & 0 & 0 \\ 0 & 0 & -0,8 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = 32 > 0.$$

Отсюда следует, что $(X_0)_1$ и $(X_0)_2$ – точки минимума. Тот факт, что $(X_0)_3$ не удовлетворяет достаточным условиям наличия максимума или минимума, не означает, что данная точка не является экстремальной. Ранее было отмечено, что используемые условия могут не выполняться в каждой экстремальной точке, хотя они и оказываются достаточными. В таком случае следует обратиться к достаточным условиям в другой формулировке.

Чтобы проиллюстрировать схему проверки достаточных условий наличия экстремума с использованием корней полинома $|\Delta|$ рассмотрим матрицу

$$\Delta = \begin{pmatrix} 0 & 4 & \frac{2x_2}{2} & 2 \\ 4 & 2 - \mu & 0 & 0 \\ 2x_2 & 0 & 2 - 2\lambda - \mu & 0 \\ 2 & 0 & 0 & 2 - \mu \end{pmatrix}$$

В точке $(X_0, \lambda_0)_1 = (2, 2, 1, 1)$ имеем $|\Delta| = 9\mu^2 - 26\mu + 16 = 0$, откуда $\mu = 2$ или $8/9$. Так как все $\mu > 0$, то $(X_0)_1 = (2, 2, 1)$ – точка минимума. Далее, в точке $(X_0, \lambda_0)_2 = (2, -2, 1, 1)$, $|\Delta| = 9\mu^2 - 26\mu + 16 = 0$, что совпадает с рассмотренным выше случаем.

Следовательно, $(X_0)_2 = (2, -2, 1)$ – точка минимума. Наконец, в точке $(X_0, \lambda_0)_3 = (2, 8, 0, 1, 4, 1, 4)$ $|\Delta| = 5\mu^2 - 6\mu - 8 = 0$, откуда $\mu = 2$ и $-0,8$. Это означает, что $(X_0)_3 = (2, 8, 0, 1, 4)$ не является экстремальной точкой.

[К содержанию](#)

5.2.2. Ограничения в виде неравенств

Ниже коротко рассматривается лишь небольшая часть методов решения задач нелинейного программирования с ограничениями, как в форме равенств, так и неравенств. Однако приводимые ниже алгоритмы имеют в настоящее время наибольшее практическое применение. Относится это, прежде всего, к методам штрафных и барьерных функций, описанию которых посвящен отдельный подраздел.

5.2.2.1. Метод Франка-Вулфа

Пусть требуется найти максимальное значение вогнутой функции:

$$\max f(X),$$

при ограничениях

$$AX \leq b$$

$$X \geq 0$$

Характерной особенностью этой задачи является то, что ее система ограничений содержит только линейные неравенства. Эта особенность является основой для замены в окрестности исследуемой точки нелинейной целевой функции линейной, благодаря чему решение исходной задачи сводится к последовательному решению задач линейного программирования.

Процесс нахождения решения задачи начинается с определения точки, принадлежащей области допустимых решений; задачи. Пусть это точка X^k , тогда в этой точке вычисляется градиент функции $f(X)$

$$\nabla f(X^k) = \left(\frac{\partial f(X^k)}{\partial x_1}, \frac{\partial f(X^k)}{\partial x_2}, \dots, \frac{\partial f(X^k)}{\partial x_n} \right)$$

и определяется линейная функция

$$F^k = \frac{\partial f(X^k)}{\partial x_1} x_1 + \frac{\partial f(X^k)}{\partial x_2} x_2 + \dots + \frac{\partial f(X^k)}{\partial x_n} x_n$$

Затем определяется максимальное значение этой функции при ограничениях $AX \leq b$ и $X \geq 0$. Пусть решение данной задачи определяется точкой $\tilde{X}^k$. Тогда за новое допустимое решение исходной задачи принимаются координаты точки $X^{(k+1)}$:

$$X^{(k+1)} = X^k + \lambda_k (\tilde{X}^k - X^k), \quad (5.3)$$

где λ_k — некоторое число, называемое шагом вычислений и заключенное между нулем и единицей ($0 \leq \lambda_k \leq 1$). Это число λ_k определяется таким образом, чтобы значение функции $f(X^{(k+1)})$ в точке $X^{(k+1)}$, зависящее от λ_k , было максимальным.

Для этого необходимо найти решение уравнения $\frac{\partial f(\lambda)}{\partial \lambda_k} = 0$ и выбрать его наименьший корень. Если значение корня больше единицы, то следует положить $\lambda_k = 1$. После определения числа λ_k находятся координаты точки $X^{(k+1)}$, вычисляется значение целевой функции в ней и выясняется необходимость перехода к новой точке $X^{(k+1)}$. Если такая необходимость имеется, то в точке $X^{(k+1)}$ вычисляется градиент целевой функции, после чего осуществляется переход к соответствующей задаче линейного программирования и определяется ее решение. Определяются координаты точки $X^{(k+2)}$ и исследуется необходимость проведения дальнейших вычислений. После конечного числа шагов получается с необходимой точностью решение исходной задачи.

Таким образом, процесс нахождения решения задачи рассмотренного класса методом Франка—Вулфа включает следующие этапы:

1. Определение исходного допустимого решения задачи.
2. Вычисление градиента функции $\nabla f(X^k)$ в точке допустимого решения.
3. Построение функции F^k и определение ее максимального значения при условиях $AX \leq b$ и $X \geq 0$.
4. Определение шага вычислений.

5. По формулам $X^{(k+1)} = X^k + \lambda_k (\tilde{X}^k - X^k)$ определение компонент нового допустимого решения.

6. Проверка необходимости перехода к последующему допустимому решению. В случае необходимости переход к этапу 2, в противном случае найдено приемлемое решение исходной задачи.

[К содержанию](#)

5.2.2.2. Условия Куна – Таккера

В данном подразделе рассматриваются *необходимые* условия Куна – Таккера, позволяющие идентифицировать стационарные точки в нелинейной задаче с ограничениями в виде неравенств. При изложении используется метод множителей Лагранжа. Указанные условия являются также достаточными, если целевая функция и область допустимых решений обладают некоторыми специальными свойствами, которые рассматриваются ниже.

Рассмотрим следующую задачу:

найти $\max \quad z = f(X)$
при ограничениях $g(X) \leq 0$.

Ограничения-неравенства можно преобразовать к виду равенств путем введения соответствующих *неотрицательных* дополнительных переменных. Таким образом, автоматически учитывая требование неотрицательности, определим $S_i^2 (\geq 0)$ как дополнительную переменную, которую следует прибавить к левой части i -го ограничения $g_i(X) \leq 0$. Пусть

$$S = (S_1, S_2, \dots, S_m)^T \text{ и } S^2 = (S_1^2, S_2^2, \dots, S_m^2)^T$$

где m – общее количество ограничений-неравенств. При этом функция Лагранжа записывается в следующем виде:

$$L(X, S, \lambda) = f(X) - \lambda [g(X) + S^2],$$

при заданных ограничениях $g(X) \leq 0$.

Необходимым условием оптимальности является неотрицательность (неположительность) λ в задаче максимизации (минимизации). Этот результат устанавливается следующим образом. Рассмотрим задачу максимизации. Множители λ выражают скорость изменения f по отношению к изменениям g , т. е. $\lambda = \partial f / \partial g$. Как только правая часть ограничения $g \leq 0$ увеличивается и становится большей нуля, область допустимых решений расширяется. Следовательно, оптимальное значение целевой функции не может уменьшиться. Это означает, что $\lambda \geq 0$. Аналогично при увеличении правой части ограничения в задаче минимизации оптимальное значение f не может увеличиться, откуда следует, что $\lambda \leq 0$. Если же ограничения заданы в виде равенств $g(X)=0$, то на знак λ никаких условий не накладывается.

Предъявленные требования к вектору λ следует рассматривать как часть необходимых условий Куна – Таккера. Далее будут установлены остальные условия.

Вычислив частные производные функции L по X , S и λ и приравняв их к нулю, получим

$$\begin{aligned}\partial L / \partial X &= \nabla f(X) - \lambda \nabla g(X) = 0, \\ \partial L / \partial S_i &= -2 \lambda_i S_i = 0, \quad i=1, 2, \dots, m, \\ \partial L / \partial \lambda &= -(g(X) + S^2) = 0.\end{aligned}$$

Исследование второй группы уравнений приводит к следующим выводам.

1. Если λ_i больше нуля, то $S_i^2=0$. Это означает, что соответствующий рассматриваемому ограничению ресурс является дефицитным и, следовательно, исчерпан полностью (ограничение становится равенством).

2. Если $S_i^2 > 0$, то $\lambda_i = 0$. Это означает, что i -й ресурс не является дефицит-

ным, и, следовательно, изменение его количества не оказывает влияния на значение $f(\lambda_i = \partial f / \partial g_i = 0)$.

Из второй и третьей групп уравнений следует, что

$$\lambda_i g_i(X) = 0, i=1, 2, \dots, m.$$

Полученные условия в основном подтверждают сделанный ранее вывод, а именно если $\lambda_i > 0$, то $g_i(X) = 0$ или $S_i^2 = 0$. Аналогично, если $g_i(X) < 0$, т. е. $S_i^2 > 0$, то $\lambda_i = 0$.

Теперь можно сформулировать необходимые условия Куна – Таккера, которым должны удовлетворять X и λ , определяющие стационарную точку в задаче максимизации:

$$\begin{aligned} \lambda &\geq 0, \\ \nabla f(X) - \lambda \nabla g(X) &= 0, \\ \lambda_i g_i &= 0, i=1, 2, \dots, m, \\ g(X) &\leq 0. \end{aligned}$$

Эти условия применимы также к задаче минимизации, если вместо условия неотрицательности λ ввести требование неположительности, необходимость которого была установлена ранее. Если же ограничения заданы в виде *равенств*, то на знак соответствующих множителей Лагранжа никаких условий не накладывается как в случае максимизации, так и в случае минимизации.

Достаточность условий Куна – Таккера

Необходимые условия Куна – Таккера являются также достаточными, если целевая функция и область допустимых решений обладают определенными свойствами, связанными с выпуклостью и вогнутостью. Эти свойства указаны в табл. 5.1.

Таблица 5.1

Тип оптимизации	Требуемые свойства	
	целевая функция	область допустимых решений
Максимизация	Вогнутая	Выпуклое множество
Минимизация	Выпуклая	Выпуклое множество

Обычно легче определить, является ли целевая функция выпуклой или вогнутой, чем доказать, что область допустимых решений представляет собой выпуклое множество. Между тем выпуклость области допустимых решений может быть установлена путем исследования функций, формирующих ограничения. В связи с этим ниже приводится перечень соответствующих требований, проверка выполнения которых на практике не представляет больших трудностей. Для того чтобы установить эти требования, сформулируем задачу нелинейного программирования в общей постановке:

(максимизировать или минимизировать)

$$Z=f(X)$$

при ограничениях

$$g_i(X) \leq 0, \quad i=1, 2, \dots, r,$$

$$g_i(X) \geq 0, \quad i=r+1, \dots, p,$$

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

При этом

$$L(X, S, \lambda) = f(X) - \sum_{i=1}^r \lambda_i [g_i(X) + S_i^2] - \sum_{i=r+1}^p \lambda_i [g_i(X) - S_i^2] - \sum_{i=p+1}^m \lambda_i g_i(X),$$

где λ_i – множитель Лагранжа, ассоциированный с i -м ограничением. Требования, которые устанавливают достаточность условий Куна – Таккера, приведены в табл. 5.2.

Таблица 5.2

Тип оптимизации	Требования		
	$f(X)$	$g_i(X)$	λ_i
Максимизация	Вогнутая	Выпуклая	$\geq 0 \quad (1 \leq i \leq r)$
		Вогнутая	$\leq 0 \quad (r+1 \leq i \leq p)$
		Линейная	Нет ограничений $(p+1 \leq i \leq m)$
Минимизация	Выпуклая	Выпуклая	$\leq 0 \quad (1 \leq i \leq r)$
		Вогнутая	$\geq 0 \quad (r+1 \leq i \leq p)$
		Линейная	Нет ограничений $(p+1 \leq i \leq m)$

Следует заметить, что табл. 5.2 не охватывает все случаи, упомянутые в табл. 5.1. Это связано с тем обстоятельством, что область допустимых решений может быть выпуклой и в то же время не соответствовать перечисленным в табл. 5.2 требованиям к функциям $g_i(X)$.

Требования, приведенные в табл. 5.2, обусловлены тем, что соответствующие ограничения порождают вогнутую функцию Лагранжа $L(X, S, \lambda)$ в случае максимизации и выпуклую функцию $L(X, S, \lambda)$ в случае минимизации. Справедливость этого утверждения можно проверить непосредственно, учитывая следующий очевидный факт: если $g_i(X)$ – выпуклая функция, то функция $\lambda_i g_i(X)$ – оказывается выпуклой при $\lambda_i \geq 0$ и вогнутой при $\lambda_i \leq 0$. С помощью аналогичных рассуждений нетрудно обосновать все остальные требования. Следует особо отметить, что линейная функция по определению является как выпуклой, так и вогнутой. Заметим также, что функция $-f$ выпуклая, если функция f вогнутая, и наоборот.

Пример 5.2.5. Рассмотрим следующую задачу минимизации:

$$f(X) = x_1^2 + x_2^2 + x_3^2, \rightarrow \min$$

при ограничениях

$$g_1(X) = 2x_1 + x_2 - 5 \leq 0,$$

$$g_2(X) = x_1 + x_3 - 2 \leq 0,$$

$$g_3(X) = 1 - x_1 \leq 0,$$

$$g_4(X) = 2 - x_2 \leq 0,$$

$$g_5(X) = -x_3 \leq 0.$$

Поскольку мы имеем дело с задачей минимизации, то $\lambda \leq 0$. Запишем условия Куна – Таккера:

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \leq 0$$

$$(2x_1, 2x_2, 2x_3) - (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = 0$$

$$\lambda_1 g_1 = \lambda_2 g_2 = \dots = \lambda_5 g_5 = 0$$

$$g(X) \leq 0.$$

После упрощений получаем

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \leq 0, \quad \lambda_3(1 - x_1) = 0$$

$$2x_1 - 2\lambda_1 - \lambda_2 + \lambda_3 = 0, \quad \lambda_4(2 - x_2) = 0,$$

$$2x_2 - \lambda_1 + \lambda_4 = 0, \quad \lambda_5 x_3 = 0,$$

$$2x_3 - \lambda_2 + \lambda_5 = 0 \quad 2x_1 + x_2 \leq 5$$

$$\lambda_1(2x_1 + x_2 - 5) = 0, \quad x_1 + x_3 \leq 2.$$

$$\lambda_2(x_1 + x_3 - 2) = 0$$

$$x_1 \geq 1, \quad x_2 \geq 2, \quad x_3 \geq 0,$$

откуда $x_1 = 1, \quad x_2 = 2, \quad x_3 = 0$; $\lambda_1 = \lambda_2 = \lambda_5 = 0, \quad \lambda_3 = -2, \quad \lambda_4 = -4$. Так как и функция $f(X)$, и область допустимых решений $g(X) \leq 0$ являются выпуклыми, то функция $L(X, S, \lambda)$ также должна быть выпуклой, и найденная стационарная точка определяет глобальный условный минимум.

[К содержанию](#)

5.2.3. Методы штрафных и барьерных функций

Дополним принятые ранее обозначения.

- Пусть $x, s, g(x)$ – n мерные векторы.
- Через X и $G(x)$ обозначим диагональные матрицы $diag(x)$ и $diag(g(x))$.
- Векторы с координатами $1/x_i$ и $1/g_i(x)$ обозначим соответственно x^{-1} и $g(x)^{-1}$.
- Через e обозначим единичный вектор, размерность e явствует из контекста.
- Скалярное произведение векторов x и s , как и ранее обозначим – xs . Этому произведению эквивалентны следующие: XSe, SXe, Xs и Sx .

Пусть $g_i \ i=1,...,l$ множество функций, определенных на векторах из R^n . И пусть x – допустимая точка на множестве ограничений $\{g_i(x) \leq 0, i=1,...,l\}$.

- Через g обозначим вектор функций векторного действительного аргумента $g_i \ i \in \{1,...,l\}$.
- g_L обозначим вектор функций векторного действительного аргумента g_i для $i \in L$, где L – подмножество из $\{1,...,l\}$.
- ∇g обозначим транспонированный якобиан, т.е. $\frac{\partial g^T}{\partial x}$.
- $A(x)$ обозначим множество индексов i для которых ограничение g_i активно на x . Другими словами $A(x) = \{i=1,...,l \mid g_i(x) = 0\}$.

Основная идея методов штрафных функций, как и их частного случая – барьерных методов, заключается в том, что исходная задача оптимизации с ограничениями преобразуется к эквивалентной задаче без ограничений, либо к задаче с простыми ограничениями.

Проиллюстрируем данную схему на примере следующей общей задачи нелинейного программирования (НЛП):

найти

$$\min f(x),$$

$$\begin{aligned} \text{при ограничениях} \quad & g(x) \leq 0, \\ & h(x) = 0, \\ & x \in R^n. \end{aligned}$$

В любом барьерном методе строится вспомогательная целевая функция $f(x) + \mu \cdot \alpha(x)$, где μ - барьерный параметр и $\alpha(x)$ определяется как

$$\alpha(x) = \sum_{i=1}^m \Phi(g_i(x)) + \sum_{i=1}^l \Psi(h_i(x)).$$

Здесь Φ и Ψ - конечные функции, для которых выполняются условия:

$\Phi(y) = 0$ если $y \leq 0$ и $\Phi(y) > 0$ если $y > 0$,

$\Psi(y) = 0$ если $y = 0$ и $\Psi(y) > 0$ если $y \neq 0$.

В качестве примера функции, удовлетворяющей перечисленным свойствам можно привести следующую функцию:

$$\alpha(x) = \sum_{i=1}^m [\max(0, g_i(x))]^p + \sum_{i=1}^l |h_i(x)|^p, \text{ где } p - \text{положительное целое.}$$

Посредством этой вспомогательной функции ограничения включаются в состав целевой функции и взвешиваются с помощью положительного барьерного параметра μ , который накладывает штраф на любое нарушение ограничений таким образом, что решения последовательности образующихся задач без ограничений сходятся к решению исходной задачи. При этом используются хорошо известные методы оптимизации без ограничений (см. выше).

Интерес представляют такие последовательности, что:

- Каждая задача последовательности решается проще исходной задачи. В частности методы конструирования задач без ограничений должны позволять достаточно просто решать эти задачи.
- Решение любой задачи из последовательности должно служить хорошей стартовой точкой для поиска решения последующих задач, вплоть до получения решения исходной задачи.

Особого внимания заслуживают методы, использующие барьерные функции специального вида (так называемые самосогласующиеся и саморегулируемые функции). Исторически так случилось, что большая группа барьерных методов, в которых применены упомянутые функции, получила общее название методов внутренних точек (МВТ).

[К содержанию](#)

5.2.3.1. Введение в методы внутренних точек

Первоначально в МВТ использовалась логарифмическая барьерная функция $\alpha(x)$ ¹:

$$f(x) - \mu \sum \log(-g_i(x)) \rightarrow \min,$$

$$\text{при ограничениях:} \quad h(x) = 0.$$

В частности, для задач ЛП:

$$\begin{aligned} c^T x &\rightarrow \min \\ \text{при ограничениях} \quad Ax &= b \\ x &\geq 0, \end{aligned} \tag{5.1}$$

где $c, x \in R^n$, $b \in R^m$, $m \leq n$, и A – матрица размерностью $m \times n$ ранга m , будем иметь:

$$\begin{aligned} c^T x - \mu \sum_{i=1}^n \ln x_i &\rightarrow \min \\ \text{при ограничениях} \quad Ax &= b \\ x &> 0 \end{aligned} \tag{5.2}$$

К настоящему времени в качестве обобщения была разработана теория самосогласующихся штрафных функций[4] и саморегулируемых барьерных функций [5].

Далее мы сосредоточимся на задаче выпуклого программирования, в которой ограничения – равенства $h(x)$ имеют специальную форму линейных ограничений. Для начала без потери общности мы будем считать эти линейные ограничения количественно малосущественными, в связи с чем будем рассматривать следующую задачу выпуклого программирования.

Найти:

$$\min \quad f(x),$$

при ограничениях
$$g(x) \leq 0 \quad (5.4)$$

В этом случае применение МВТ к (5.4) состоит в формировании последовательности задач минимизации без ограничений следующего вида:

Найти:

$$\min f_{\mu}(x), \quad (5.5)$$

где $f_{\mu}(x) = f(x) - \mu \sum \log(-g_i(x))$. Строгая положительность $-g_i(x)$ не вызывает сомнений и определяет допустимую область, которая является открытой.

Анализ барьерных методов удобно проводить в терминах вектора – градиента ∇f_{μ} и гессиана $\nabla^2 f_{\mu}$. Определим данные объекты для барьерной задачи с параметром μ :

$$\nabla f_{\mu}(x) = \nabla f(x) - \mu \nabla g(x) G(x)^{-1} e = \nabla f(x) - \mu \sum_i \frac{\nabla g_i(x)}{g_i(x)},$$

$$\nabla^2 f_{\mu}(x) = \nabla^2 f(x) + \mu \sum_i \left(\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right).$$

5.2.3.2. Метод внутренних точек, как преобразование условий Куна – Таккера

К содержанию

Рассмотрим МВТ с другой точки зрения, а именно, как непосредственное применение необходимых и достаточных условий существования оптимального решения. Начнем анализ с задачи ЛП (5.1), для которой условия оптимальности можно записать следующим образом:

¹ Впервые введена Фришем[1]

$$\begin{aligned}
Ax &= b, \\
A^T y + s &= c, \\
x^T s &= 0, \\
x, s &\geq 0.
\end{aligned} \tag{5.6}$$

В то же время условия Куна-Таккера для барьерной задачи (5.2) запишутся следующим образом:

$$\begin{aligned}
Ax &= b, \\
A^T y + \mu X^l e &= c, \\
x &> 0,
\end{aligned} \tag{5.7}$$

Следовательно, если мы определим s как неотрицательный вектор, такой, что $s = \mu X^l e$, то (5.7) можно привести к следующей системе уравнений:

$$\begin{aligned}
Ax &= b, \\
A^T y + s &= c, \\
Xs &= \mu e, \\
x, s &\geq 0,
\end{aligned} \tag{5.8}$$

которая, на самом деле, является преобразованием системы (5.6). Заметим, что условия дополняющей нежесткости $Xs = \mu e$ влекут равенство $x^T s = n \mu$, и таким образом значение μ для точки, удовлетворяющей системе (5.8), является мерой близости к решению (5.6).

В случае задачи НЛП вспомним, что ограничениям, обращающимся в строгие равенства в точке оптимума x^* , единственным образом соответствует вектор y^* , такой, что выполняются следующие условия оптимальности для задачи (5.4):

$$\begin{aligned}
g(x^*) &\leq 0, \\
\nabla f(x^*) - \sum y_i^* \nabla g_i(x^*) &= 0,
\end{aligned} \tag{5.9}$$

$$\begin{aligned} Y^* g(x^*) &= 0, \\ y^* &\leq 0. \end{aligned}$$

Условия Куна - Таккера для барьерной задачи (5.5) запишутся следующим образом:

$$\begin{aligned} g(x_\mu^*) &< 0, \\ \nabla f(x_\mu^*) - \sum \frac{\mu}{g_i(x_\mu^*)} \nabla g_i(x_\mu^*) &= 0. \end{aligned} \quad (5.10)$$

Очевидно, что y является отрицательным вектором, таким, что $y = \mu g(x_\mu^*)^{-1}$, что эквивалентно $Yg(x_\mu^*) = \mu e$. С учетом этого перепишем (5.10):

$$\begin{aligned} g(x_\mu^*) &< 0, \\ \nabla f(x_\mu^*) - \sum y_i^* \nabla g_i(x_\mu^*) &= 0, \\ Yg(x_\mu^*) &= \mu e. \end{aligned} \quad (5.11)$$

Система (5.11) является преобразованной системой (5.9) с параметром μ в условиях дополняющей нежесткости.

Резюмируя сказанное, можно сделать вывод о том, что МВТ можно рассматривать с двух точек зрения – и как метод штрафных функций, и как гомотопный² метод. В обоих случаях решение задачи (5.5) по параметру μ сходится к решению (5.4).

Решение (5.5) носит название μ – центра.

МВТ использует μ – центры для различных значений μ вплоть до получения оптимального решения. В следующем разделе вводятся понятия w -центров и взвешенного центрального пути, называемого (w) -центральной путем, который является обобщением ОДР для (5.5) и определяет условия, при которых

² Гомотопия – свойство непрерывных отображений принадлежать к одному семейству непрерывных отображений, непрерывно зависящих от параметра.

последовательность w -центров сходится к оптимальному решению.

5.2.3.3. (w)-центральный путь

В предыдущем подразделе пояснялся алгоритм МВТ с точки зрения преобразований условий Куна – Таккера. На практике эти уравнения многократно модифицируются путем преобразования вектора μe из (5.11) в вектор w . Соответствующее решение (5.11) называется мишенью и чаще всего не лежит на центральном пути. Смысл данного понятия поясняется ниже. Понятие мишени было введено для линейного программирования [6] и затем обобщено на случай выпуклого программирования [7]. Нижеследующие определения и теоремы³ связаны с задачей, которую мы назовем задачей $\mathcal{P}$, и которая записывается следующим образом.

Найти

$$\begin{aligned} \min \quad & f(x), \\ \text{при ограничениях:} \quad & Ax = b, \\ & g(x) \leq 0, \\ & x \in R^n, \end{aligned} \tag{5.12}$$

где f и g – выпуклые функции, $x \in R^n$, $A \in R^{m_1 \times n}$, $b \in R^{m_1}$, $g(x) \in R^{m_2}$.

Решение задачи (5.12) обозначим $opt(\mathcal{P})$.

Определение 5.2.3.1 (Взвешенный центральный путь и μ - w -центры).

Пусть имеется вектор фиксированных весов $w \in R_{++}^{m_2}$. w -центральный путь из (5.12) порождает следующее параметризованное семейство логарифмических барьерных задач:

$$\mathcal{P}_w(\mu) \quad \inf \left\{ f(x) - \mu \sum_{j=1}^{m_2} w_j \log(-g_j(x)) \right\} \quad \text{над } x \in P^0 \},$$

где $\mu \in R_{++}$ - параметр семейства и $P^0 = \{x \in R^n \mid Ax = b, g(x) < 0\}$. Решения

$P_w(\mu)$ называются μ - w -центрами. Если каждая задача $P_w(\mu)$ имеет единственное решение $x(\mu)$, то это решение называется μ - w -центром (5.12) и путь $\mu > 0 \rightarrow x(\mu) \in P^0$ называется w -центральной путем, ассоциированным с (5.12).

При $\mu \rightarrow \infty$ в пределе задача может иметь единственное решение, интерпретируемое как w -центр (5.12), определяемое ниже. Если w – единичный вектор, то w -центральный путь называется центральным путем, а w -центр называется аналитическим центром P^0 .

Определение 5.2.3.2 (w -центр $opt(P)$) Вернемся к (5.12) и предположим, что $opt(P)$ не пусто и ограничено. w -центр $opt(P)$ определяет единственное оптимальное решение следующей задачи выпуклого программирования:

$$x^* = \max \sum_{j \in B} w_j \log |g_j(x)| \quad \text{над } \{x \in opt(P) \mid g_B(x) < 0\},$$

где B – множество индексов $\{j \mid g_j(x) < 0 \text{ для некоторого } x \in opt(P)\}$.

Понятие w -центра имеет хорошую геометрическую интерпретацию. Для упрощения описания, положим w – единичным вектором и вернемся к рассмотрению задачи ЛП (5.1), предположив, что МДР последней ограничен, и что существует $\tilde{x}$, такой, что $Ax=b$ и $x>0$.

Оптимальное решение (5.2), соответствующее $\mu \rightarrow \infty$ является аналитическим центром (5.2). Обозначим его, как x_∞ . Геометрически x_∞ максимизирует скалярное произведение расстояний от любой допустимой точки до границ МДР (5.1). Для конечного значения μ можно показать, что соответствующее решение (5.2) также максимизирует скалярное произведение всех расстояний от любой точки S до границ некоторого связанного с ней многогранника. В этом случае x_μ – также аналитический центр, хотя и не аналитический центр внутренности области допустимых решений (ОДР).

³ Теоремы приводятся только с ознакомительной целью и без доказательств.

В случае НЛП интерпретация сложнее, т.к. ОДР не обязательно ограничена гранями многогранника. Однако эта интерпретация может быть обобщена заменой логарифмической барьерной функции на барьерную функцию B , которая является ν – само регулируемой для P^0 , где ν – скалярная характеристика барьера. Нестеров и Немировский [3] показали, что если P^0 не пусто и если допустимое множество ограничено, то аналитический центр x_∞ на P^0 может быть определен как минимум B над P^0 . Эти посылки означают, что любая допустимая точка x содержится в шаре с центром в x_∞ и радиусом, зависящим только от ν , или более точно $\|x - x_\infty\|_{x_\infty} \leq \nu + 2\nu^{1/2}$, где $\|\bullet\|_{x_\infty}$ – норма гессиана барьерной функции F .

Ниже представлен ряд условий, которые определяют μ -w-центры, основанные на логарифмических барьерах. Теорема 5.2.3.1 представляет практический интерес, т.к. определяет достаточные условия, которые могут быть проверены в прямом пространстве решений.

Теорема 5.2.3.1 (Достаточные условия)

$$\left[\begin{array}{l} f, g \text{ выпуклы и дифференцируемы} \\ A \text{ имеет полный ранг по строкам} \\ \exists x \mid Ax = b, g(x) < 0 \\ \{x \mid g(x) \leq 0\} \text{ ограничена} \end{array} \right] \Rightarrow \left[\begin{array}{l} P_w(\mu) \text{ имеет единственное} \\ \text{решение для каждого } \mu > 0 \text{ и усло-} \\ \text{вия Куна-Таккера гарантируют оп-} \\ \text{тимальность} \end{array} \right]$$

Теорема 5.2.3.2 (Необходимые и достаточные условия)

Предположим, что A имеет полный по строкам ранг, и что $opt(P)$ не пусто и ограничено, тогда имеет место следующие эквивалентные отношения:

$$(PD^0 \neq \emptyset) \Leftrightarrow (opt(P_w(\mu)) \neq \emptyset \forall \mu > 0),$$

где PD^0 – множество прямо-двойственных внутренних допустимых решений (5.12):

$$PD^0 = \{(x, v, y) : x \in P^0, v \in R^{m_1}, y \in R_{++}^{m_2} \mid Ax = b, g(x) < 0, \nabla f(x) - A^T v + \nabla g(x)y = 0\}.$$

Заметим, что в ЛП хорошо известен результат, что если PD^0 не пусто, то МДР обязательно не пуст и ограничен. Это, однако, не распространяется на случай НЛП.

Если далее допустить, что f и g бесконечно дифференцируемы, то по утверждению следующей теоремы, $\{x(\mu) : \mu \rightarrow 0_+\}$ сходится к w -центру $opt(\mathcal{P})$.

Теорема 5.2.3.3 [19] (Сходимость центрального пути). Предположим, что A имеет полный ранг по строкам, ОДР (5.12) не пуста и ограничена, f и g бесконечно дифференцируемы, тогда $(PD^0 \neq \emptyset) \Rightarrow \{x(\mu) : \mu \rightarrow 0_+\}$ сходится к w -центру ОДР (5.12).

Теорема 5.2.3.4 [19] (Сходимость центрального пути). Предположим, что A имеет полный ранг по строкам, ОДР (5.12) содержит точно дополняющие решения⁴, $\Rightarrow \{x(\mu) : \mu \rightarrow 0_+\}$ сходятся к w -центру ОДР (5.12).

Обе теоремы объясняют использование барьерного параметра μ , который приближается к нулю в процессе работы алгоритма МВТ: μ -центров или μ - w -центров, где w – единичный вектор, сходящихся к решению при некоторых нестрогих допущениях.

На практике МВТ не находят точных μ -центров для текущих значений μ начиная с исходной мишени и до получения оптимального решения. В этом нет необходимости, т.к. оптимальные решения чаще всего не лежат на центральном пути. Для данного μ МВТ генерируются точки, которые покрывают μ -центр и затем меняются значения барьерного параметра. В связи с этим на каждом шаге изменяется определение окрестности центрального пути.

С целью упрощения во всех нижеследующих определениях в качестве w используем единичный вектор.

Определение 5.2.3.3 (2–норма окрестности) Для некоторого $\theta > 0$ 2–норма окрестности $N_2(\theta)$ определяется следующим образом:

$$N_2(\theta) = \{ (x, v, y) \in PD^0 \mid \|G(x)y - \mu e\|_2 \leq \theta \mu \},$$

Определение 5.2.3.4 (∞ –норма окрестности) Для некоторого $\gamma \in (0, 1)$, односторонняя ∞ -норма окрестности $N_\infty(\gamma)$ определяется следующим образом:

$$N_\infty(\gamma) = \{ (x, v, y) \in PD^0 \mid G(x)y \geq \gamma \mu e \}.$$

Определение 5.2.3.5 (недопустимая⁵ ∞ – норма окрестности) Для некоторого $\gamma \in (0, 1)$ и $\beta > 0$ расширенная односторонняя ∞ – норма окрестности определяется следующим образом:

$$N_\infty(\gamma, \beta) = \{ (x, v, y) \in N_\infty(\gamma), s > 0 \mid \|g(x) + s\| \leq \beta \mu, \|f(x) - A^T v + \nabla g(x)y\| \leq \beta \mu \}.$$

[К содержанию](#)

5.2.3.4. Разновидности МВТ

Существует устоявшаяся классификация алгоритмов МВТ. В данном подразделе мы коротко охарактеризованы наиболее значимые разновидности МВТ, а именно:

- методы потенциальной редукции и методы следования центральному пути,
- прямые и двойственные методы, прямо-двойственные методы,
- методы допустимых и недопустимых областей.

Предварим описание методов каждой из перечисленных разновидностей замечанием о том, что по сути дела используемая классификация является классификацией МВТ по их отдельным свойствам. Это означает, что один и тот же метод может принадлежать разным разделам классификации. В частности рас-

⁴ Здесь и далее решения, на которых строго выполняются условия дополняющей нежесткости.

смаатриваемый ниже численный пример иллюстрирует работу алгоритма, который можно охарактеризовать как прямо-двойственный метод следования центральному пути в недопустимой области.

Методы потенциальной редукции. Методы следования центральному пути

Исторически методы потенциальной редукции появились первыми. Таковым являлся, например, алгоритм Кармаркара [2]. Он был основан на использовании следующей логарифмической потенциальной функции.

$$\Phi(x) = \rho \log(c^T x - Z) - \sum \log x_i,$$

где $\rho = n + 1$ и Z – нижняя оценка оптимального значения целевой функции.

Методы следования центральному пути появились позднее и также нашли широкое применение. В обоих случаях генерируются ньютоновские направления. Основное различие между ними состоит в том, что размер шага в выбранном направлении определяется, опираясь на две различные стратегии. В методах потенциальной редукции длина шага регулируется с помощью функций, называемых потенциальными; в то время как в методах следования центральному пути длина шага выбирается таким образом, чтобы вновь получаемая точка оставалась в заданной окрестности центрального пути.

Методы потенциальной редукции

Как отмечалось выше, целью методов потенциальной редукции является уменьшение некоторого потенциала на максимально возможную величину на каждом шаге. Наилучшим применением методов потенциальной редукции явилось решение задач ЛП. Коротко опишем алгоритм, для которого получена наилучшая на сегодняшний день оценка сходимости МВТ $O(\sqrt{n} |\log \varepsilon|)$ итера-

⁵ Применяется для так называемых МВТ недопустимых областей – весьма эффективных алгоритмов МВТ.

ций метода Ньютона до достижения величины двойственной невязки $x^T s$ в задаче (5.1) меньшей, чем наперед заданное, малое значение ε . В вычислительной схеме алгоритма использована следующая потенциальная функция.

$$\Phi(x, y, s) = \rho \log(c^T x - b^T y) - \sum \log x_i s_i,$$

где $\rho > n$ - постоянный параметр.

Алгоритм потенциальной редукции генерирует последовательность допустимых решений (x^k, y^k, s^k) , где x^k и s^k положительные, такие что $\Phi(x^k, y^k, s^k)$ сходится к $-\infty$. Можно показать, что $\Phi(x^k, y^k, s^k)$ сходится к $-\infty$ только тогда, когда слагаемое $\rho \log(c^T x)$ приближается к $-\infty$, а, следовательно, $x^T s$ приближается к нулю, что свидетельствует о приближении к оптимуму.

Направление перемещения вычисляется для установленного значения параметра μ^k эквивалентного $\frac{n}{\rho} \frac{x^T s}{n}$ в (5.8). Для этого находится решение соответствующей ньютоновской системы. Длина шага α_k вдоль ньютоновского направления определяется путем минимизации Φ до сохранения его положительности на следующей итерации. Такой шаг уменьшает Φ на фиксированную величину на каждой итерации, вследствие чего Φ сходится к $-\infty$. Кроме того, можно показать, что генерируемая последовательность (x^k, y^k, s^k) ограничена и сходится к точно дополняющему решению⁶, которое и является оптимальным решением. При этом выполняется неравенство $x^* + s^* > 0$.

Методы следования центральному пути

В методах следования центральному пути направления движения определяются также на основе решения параметризованной ньютоновской системы нелинейных уравнений (5.8).

Отличие от методов потенциальной редукции состоит в том, что длина

⁶ Решение на котором выполняются условия дополняющей нежесткости.

шага α_k выбирается такой, чтобы на следующей итерации найденная точка оставалась в окрестности центрального пути (используются меры $N_2(\theta)$ и $N_\infty(\gamma)$). В действительности в этом нет необходимости и зачастую вычислительные усилия в этом направлении нельзя признать рациональными.

Алгоритмы следования центральному пути обычно содержат два вида шагов: предиктор (предсказатель), цель которого приблизиться к оптимальному решению, уменьшая двойственную невязку, и корректор, целью которого является возвращение в окрестность центрального пути.

Прямые (двойственные) и прямо-двойственные методы

Вычислительные эксперименты показали, что прямо-двойственные методы значительно лучше, как по устойчивости, так и по быстродействию.

Прямые методы

Вернемся к задаче ЛП (5.1) и соответствующей барьерной задаче (5.2) с параметром μ . Предположим, что ОДР задачи не пуста. Применив теорему (5.2.3.2), определим, что (5.2) имеет единственное решение x^* . Из этого с необходимостью следует существование y^* и s^* , таких, что:

$$A x^* = b, \quad (5.13a)$$

$$A^T y^* + s^* = c, \quad (5.13b)$$

$$X^* s^* = \mu e, \quad (5.13c)$$

$$x^* > 0, \quad (5.13d)$$

$$s^* > 0, \quad (5.13e)$$

Прямые методы состоят в применении метода Ньютона к (5.2). Обычно считается, что целевая функция аппроксимируется первыми двумя членами ее разложения в ряд Тейлора на текущей итерации в точке x . Поэтому задача (5.2) может быть решена только приближенно и требуется сделать несколько ньюто-

новских шагов до выполнения критерия останова, такого как $\|Xs - \mu e\| < \theta$ для допустимой пары (x, s) , где θ - малый скаляр.

Для более точного описания предположим, что x – допустимое решение и пусть мы рассматриваем следующую квадратичную аппроксимацию:

$$c^T(x + \Delta x) - \mu \sum \log(x_i + \Delta x_i) \approx c^T x - \mu \sum \log x_i + c^T \Delta x - \mu x^{-T} \Delta x + \frac{\mu}{2} \Delta x^T X^{-2} \Delta x$$
, где x^{-1} – обозначение вектора – столбца с компонентами $1/x_i$ и X обозначает диагональную матрицу с компонентами x_i .

Ньютоновская задача оптимизации формулируется для данного μ , является задачей выпуклого программирования и имеет следующую структуру:

$$\begin{aligned} \text{найти} \quad & \min \frac{1}{2} \Delta x^T X^{-2} \Delta x + \left(\frac{c}{\mu} - x^{-1} \right) \Delta x \\ \text{при условиях:} \quad & A \Delta x = 0, \\ & \Delta x \geq -x \end{aligned} \tag{5.14}$$

Если не брать в расчет последнее ограничивающее неравенство, то решением этой системы Δx_N будет:

$$\Delta x_N = X^2 \left(x^{-1} - \frac{c}{\mu} + A^T (A X^2 A^T)^{-1} A X^2 \left(x^{-1} - \frac{c}{\mu} \right) \right). \tag{5.15}$$

Выразим также Δx_N следующим образом:

$$X^{-1} \Delta x_N = e - \frac{Xs(x)}{\mu},$$

где $s(x) = c - A^T y(x)$ и $y(x) = -(A X^2 A^T)^{-1} A X (Xc - \mu e)$.

Легко видеть, что данным x , $s(x)$, $y(x)$ соответствует следующая задача с переменными s и y :

$$\begin{aligned} \text{найти} \quad & \min_{s, y} \|Xs - \mu e\|_2 \\ \text{при условиях} \quad & s = c - A^T y \end{aligned} \tag{5.16}$$

Другими словами в прямом методе ньютоновский шаг определяет новые

переменные прямой задачи $x^+ = x + \Delta x_N$ и пары двойственных переменных (y, s) , которые находятся в окрестности возможного аналитического μ – центра при условии, что норма $\|Xs - \mu e\|_2$ минимальна. Однако только условия (5.13a) и (5.13b) выполняются для (x^+, s, y) . Априори ни x^+ , ни $s(x)$ не являются положительными. При этом легко видеть, что если $\|X^{-1}\Delta x_N\|_2 < 1$, то (5.13d) и (5.13e) выполняются для x^+ и $s(x)$ и, таким образом, полный ньютоновский шаг может быть выполнен. Более подробно, если евклидова норма $\|X^{-1}\Delta x_N\|_2$, обозначаемая как $\lambda(x)$. эквивалентна нулю, то условия (5.13c) выполняются непосредственно, (5.13d) выполняются потому, что x предполагается допустимым и, наконец, (5.13e) выполняются потому, что $s_i = \mu / x_i$ для всех i .

С точки зрения рассматриваемой нами классификации важно отметить, что прямой алгоритм может быть интерпретирован в случае ЛП как прямо-двойственный алгоритм, т.к. он вычисляет как переменные прямой задачи x , так и двойственные переменные $(s(x), y(x))$. Прямым он является в том смысле, что шаг метода Ньютона использует только данные прямой задачи.

В случае НЛП развивается этот же подход. Для задачи выпуклого программирования (5.12) прямые алгоритмы рассматривают параметризованную барьерную задачу:

$$\begin{aligned} \text{найти} \quad & \min \quad f(x) - \mu \sum \log(-g_i(x)), \\ \text{при ограничениях} \quad & Ax = b \\ & x \in R^n. \end{aligned} \tag{5.17}$$

Такой барьер определяет только допустимые точки и, следовательно, прямые алгоритмы предполагают, что существует по крайней мере одна точка $\bar{x}$, такая что $g(\bar{x}) < 0$.

Далее строится квадратичная модель нелинейной целевой функции, при

этом используются первые члены ее разложения в ряд Тейлора. Задача (5.17) преобразуется к следующему виду:

Найти

$$\min_{\Delta x} \frac{1}{2} \Delta x^T [\nabla^2 f(x) + \mu \sum (\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)})] \Delta x + \nabla f(x) - \mu \nabla g(x) G(x)^{-1} e)^T \Delta x$$

при ограничениях

$$A \Delta x = b. \quad (5.17)$$

Для того, чтобы остаться внутри допустимой области, определяемой ограничениями – неравенствами, полный ньютоновский шаг может быть ослаблен. Далее процедура повторяется с уменьшением значения μ до тех пор, пока не будет получено приближенное решение задачи НЛП.

Прямо-двойственные методы

Целью прямо-двойственных алгоритмов является одновременное решение пары двойственных задач. Они основываются на применении метода Ньютона для поиска решений системы Куна – Таккера (5.12), в которых условия дополняющей нежесткости преобразованы с использованием положительного параметра μ :

$$\begin{aligned} Ax_1 &= b, \\ g(x_1) + x_2 &= 0, \\ \nabla f(x_1) + A^T y_1 + \frac{\partial g}{\partial x}(x_1)^T y_2 &= 0, \\ Y_2 x_2 &= \mu e, \\ x_2 &\geq 0. \end{aligned} \quad (5.19)$$

Решение этой системы лежит на центральном пути. Часто его также называют центральной мишенью. Приближенное решение задачи НЛП основано на применении метода Ньютона в один или несколько этапов, чему соответствует решение одной или нескольких систем уравнений следующего вида:

$$\begin{bmatrix} A & 0 & 0 & 0 \\ \frac{\partial g}{\partial x} & I & 0 & 0 \\ \nabla^2 f(x_1^k) + \sum y_{2j}^k \nabla^2 g_j(x_1^k) & 0 & A^T & \frac{\partial g}{\partial x}^T \\ 0 & Y_2^k & 0 & X_2^k \end{bmatrix} \begin{bmatrix} \Delta x_1^{k+1} \\ \Delta x_2^{k+1} \\ \Delta y_1^{k+1} \\ \Delta y_2^{k+1} \end{bmatrix} = \begin{bmatrix} -(Ax_1^k - b) \\ -(g(x_1^k) + x_2^k) \\ -(\nabla f(x_1^k) + A^T y_1^k + \frac{\partial g}{\partial x}(x_1^k) y_2^k) \\ -(Y_2^k x_2^k - \mu e) \end{bmatrix} \quad (5.20)$$

На разных шагах алгоритма используются разные значения правых частей системы уравнений, которые обозначим $r_{p1}, r_{p2}, r_{d1}, r_c$ и которые соответствуют прямым, двойственным остаткам и невязке условия центральности соответственно. При этом систему (5.20) можно переписать следующим образом:

$$\begin{bmatrix} A & 0 & 0 & 0 \\ \frac{\partial g}{\partial x} & I & 0 & 0 \\ \nabla^2 f(x_1^k) + \sum y_{2j}^k \nabla^2 g_j(x_1^k) & 0 & A^T & \frac{\partial g}{\partial x}^T \\ 0 & Y_2^k & 0 & X_2^k \end{bmatrix} \begin{bmatrix} \Delta x_1^{k+1} \\ \Delta x_2^{k+1} \\ \Delta y_1^{k+1} \\ \Delta y_2^{k+1} \end{bmatrix} = \begin{bmatrix} r_{p1} \\ r_{p2} \\ r_{d1} \\ r_c \end{bmatrix} \quad (5.21)$$

Ньютоновские шаги ослабляют так, чтобы новые точки оставались внутри ОДР и по возможности так, чтобы используемая функция качества достаточно уменьшалась.

Как отмечалось выше, обычно применяются два вида шагов:

- шаг предиктор (предсказатель), соответствующий решению системы (5.20), параметр μ при этом может быть принят равным нулю;
- шаг корректор, соответствующий решению системы (5.21) при следующих значениях правых частей: $r_{p1}=0, \quad r_{p2}=0, \quad r_{d1}=0, \quad r_c=(\mu_c - \mu)e - \Delta_p X_2 \Delta_p y_2$, где $\Delta_p X_2$ и $\Delta_p y_2$ – компоненты решения, полученного на шаге предиктор и μ_c – положительный скаляр, выбранный соответствующим образом.

Эта схема была применена с большим практическим успехом Мехротрой

[9]. В настоящее время используется в большинстве реализаций МВТ.

Методы допустимых и недопустимых областей

Как уже упоминалось, существует два класса МВТ – методы допустимых и недопустимых областей. В методах допустимых областей начальная точка должна: принадлежать ОДР и быть найдена до начала работы алгоритма. Напротив, в методах недопустимых областей нарушение допустимости и величины невязок условий дополняющей нежесткости уменьшаются одновременно.

Методы допустимых областей

Вернемся к задаче ЛП в стандартной формулировке (5.1). Для МВТ допустимых областей предполагается существование допустимой внутренней точки, т.е. такой допустимой точки, которая находится внутри ОДР, определяемой ограничениями – неравенствами (ограничения выполняются как строгие неравенства). В случае прямо-двойственной системы данный метод означает возможность найти (x^0, y^0, s^0) , таких, что:

$$\begin{aligned} Ax^0 &= b, \\ A^T y^0 + s^0 &= c, \\ x^0 &> 0, \\ s^0 &> 0, \end{aligned} \tag{5.22}$$

с гарантией существования такой точки. На практике методы допустимых областей включают фазу 1, которая позволяет найти стартовую допустимую точку; фаза 2 состоит в приближении к оптимуму с помощью генерации допустимых точек. Мотивацией использования данного варианта МВТ является то, что если фаза 2 не сходится, пользователь может опереться на допустимую точку, полученную в первой фазе алгоритма.

Таким образом, стартовая точка (x^k, y^k, s^k) предполагается допустимой и во второй фазе применяется метод Ньютона для поиска решения системы:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mu e - S^k X^k e \end{bmatrix}. \quad (5.23)$$

Т.к. компоненты векторов x и s должны оставаться положительными, ньютоновский шаг ослабляется с помощью положительного параметра $\alpha \leq 1$:

$$x^{k+1} = x^k + \alpha \times \Delta x,$$

$$y^{k+1} = y^k + \alpha \times \Delta y,$$

$$s^{k+1} = s^k + \alpha \times \Delta s.$$

Легко видеть, что при наличии исходной допустимой точки все подпоследовательности генерируемых методом Ньютона решений также допустимы в том случае, когда все ограничения линейны. Действительно, из $Ax^k = b$ и $A\Delta x = 0$, $Ax^{k+1} = A(x^k + \alpha \times \Delta x) = Ax^k + \alpha \times \Delta x = b + \alpha \times 0 = b$ следует, что x^{k+1} допустим и, таким образом, выполняются ограничения $Ax=b$, при любой длине шага α . То же самое можно сказать относительно компонент y^{k+1} и s^{k+1} .

Недостатком МВТ допустимых областей является необходимость поиска стартовой внутренней точки ОДР. Таких точек может не существовать, а ОДР иметь пустую внутренность, хотя сама при этом будет непустой.

Методы недопустимых областей лишены этого недостатка⁷.

Методы недопустимых областей

Первый подход, в случае, когда приходится иметь дело с ОДР с пустой внутренностью, алгоритм МВТ (недопустимых областей) заключается в порождении последовательности недопустимых точек, приближающихся к ОДР извне.

Вернемся к задаче ЛП в стандартной формулировке и ассоциированной с ней системе линейных уравнений (5.8), где x и s принадлежат положительному

ортанту. Для старта с недопустимой точки методом Ньютона необходимо вначале решить следующую систему:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax^k \\ c - A^T y^k - s^k \\ \mu e - S^k X^k e \end{bmatrix}. \quad (5.24)$$

Замечательным фактом является то, что как бы ни нарушались линейные ограничения, полный ньютоновский шаг восстанавливает допустимость (находит точку, принадлежащую ОДР), тогда как ослабленный ньютоновский шаг уменьшает недопустимость (сокращает невязку) пропорционально величине $(1-\alpha)$. Более точно, имеем:

$$\begin{aligned} b - Ax^{k+1} &= (1-\alpha) \times (b - Ax^k), \\ c - A^T y^{k+1} - s^{k+1} &= (1-\alpha) \times (c - A^T y^k - s^k). \end{aligned}$$

В результате уменьшатся нарушение ограничений пропорционально длине шага.

Вернемся к задаче НЛП.

Рассмотрим нелинейную систему уравнений $g(x_1) + x_2 = 0$, где x_2 — неотрицательные ослабляющие переменные и положим, что $g(x_1^k) + x_2^k \neq 0$ на текущей итерации. В противоположность линейному случаю, уменьшение нарушения g может не иметь места, если ньютоновский шаг $(\Delta x_1, \Delta x_2)$ недостаточно ослаблен. Для определения величины ослабляющего фактора α члены разложения g по x_1^k в ряд Тейлора до второго порядка могут быть использованы для оценки нарушения ограничений, которые приблизительно равны $\frac{\alpha}{2} \Delta x_1^T \nabla^2 g \Delta x_1$.

Восстановлением допустимости нелинейных ограничений значительно сложнее управлять, чем соответствующей процедурой для линейного случая.

⁷ Следует заметить, что в этом случае само название метод внутренних точек теряет первоначальный смысл и

Второй подход, применяемый в МВТ недопустимых областей, заключается в использовании принципа инвариантного погружения, суть которого в данном случае состоит в том, что исходная задача переформулируется так, чтобы:

- стартовая внутренняя точка всегда была известна;
- решение могло быть найдено с помощью МВТ допустимых областей.

Инвариантное погружение является основным подходом, применяемым для обобщения проблемы в целях ее изучения и часто помогающим в поиске решений. Данный подход для МВТ в линейном случае базируется на применении кососимметрической самодвойственной модели, которая была изучена еще в 60-х годах прошлого века Голдманом и Таккером [10].

В случае ЛП для любой положительной пары (x^0, s^0) , введем новые структурные переменные θ и τ , которые приводят к расширенной задаче, называемой однородной самодвойственной (ОСД):

$$\text{найти} \quad \min_{(x, s, \lambda, \theta)} ((x^0)^T s^0 + 1)\theta$$

при ограничениях:

$$\begin{aligned} -c^T x + b^T \lambda + \bar{z} \theta &\geq 0, \\ c\tau - A^T \lambda - \bar{c} \theta &\geq 0, \\ -b\tau + Ax + \bar{b} \theta &= 0, \\ \bar{z}\tau + \bar{c}^T x - \bar{b}^T \lambda &= -((x^0)^T s^0 + 1), \\ (\tau, x) &\geq 0, \\ (\lambda, \theta) &\text{ свободные.} \end{aligned} \tag{5.25}$$

где $\bar{b} = b - Ax^0$, $\bar{c} = c - A^T \lambda^0 - s^0$, $\bar{z} = c^T x^0 + 1 - b^T \lambda^0$. Достаточно просто увидеть, что для всех x^0, s^0, λ^0 , точка $(x^0, s^0, \lambda^0, 1)$ является допустимой для данной ОСД формулировки. Из этого следует, что эквивалентная монотонная смешанная линейная дополняющая задача (МСЛД) к данной самодвойственной

задаче, получаемая введением дополнительных переменных для ограничений – неравенств имеет оптимальное решение. Оптимальное решение может быть найдено с помощью стандартного МВТ допустимых областей, стартовой точкой для которого является определенное выше допустимое решение.

Можно показать, что $\theta = 0$ для любого оптимального решения (5.25) и, следовательно:

$$\begin{array}{rcll} -c^T x & + b^T \lambda & \geq & 0, \\ c\tau & & - A^T \lambda & \geq 0, \\ -b\tau & + Ax & = & 0, \\ (\tau, x) & & \geq & 0, \\ \lambda & & & \text{свободные.} \end{array} \quad (5.26)$$

Если $\tau > 0$, то пара $(\frac{x}{\tau}, \frac{\lambda}{\tau})$ является прямо-двойственным оптимальным решением, т.к. является прямо и двойственно допустимой с нулевой двойственной невязкой. Обратное тоже верно: если ОДР задачи (5.1) непуста, то $\tau > 0$. Если же $\tau = 0$, это свидетельствует о несовместности ОДР прямой или двойственной задачи (или и той и другой одновременно).

Для НЛП, используя рассматриваемый подход, введем новые структурные переменные τ и k . Ориентируясь на приведенную ниже задачу НЛП в стандартной формулировке:

$$\begin{array}{ll} \text{Найти} & \min_x f(x) \\ \text{при ограничениях} & g(x) \geq 0, \end{array} \quad (5.27)$$

Используем следующую однородную коническую формулировку:

$$\text{Найти} \quad \min \quad z^T y + \tau k$$

при ограничениях:

$$\begin{aligned}
\tau \nabla f\left(\frac{x}{\tau}\right) + \frac{\partial g^T}{\partial x}\left(\frac{x}{\tau}\right)y &= 0, \\
g\left(\frac{x}{\tau}\right) &= \frac{z}{\tau}, \\
-x^T \nabla f\left(\frac{x}{\tau}\right) - x^T \frac{\partial g^T}{\partial x}\left(\frac{x}{\tau}\right)\frac{y}{\tau} - y^T g\left(\frac{x}{\tau}\right) &= k, \\
x, \tau, y, k, z &\geq 0.
\end{aligned} \tag{5.28}$$

В этой формулировке τ и k положительные скалярные переменные. Доказано [12], что алгоритм, базирующийся на приведенных соотношениях, имеет возможность распознавать несовместность ОДР задачи НЛП.

[К содержанию](#)

5.2.3.5. Глобально сходящийся алгоритма на основе барьера Фриша

При рассмотрении алгоритма делается нестрогое предположение о гладкости и регулярности функций, определяющих задачу, что дает возможность определения глобального оптимума и является необходимым условием для обеспечения глобальной сходимости алгоритма [13]. При условии, что стартовая точка определена, алгоритм порождает последовательность точек, сходящихся к стационарному решению задачи.

Рассматривается следующая задача:

$$\begin{aligned}
&\text{найти} && \min_x f(x) \\
&&& g(x) + s = 0, \\
&\text{при ограничениях} && x \in R^n, \\
&&& s \geq 0.
\end{aligned} \tag{5.29}$$

Функции g и f выпуклы, дифференцируемы и удовлетворяют допущениям теоремы (5.2.3.1). В соответствии с этими предположениями определена следующая μ – параметризованная барьерная задача:

$$\begin{aligned}
&\text{найти} && \min_x f(x) - \mu \sum \ln s_i \\
&&& g(x) + s = 0, \\
&\text{при ограничениях} && x \in R^n.
\end{aligned} \tag{5.30}$$

(5.30) имеет единственное решение для каждого $\mu > 0$. Условия Куна – Таккера для нее запишутся следующим образом:

$$\begin{aligned}
g(x) + s &= 0, \\
\nabla f(x) + \frac{\partial g^T}{\partial x}(x)y &= 0, \\
YSe - \mu e &= 0.
\end{aligned} \tag{5.31}$$

Стартовая точка алгоритма $z=(x,y,s)$, $y>0$, $s>0$ является прямо-двойственным внутренним решением. Последующие решения порождаются методом Ньютона применительно к следующей прямо-двойственной системе уравнений:

$$F(z) = \begin{pmatrix} g(x) + s \\ \nabla f(x) + \frac{\partial g^T}{\partial x}(x)y \\ YSe - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Направление движения dz определяются с помощью решения следующей линейной системы уравнений:

$$\nabla F(z)dz + F(z) = 0 \tag{5.32}$$

В алгоритме Виала параметр μ выбирается адаптивно на каждой итерации:

- если $\min_i (y_i s_i) > \delta \frac{y^T s}{m}$, где δ обычно выбирается равным 1.e-5, то ис-

пользуется нормальная величина параметра $\mu = \frac{1}{m^2} \frac{y^T s}{m}$;

- в противном случае используется значение $\mu = \frac{y^T s}{m}$.

Особенностью алгоритма Виала является необходимость выполнения множества итераций до получения приближенного для данного μ решения. Когда такое решение найдено, μ уменьшается и процесс повторяется. Алгоритм глобально сходится независимо от стартовой точки в предположении, что приближенное решение для данного μ существует.

Для данного μ длина шага может выбираться на каждой итерации. Виал ввел функцию качества P_μ и требование к алгоритму, чтобы величина шага α удовлетворяла условию Голдштейна-Армийо в следующей форме:

$$\lambda_1 \alpha \left| \nabla P_\mu(z) dz \right| \leq P_\mu(z) - P_\mu(z + \alpha dz) \leq \lambda_2 \alpha \left| \nabla P_\mu(z) dz \right|,$$

где λ_1 и λ_2 - константы, такие что $0 < \lambda_1 < \lambda_2 < 1$. Для данного $\mu > 0$, это показывает, что для любой предельной точки $\bar{z}$ любой последовательности z^k , где α^k выбирается в соответствии с приведенным ниже правилом:

$$\nabla P_\mu(z) dz = 0.$$

Правило Голдштейна-Армийо может выполняться при условии, что направление dz уменьшает P_μ . Функция качества имеет следующий вид:

$$P_\mu(z) = y^T s + \frac{1}{2} \|g(x) + s\|^2 + \frac{1}{2} \left\| \frac{\partial g^T}{\partial x} y + \nabla f \right\|^2 - \mu \sum \log y_i s_i.$$

Принимая во внимание ньютоновские направления dz из (5.32), элементарные алгебраические преобразования показывают, что

$$\nabla P_\mu(z) dz = -\|g(x) + s\|^2 - \left\| \frac{\partial g^T}{\partial x} y + \nabla f \right\|^2 - \|(YS)^{1/2} - \mu(YS)^{-1/2}\|^2.$$

Т.к. все члены выражения отрицательны, то ньютоновское направление, соответствующее (5.31) уменьшает значение P_μ .

Кроме того, каждый член выражения является нормой вектора, определенного условиями Куна – Таккера (5.30). В результате любая предельная точка $\bar{z}$ любой последовательности z^k , порождаемой описанной выше процедурой, является решением системы (5.30)⁸.

Описание теоретической схемы данного алгоритма можно дополнить еще двумя правилами, применяемыми на практике:

- использованием различной длины шага для пространства решений прямой и двойственной задач; при этом применяется эвристическая процедура линейного поиска.
- В алгоритмах, основанных на самосогласующихся барьерах, длина шага детально определяется на каждой итерации. Уменьшение функции качества может быть ограничено снизу положительным числом, следовательно, появляется возможность вычислить максимальное число итераций, необходимых для поиска оптимума.

Оценки трудоемкости алгоритмов⁹

Существует множество модификаций приведенной вычислительной схемы, зависящих от параметров. Опишем для сравнения две наиболее простые.

1. Прямо-двойственный алгоритм с полным ньютоновским шагом или алгоритм с коротким шагом:

- Используется мера $\delta(\bar{x}\bar{s}, \mu) := \delta_0(\bar{x}\bar{s}, \mu)$;
- $\mu^0 := 1$;
- $\theta := \frac{1}{2\sqrt{n}}$;
- $k = \sqrt{2}$;

⁸ Здесь опущены некоторые детали, которые вытекают из того, что последовательность z^k не обязательно сходится.

⁹ Оценки приводятся для задач ЛП, для задач выпуклого программирования также получены полиномиальные оценки..

- $\alpha = 1$.

Для данной модификации получена наилучшая теоретическая оценка трудоемкости решения задач ЛП: $O(\sqrt{n} \log \frac{n}{\varepsilon})$.

2. Прямо-двойственный алгоритм с длинным шагом:

- Используется мера $\delta(\bar{x}\bar{s}, \mu) := \delta_0(\bar{x}\bar{s}, \mu)$;
- $\mu^0 := 1$;
- $0 < \theta < \frac{n}{n + \sqrt{n}}$;
- $k = \frac{\sqrt{R}}{2\sqrt{1 + \sqrt{R}}}$, где $R = \frac{\theta\sqrt{n}}{1 - \theta}$;
- α вычисляется в соответствии с (5.37).

Для данной модификации доказанная теоретическая оценка трудоемкости (количество шагов до получения допустимого решения, такого, что

$$\delta_0(\bar{x}\bar{s}, \mu) \leq k \text{ и } (n+1)\mu \leq \varepsilon): \left[\frac{1}{\theta} \left[2 \left(1 + \sqrt{\frac{\theta\sqrt{n}}{1 - \theta}} \right)^4 \right] \log \frac{n}{\varepsilon} \right].$$

При $\theta := \frac{1}{2}$ получается оценка трудоемкости $O(n \log \frac{n}{\varepsilon})$, для $\theta := \frac{C}{\sqrt{n}}$, при любом

фиксированном положительном C , оценка трудоемкости $O(\sqrt{n} \log \frac{n}{\varepsilon})$ [18].

Существует серьезное расхождение между теоретическими и практическими результатами. Так наилучшая теоретическая оценка трудоемкости методов с длинным шагом [14] равна $O(\sqrt{n}(\log n) \log \frac{n}{\varepsilon})$, что хуже оценки методов с коротким шагом, практически же методы с длинным шагом сходятся значи-

тельно быстрее (имеется статистическая оценка $O(\log n \log \frac{n}{\varepsilon})$, где $O > 1$ – не-
большая константа).

[К содержанию](#)

5.2.3.6. Пример применения метода внутренних точек

Поскольку все современные алгоритмы МВТ не отличаются наглядностью и не предназначены для ручного счета, ограничимся иллюстрацией одного из рассмотренных выше алгоритмов на решении задачи ЛП предельно малой размерности.

Пример 5.2.6. Рассмотрим задачу ЛП в стандартной форме записи:

$$\begin{array}{ll} \text{найти:} & \max \quad 3x_1 + 2,5x_2 \\ & 2x_1 + 0,5x_2 \leq 3 \\ \text{при ограничениях} & x_1 + x_2 \leq 2 \\ & x_1 \geq 0, \quad x_2 \geq 0 \end{array}$$

Назовем ее задачей Р.

Соответствующая двойственная задача запишется следующим образом:

$$\begin{array}{ll} \text{найти:} & \min \quad 3u_1 + 2u_2 \\ & 2u_1 + u_2 \geq 3 \\ \text{при ограничениях} & 0,5u_1 + u_2 \geq 2,5 \\ & u_1 \geq 0, \quad u_2 \geq 0 \end{array}$$

Обозначим данную задачу, как задачу D.

Для поиска решений пары задач Р и D воспользуемся рассмотренной выше (раздел 5.2.3.5) эффективной разновидностью МВТ прямо-двойственным методом следования центральному пути в недопустимой области. Для этого применим модель Голдмана-Таккера, составив ОСД – задачу (5.26). Введем необходимые структурные переменные и ослабляющие переменные для ограничений

пары задач Р и D:

Задача Р

$$\begin{array}{rclcl} 2x_1 & + & 0,5x_2 & + & y_1 & = & 3 \\ x_1 & & + & x_2 & & + & y_2 & = & 2 \end{array}$$

$$\Downarrow$$

$$\begin{array}{rclcl} 2x_1 & + & 0,5x_2 & - & 3\tau & + & y_1 & = & 0 \\ x_1 & & + & x_2 & - & 2\tau & & + & y_2 & = & 0 \end{array}$$

$\Downarrow$

$$\begin{array}{rclcl} -2x_1 & - & 0,5x_2 & + & 3\tau & - & y_1 & = & 0 \\ -x_1 & & - & x_2 & + & 2\tau & & - & y_2 & = & 0 \end{array}$$

$$3x_1 + 2,5x_2 - 3u_1 - 2u_2 - \rho$$

$$\Downarrow$$

ОСД – задача:

$$\begin{array}{rclcl} 2u_1 & + & u_2 & - & 3\tau & - & v_1 & = & 0 \\ 0,5u_1 & + & u_2 & - & 2,5\tau & & - & v_2 & = & 0 \\ -2x_1 & - & 0,5x_2 & & + & 3\tau & & - & y_1 & = & 0 \\ -x_1 & & - & x_2 & & + & 2\tau & & - & y_2 & = & 0 \\ 3x_1 & + & 2,5x_2 & - & 3u_1 & - & 2u_2 & - & \rho & = & 0 \end{array} \quad (5.33)$$

Для удобства запишем (5.33) в матричном виде. Определим матрицу системы M и векторы x и $s(x)$:

$$M = \begin{pmatrix} 0 & 0 & 2 & 1 & -3 \\ 0 & 0 & 0,5 & 1 & -2,5 \\ -2 & -0,5 & 0 & 0 & 3 \\ -1 & -1 & 0 & 0 & 2 \\ 3 & 2,5 & -3 & -2 & 0 \end{pmatrix}; \quad x = \begin{pmatrix} x_1 \\ x_2 \\ u_1 \\ u_2 \\ \tau \end{pmatrix}; \quad s(x) = \begin{pmatrix} v_1 \\ v_2 \\ y_1 \\ y_2 \\ \rho \end{pmatrix}$$

В соответствии с (5.26) можно записать:

$$Mx \geq 0; \quad x \geq 0; \quad s(x) = Mx$$

Определим вектор $r=e-Me$, матрицу $\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}$, и число $\lambda=n+1$. Тогда:

$$r=\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}-\begin{pmatrix} 0 & 0 & 2 & 1 & -3 \\ 0 & 0 & 0,5 & 1 & -2,5 \\ -2 & -0,5 & 0 & 0 & 3 \\ -1 & -1 & 0 & 0 & 2 \\ 3 & 2,5 & -3 & -2 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}=\begin{pmatrix} 1 \\ 2 \\ 0,5 \\ 1 \\ 0,5 \end{pmatrix}, \lambda=6.$$

Дополним переменные задачи парой ϑ и ν . Система, определяющая центральный путь, запишется следующим образом:

$$-\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \vartheta \end{pmatrix}+\begin{pmatrix} s \\ \nu \end{pmatrix}=\begin{pmatrix} 0 \\ \lambda \end{pmatrix}; \begin{pmatrix} x \\ \vartheta \end{pmatrix}, \begin{pmatrix} s \\ \nu \end{pmatrix} \geq 0, \quad (5.34)$$

$$\begin{pmatrix} x \\ \vartheta \end{pmatrix}\begin{pmatrix} s \\ \nu \end{pmatrix}=\begin{pmatrix} \mu e \\ \mu \end{pmatrix}.$$

В качестве стартовой точки можно выбрать единичный вектор $(x^0, \vartheta^0, s^0, \nu^0)=(e, 1, e, 1)$, который всегда является допустимым решением (5.34), т.е. принадлежит центральному пути. Выберем также начальное значение параметра $\mu=1$.

Шаг метода Ньютона состоит в поиске приращений $(\Delta x, \Delta \vartheta, \Delta s, \Delta \nu)$ после редукции параметра μ , таких, что:

$$-\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}\begin{pmatrix} x + \Delta x \\ \vartheta + \Delta \vartheta \end{pmatrix}+\begin{pmatrix} s + \Delta s \\ \nu + \Delta \nu \end{pmatrix}=\begin{pmatrix} 0 \\ \lambda \end{pmatrix}; \begin{pmatrix} x + \Delta x \\ \vartheta + \Delta \vartheta \end{pmatrix}, \begin{pmatrix} s + \Delta s \\ \nu + \Delta \nu \end{pmatrix} \geq 0, \quad (5.35)$$

$$\begin{pmatrix} x + \Delta x \\ \vartheta + \Delta \vartheta \end{pmatrix}\begin{pmatrix} s + \Delta s \\ \nu + \Delta \nu \end{pmatrix}=\begin{pmatrix} \mu e \\ \mu \end{pmatrix}.$$

Пренебрегая квадратичными слагаемыми $\Delta x \Delta s$ и $\Delta \vartheta \Delta \nu$ системы (5.35), получаем ее линейную аппроксимацию:

$$\begin{array}{rclcl}
-M\Delta x & -r\Delta g & +\Delta s & = & 0 \\
r^T\Delta x & & +\Delta v & = & 0 \\
s\Delta x & +x\Delta s & & = & \mu e - xs \\
v\Delta g & +g\Delta v & & = & \mu - gv
\end{array} \tag{5.36}$$

Размерность системы (5.36) легко уменьшить, произведя элементарные преобразования. Для этого используем нижеследующие наблюдения.

Для любого вектора (x, g, s, v) , удовлетворяющего системе (5.34), справедливо следующее соотношение:

$$x^T s + gv = g\lambda.$$

Новое решение после шага метода Ньютона, следовательно, удовлетворяет соотношению:

$$(x + \Delta x)^T (s + \Delta s) + (g + \Delta g)^T (v + \Delta v) = (g + \Delta g)\lambda.$$

Перегруппировав слагаемые в последнем выражении, будем иметь:

$$(x^T s + gv) + (\Delta x^T \Delta s + \Delta g \Delta v) + (x^T \Delta s + s^T \Delta x + g \Delta v + v \Delta g) = g\lambda + \Delta g\lambda$$

Как показано выше, первое слагаемое левой части уравнения эквивалентно $g\lambda$, из (5.36) выводим, что второе слагаемое равно нулю и из последнего уравнения (5.36) вытекает, что третье слагаемое эквивалентно $\mu(n+1) - x^T s - gv = \mu\lambda - g\lambda$. В результате получаем выражение $\mu\lambda - g\lambda = \Delta g\lambda$. Окончательно получаем выражение для Δg :

$$\Delta g = \mu - g. \tag{5.36a}$$

Это выражение подставленное в последнее уравнение (5.36) дает значение

$$\Delta v: \quad \Delta v = \frac{\mu}{g} - v - \frac{v(\mu - g)}{g}, \text{ или}$$

$$\Delta v = \frac{\mu(1-v)}{g}. \tag{5.36b}$$

Δs может быть получено из третьего уравнения (5.36), как

$$\Delta s = \mu X^{-1} e - s - X^{-1} S \Delta x. \tag{5.36c}$$

В конце концов, подставив все полученные значения в первое уравнение (5.36) будем иметь:

$$M\Delta x + X^{-1}S\Delta x = \mu X^{-1}e - s - (\mu - \vartheta)r,$$

т.е. Δx является единственным решением положительно определенной системы:

$$(M + X^{-1}S)\Delta x = \mu X^{-1}e - s - (\mu - \vartheta)r. \quad (3,36d)$$

После определения $(\Delta x, \Delta \vartheta, \Delta s, \Delta v)$, следует скорректировать текущее решение, используя при необходимости ослабляющий фактор α :

$$x := x + \alpha \Delta x$$

$$\vartheta := \vartheta + \alpha \Delta \vartheta$$

$$s := s + \alpha \Delta s$$

$$v := v + \alpha \Delta v$$

В качестве меры близости $\delta_0(\bar{x}\bar{s}, \mu)$ к центральному пути при использовании барьера Фриша рекомендуется следующее выражение:

$$\delta_0(\bar{x}\bar{s}, \mu) = \frac{1}{2} \left\| \left(\frac{\bar{x}\bar{s}}{\mu} \right)^{\frac{1}{2}} - \left(\frac{\mu}{\bar{x}\bar{s}} \right)^{\frac{1}{2}} \right\|_2,$$

где $\bar{x}, \bar{s}$ - векторы, составленные из x, ϑ и s, v соответственно, а $\bar{x}\bar{s}$ - их скалярное произведение.

Простейший способ определения величины ослабляющего фактора α :

$$\alpha = \min \left\{ 1, \min_j \left(\frac{-\bar{x}_j}{\Delta \bar{x}_j} : \Delta \bar{x}_j < 0 \right), \min_i \left(\frac{-\bar{s}_i}{\Delta \bar{s}_i} : \Delta \bar{s}_i < 0 \right) \right\} \quad (5.37)$$

С учетом введенных обозначений общая схема алгоритма выглядит следующим образом:

Input:

Предельная мера близости к центральному пути k ;

параметр точности решения $\varepsilon > 0$;

параметр величины шага θ , $0 < \theta < 1$;

$(\bar{x}^0, \bar{s}^0)$, $\mu^0 \leq 1$, такие, что $\delta_0(\bar{x}\bar{s}, \mu) \leq k$.

do

$\bar{x} := \bar{x}^0$; $\bar{s} := \bar{s}^0$; $\mu := \mu^0$.

while $(n+1)\mu \geq \varepsilon$ *do*

$\mu := (1 - \theta)\mu$;

while $\delta_0(\bar{x}\bar{s}, \mu) \geq k$ *do*

решение системы (5.36);

(5.38)

вычисление α (5.37);

$\bar{x} := \bar{x} + \alpha\Delta\bar{x}$;

$\bar{s} := \bar{s} + \alpha\Delta\bar{s}$;

end do

end do

end do

[К содержанию](#)

5.2.3.6. 1. Факторизация по Холецкому

Из всего многообразия методов решения систем линейных уравнений в МВТ используется их небольшая часть, имеющая наилучшую численную устойчивость. Чаще всего применяются различные модификации схемы Холецкого. Следует заметить, что в западных источниках под факторизацией по Холецкому подразумевается метод, называемый методом квадратного корня в отечественной литературе [15]. Поскольку метод квадратного корня применяется для симметричных положительно определенных систем уравнений, сведе-

ние к которым (5.36d) не всегда оправдано, рассмотрим более общий метод, который собственно в отечественной литературе называется схемой Холесского.

В схеме Холесского производится разложение (факторизация) исходной (произвольной) квадратной матрицы системы на два сомножителя:

$A=BC$, где B – нижнетреугольная, а C – верхнетреугольная матрицы вида

$$B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ & \dots & \dots & \\ b_{n1} & b_{n2} & & b_{nn} \end{pmatrix} \quad C = \begin{pmatrix} 1 & c_{12} & \dots & c_{1n} \\ 0 & 1 & \dots & c_{2n} \\ & \dots & \dots & \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

где элементы матриц B и C вычисляются в соответствии с рекуррентными соотношениями:

$$\begin{aligned} 1) \quad & b_{i1} = a_{i1}, \quad i = \overline{1, n}; \\ 2) \quad & c_{1j} = \frac{a_{1j}}{b_{11}}, \quad j = \overline{1, n}; \\ 3) \quad & b_{ij} = a_{ij} - \sum_{k=1}^{j-1} b_{ik} c_{kj} \quad i \geq j > 1; \\ 4) \quad & c_{ij} = \frac{1}{b_{ii}} (a_{ij} - \sum_{k=1}^{i-1} b_{ik} c_{kj}) \quad 1 < i < j; \end{aligned} \quad (5.39)$$

При решении уравнения $Ax=d$ факторизация $A=BC$ позволяет достаточно просто определить x . Дополним для этого соотношения (1-4):

$$\begin{aligned} 5) \quad & y_1 = \frac{a_{1,n+1}}{b_{11}}, \text{ где } a_{i,n+1} = d_i; \\ 6) \quad & y_i = \frac{1}{b_{ii}} (a_{i,n+1} - \sum_{k=1}^{i-1} b_{ik} y_k); \\ 7) \quad & x_n = y_n, \quad x_i = y_i - \sum_{k=i+1}^n c_{ik} x_k, \quad i < n. \end{aligned} \quad (5.40)$$

Если положить $A := (M + X^{-1}S)$, $x := \Delta x$, а $d := \mu X^{-1}e - s - (\mu - \vartheta)r$, то соотношения (1-7) являются алгоритмом решения системы (5.36d).

5.2.3.6.2. Численная иллюстрация шага алгоритма

Вернемся к нашему примеру. В соответствии с алгоритмом (5.38) и соотношениями (5.36a-5.36d) определяем: $\theta = 1/2$ $\mu := (1 - \theta) \mu = 1/2$;

$$\Delta \vartheta = \mu - \vartheta = 1/2 - 1 = -1/2; \quad \Delta v = \frac{\mu(1-v)}{\vartheta} = \frac{\frac{1}{2}(1-1)}{1} = 0;$$

$$\Delta s = \mu X^{-1}e - s - X^{-1}S\Delta x$$

$$\Delta s = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta x_5 \end{pmatrix} =$$

$$= \begin{pmatrix} -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \end{pmatrix} - \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta x_5 \end{pmatrix}$$

$$(M + X^{-1}S)\Delta x = \mu X^{-1}e - s - (\mu - \vartheta)r$$

$$(M + X^{-1}S)\Delta x = \begin{pmatrix} 1 & 0 & 2 & 1 & -3 \\ 0 & 1 & 0,5 & 1 & -2,5 \\ -2 & -0,5 & 1 & 0 & 3 \\ -1 & -1 & 0 & 1 & 2 \\ 3 & 2,5 & -3 & -2 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta x_5 \end{pmatrix} = A\Delta x;$$

$$\mu X^{-1}e - s - (\mu - \vartheta)r = \begin{pmatrix} -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 2 \\ 0,5 \\ 1 \\ 0,5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0,5 \\ -0,25 \\ 0 \\ -0,25 \end{pmatrix} = d.$$

Решаем систему (5.36d), используя схему Холесского (5.39)-(5.40) (для удобства все вычисления сведены в таблицу 5.3.

Факторизация Холесского

Таблица 5.3

Δx_1	Δx_2	Δx_3	Δx_4	Δx_5	d
1	0	2	1	-3	0
0	1	0,5	1	-2,5	0,5
-2	-0,5	1	0	3	-0,25
-1	-1	0	1	2	0
3	2,5	-3	-2	1	-0,25
1	0	2	1	-3	0
0	1	0,5	1	-2,5	0,5
-2	-0,5	5,25	0,4761	-0,8095	0
-1	-1	2,5	1,8095	-0,8158	0,2763
3	2,5	-10,25	-2,62	5,8152	-0,1335
					y
					Δx
					0
					-0,1923
					0,5
					0,0928
					0
					-0,1878
					0,2763
					-0,1674
					-0,1335
					-0,1335

Отсюда получаем разложение $A=BC$

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -2 & -0,5 & 5,25 & 0 & 0 \\ -1 & -1 & 2,5 & 1,8095 & 0 \\ 3 & 2,5 & -10,25 & -2,62 & 5,8152 \end{pmatrix}; C = \begin{pmatrix} 1 & 0 & 2 & 1 & -3 \\ 0 & 1 & 0,5 & 1 & -2,5 \\ 0 & 0 & 1 & 0,4761 & -0,8095 \\ 0 & 0 & 0 & 1 & -0,8158 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

и решение системы (5.36d)

$$\Delta x = \begin{pmatrix} -0,1923 \\ 0,0928 \\ -0,1878 \\ -0,1674 \\ -0,1335 \end{pmatrix}; \quad \Delta s = \begin{pmatrix} -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \\ -0,5 \end{pmatrix} - \begin{pmatrix} -0,1923 \\ 0,0928 \\ -0,1878 \\ -0,1674 \\ -0,1335 \end{pmatrix} = \begin{pmatrix} -0,3077 \\ -0,5928 \\ -0,3122 \\ -0,6674 \\ -0,3665 \end{pmatrix}; \text{ вычисляем } \alpha = 1 \text{ и согласно}$$

(5.37) и переходим к новой точке:

$$\vartheta = 1 - \frac{1}{2} = \frac{1}{2}; \quad v = 1 + 0 = 1; \quad x = \begin{pmatrix} 0,8077 \\ 1,0928 \\ 0,8122 \\ 1,1674 \\ 0,8665 \end{pmatrix}; \quad s = \begin{pmatrix} 0,6923 \\ 0,4072 \\ 0,6878 \\ 0,3326 \\ 0,6335 \end{pmatrix}.$$

Проверяем значение $\delta_0(\bar{x}, \mu) = 0,866$ и переходим к следующему шагу.

После 6 шагов получаем следующие результаты:

$$x = \begin{pmatrix} 1,1765 \\ 0,5882 \\ 0,2941 \\ 2,0588 \\ 0,8824 \end{pmatrix}; \quad s = \begin{pmatrix} 0,0012 \\ 0,0000 \\ 0,0008 \\ 0,0000 \\ 0,0000 \end{pmatrix}. \text{ При этом выполняются условия оптимальности. Окон-}$$

чательное решение получаем согласно (5.26) $\bar{x}^* := (\frac{\bar{x}}{\tau}, \tau)$, а т.к. $\tau = 0,8824$, то

$$\bar{x}^* = \begin{pmatrix} 1,3333 \\ 0,6667 \\ 0,3333 \\ 2,3333 \\ 1,0 \end{pmatrix}. \text{ Возвращаясь к исходным обозначениям, } x = \begin{pmatrix} x_1 \\ x_2 \\ u_1 \\ u_2 \\ \tau \end{pmatrix}; \quad s(x) = \begin{pmatrix} v_1 \\ v_2 \\ y_1 \\ y_2 \\ \rho \end{pmatrix}, \text{ по-}$$

лучаем оптимальные решения пары задач P и D:

$$x^* = \begin{pmatrix} 1,3333 \\ 0,6667 \end{pmatrix}; \quad u^* = \begin{pmatrix} 0,3333 \\ 2,3333 \end{pmatrix}. \text{ Выполнение условий оптимальности легко}$$

проверяется.

Графическая иллюстрация алгоритма

Рис. 5.3 иллюстрирует работу прямого длинношагового МВТ при решении задачи выпуклого программирования для двумерного случая. На рисунке центральный путь изображен кривой внутри ОДР, целевая функция обозначена касательной к ОДР в точке оптимума, шагам алгоритма соответствуют жирные точки, соединенные отрезками.

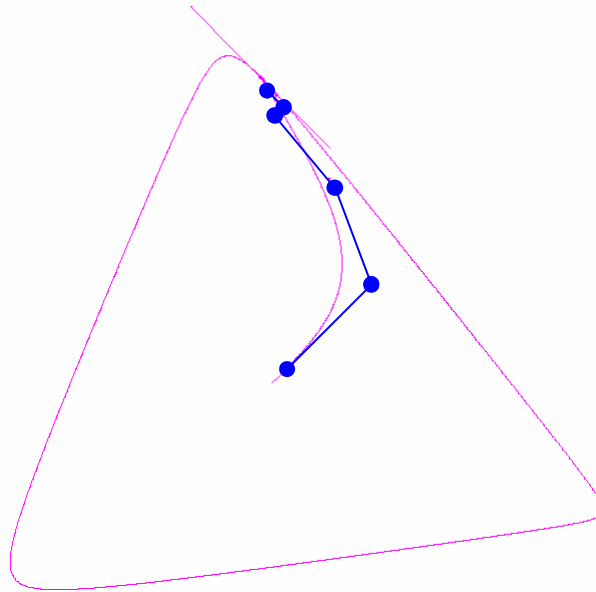


Рис. 5.3. Графическая иллюстрация прямого МВТ при решении задачи выпуклого программирования.

К содержанию

5.3. Упражнения

5.3.1. Решить методом множителей Лагранжа

Продажная цена акций А и В на начало предстоящего месяца составляет, соответственно, 26,60 и 21,55 рублей. В распоряжении инвестора имеется капи-

тал в размере 19000 р., который он решил использовать для вложений в эти ценные бумаги.

После проведения статистического анализа динамики курсов данных акций за предыдущий временной период получены оценки их средней эффективности $M1=0,4213$, $M2=0,2804$, а также оценки ковариаций: $V11=0,1267$, $V12=-0,0278$, $V22=0,0978$, отражающих взаимовлияние акций.

Какие из этих акций и в каком количестве следует приобрести по сегодняшнему курсу продаж, чтобы с минимальным риском получить в предстоящем месяце эффективность от такой структуры портфеля инвестиций 37,91% ?

Требуется:

1. Построить экономико-математическую модель расчета такой структуры инвестиций, которая обеспечит желательный уровень эффективности всего портфеля при минимуме риска не достичь этого уровня.

2. Решить полученную задачу квадратичного программирования, используя метод множителей Лагранжа. Определить количество акций А и В, которое должно находиться в минимальном по риску портфеле инвестиций.

Ответ:

M1	M2	V11	V12	V22	X1	X2	Zmin	N1	N2
0,42133	0,28044	0,1267	-0,0278	0,0978	0,7	0,3	0,05921	500	264

5.3.2. Решить методом Франка - Вулфа

Продажная цена акций А и В на начало предстоящего месяца составляет, соответственно, 37,69 и 6,69 рублей. В распоряжении инвестора имеется капитал в размере 16000 р., который он может использовать для вложений в эти ценные бумаги.

После проведения статистического анализа динамики курсов этих акций за предыдущий временной период получены оценки средней эффективности акций $M1=0,1707$, $M2=0,1667$, а также оценки ковариаций: $V11=0,1156$, $V12=0,0689$, $V22=0,1222$, отражающих взаимовлияние акций.

Какие из этих акций и в каком количестве следует приобрести по сегодняшнему курсу продаж, чтобы с минимальным риском получить в предстоящем месяце эффективность от такой структуры портфеля инвестиций не менее 15,20 % ?

Требуется:

1. Построить экономико-математическую модель расчета такой структуры инвестиций, которая обеспечит желательный уровень эффективности всего портфеля при минимуме риска не достичь этого уровня.

2. Решить полученную задачу квадратичного программирования, используя градиентный метод Франка-Вулфа с допуском по точности оптимального решения равным 0,01. Определить количество акций А и В, которое должно находиться в минимальном по риску портфеле инвестиций.

Ответ:

M1	M2	V11	V12	V22	X1	X2	Zmin	N1	N2
0,17067	0,16667	0,1156	0,0689	0,1222	0,5	0,4	0,07601	212	956

5.3.3. Решить задачу квадратичного программирования

Общие издержки фирмы на производство продукции в объеме Q единиц определяются зависимостью $TC(Q) = 3136 + 36*Q + 4*Q^2$ руб. Фирма может продать любое количество продукции по действующей рыночной цене 460 руб.

Требуется:

1. Определить объем производства Q^* , максимизирующий прибыль фирмы, найти значение максимума прибыли. Найти объем производства, минимизирующий средние издержки выпуска единицы продукции, указать значение минимально возможных издержек. Определить диапазон выпуска продукции, безубыточный для фирмы при действующей цене.

2. Представить аналитическую запись зависимости оптимального объема производства фирмы Q^* от динамики действующей рыночной цены при условии, что производственные мощности фирмы ограничены диапазоном $224 \leq Q \leq 448$. Построить график этой зависимости.

Ответ:

Q^*	ПРmax	Q^{**}	ACmin	Q1	Q2	$P < 946$	$\leq P < 1828$	$\leq P < 3620$	$\leq P$
53	8100	28	260	8	98	0	224	$0,125*(P - 36)$	448

5.3.4. Решить методом множителей Лагранжа

Предприятие производит некоторую продукцию двумя технологическими способами, причем издержки производства при первом способе изготовления x_1 тонн продукции равны $a_0 + a_1x_1 + a_2x_1^2$ руб., а при втором способе изготовления x_2 тонн продукции равны $b_0 + b_1x_2 + b_2x_2^2$ руб. Составить план производства, при котором будет произведено d тонн продукции при минимальных издержках. Решить методом множителей Лагранжа.

