МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный
университет им. Н. И. Лобачевского»

Ю. А. Кузнецов, А. В. Семенов

ОСНОВЫ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Учебно-методическое пособие

Рекомендовано методической комиссией
Института экономики и предпринимательства
для студентов ННГУ, обучающихся по направлению подготовки
38.03.05 «Бизнес-информатика», квалификация – бакалавр

Нижний Новгород
2022

УДК 519.853
ББК 22.18
К89

К-89 Кузнецов Ю. А., Семенов А. В. ОСНОВЫ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ: Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2022. – 48 с.

Рецензент: к.ф.-м.н., доцент **А. А. Тюхтина**

Учебно-методическое пособие посвящено теории условного экстремума функций многих вещественных переменных. Помимо изложения метода множителей Лагранжа и элементов теории двойственности в пособии приводится пример решения задач конечномерной оптимизации в системе MATLAB, а также задачи для самостоятельного решения.

Пособие предназначено для преподавателей ИЭП ННГУ, ведущих дисциплины «Методы оптимизации», «Методы оптимальных решений», «Исследование операций» и студентов, обучающихся по направлению подготовки 38.03.05 «Бизнес-информатика», бакалаврской образовательной программы «Аналитические методы и информационные технологии поддержки принятия решений в экономике и бизнесе».

Ответственный за выпуск:
председатель методической комиссии ИЭП ННГУ,
к.э.н., доцент **Макарова С. Д.**

УДК 519.852
ББК 22.18

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Математическое программирование – область математики, разрабатывающая теорию и численные методы решения многомерных экстремальных задач с ограничениями, т.е. задач на экстремум функции многих переменных с ограничениями на область изменения этих переменных. В отличие от классической теории экстремальных задач, которая является частью математического программирования, основное внимание уделяется тем задачам, в которых активно участвуют ограничения на область изменения переменных.

Своё название математическое программирование получило от входящего в него линейного программирования, разрабатывающего теорию и численные методы нахождения экстремума (максимума или минимума) линейной функции многих переменных при наличии линейных ограничений, т.е. линейных равенств или неравенств, связывающих эти переменные. К задачам линейного программирования приводится широкий круг вопросов планирования экономических и технико-экономических процессов, где отыскивается наилучшее (оптимальное) решение. Само возникновение и развитие линейного программирования непосредственно связано с экономической проблематикой. Термин «линейное программирование» появился в 1951 в работах американских учёных (Дж. Б. Данциг, Т. Купманс). Слово программирование выбрано потому, что набор переменных, подлежащих нахождению, обычно определяет программу (план) работы некоторого экономического объекта.

Первые исследования по линейному программированию (основные задачи и приложения, критерии оптимальности, геометрическая интерпретация методов нахождения оптимального решения, экономическая трактовка результатов математического анализа) были проведены в 30-годы в Ленинградском университете (Л. В. Канторович). Наиболее интенсивно линейное программирование развивалось в 1955-65 в СССР и США, когда оно было одним из наиболее «модных» разделов прикладной математики.

Именно после успехов линейного программирования появились новые подходы к решению задач минимизации выпуклых функций на выпуклом множестве и возрос интерес к нелинейным экстремальным задачам общего вида. Трудный класс задач оптимизации составляют дискретные задачи, в которых область изменения переменных состоит из отдельных точек. Многие такие задачи были сформулированы в период интенсивного развития линейного программирования, однако решению поддавались в основном задачи, родственные транспортной задаче линейного программирования. Впоследствии были разработаны численные методы решения достаточно обширных классов таких задач, которые в настоящее время являются одним из наиболее интенсивно развивающихся разделов математического программирования. Широкий класс нелинейных и дискретных задач может решаться с использованием идеи рекуррентного подхода (методов типа математической индукции). Такого рода методы объединяются названием динамическое программирование. Наконец, решение задач оптимизации со случайными параметрами в функциях, определяющих ограничения, составляет стохастическое программирование. К математическому программированию относят также (под названием бесконечномерное программирование) методы решения экстремальных задач с бесконечным числом переменных, например, такие, где набором переменных является функция (или несколько функций) непрерывного аргумента.

Эти разделы математического программирования возникли и первоначально развивались как отдельные дисциплины со своими особыми задачами и методами. Имевшиеся между ними связи были, как правило, односторонними (например, линейное программирование входит в выпуклое программирование, которое само входит в нелинейное программирование). В дальнейшем эти связи значительно расширились, и математическое программирование всё больше становится похожим на единую экстремальных задач теорию. Однако и сейчас в нём можно выделить несколько характерных групп задач и методов их решения.

Для решения задач на условный экстремум обычно используется метод множителей Лагранжа. Это правило было впервые сформулировано им в конце XVIII в. для исследования вариационных задач с ограничениями[*], и только потом – для конечномерных экстремальных задач. Данный метод представляет не только большой теоретический интерес, его можно использовать и для решения конкретных задач. Однако существует целый ряд препятствий применения метода Лагранжа (например, это проблема существования локальных условных экстремумов и отсутствия глобальных условных экстремумов).

[*] Первое упоминание о правиле множителей содержится в работах Эйлера по изопериметрическим задачам (1744 г.)

§1. КОНЕЧНОМЕРНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

1.1. Постановка задачи оптимизации

Пусть заданы множество X и функция $f(x)$, определенная на X. Рассматриваются конечномерные задачи оптимизации, т.е. задачи, допустимое множество которых лежит в евклидовом пространстве $\mathbf{R}^n$. Требуется найти точки минимума или максимума функции f на X. Всюду далее, если не оговорено иное, рассматривается задача

$$f(x) \to \min, \qquad x \in X. \tag{1.1}$$

Функция f называется *целевой*, X – *допустимым множеством*. Любой элемент $x \in X$ – допустимой точкой задачи (1.1).

Точка $x^* \in X$ называется:

1) *точкой глобального минимума* функции f на множестве X или глобальным решением задачи (1.1), если

$$f(x^*) \le f(x) \text{ при всех } x \in X; \tag{1.2}$$

2) *точкой локального минимума* f на X или локальным решением задачи (1.1), если существует число $\varepsilon > 0$ такое, что

$$f(x^*) \le f(x) \text{ при всех } x \in X \cap U_\varepsilon(x^*), \tag{1.3}$$

где $U_\varepsilon(x^*) = \{x \in \mathbf{R}^n \mid \| x - x^* \| < \varepsilon\}$ – шар радиуса $\varepsilon > 0$ с центром в x^*.

Если неравенство в (1.2) или (1.3) выполняется как строгое при $x \ne x^*$, то говорят, что x^* – точка *строгого* минимума (строгое решение) в глобальном или локальном смысле.

Для отражения того факта, что точка $x^* \in X$ является точкой глобального минимума функции f на X используется запись

$$f(x^*) = \min_{x \in X} f(x)$$

или эквивалентная ей

$$x^* = \arg\min_{x \in X} f(x).$$

$\operatorname{Arg}\min\limits_{x \in X} f(x) = \{x^* \in X \mid f(x^*) = f^*\}$ – множество всех точек глобального минимума f на X. Таким образом, $\arg\min\limits_{x \in X} f(x)$ – есть точка из множества $\operatorname{Arg}\min\limits_{x \in X} f(x)$.

Для задачи максимизации функции f на множестве X

$$f(x) \to \max, \qquad x \in X \tag{1.4}$$

соответствующие понятия, получаются заменой в данных выше определениях слова «минимум» на «максимум» и знака неравенства в (1.2), (1.3) на противоположный.

Решения задач (1.1), (1.4), т.е. точки минимума и максимума функции f на X, называются также *точками экстремума*, а сами задачи (1.1), (1.4) – *экстремальными* задачами.

Задача (1.4) эквивалентна задаче

$$-f(x) \to \min, \qquad x \in X$$

в том смысле, что множества глобальных и локальных, строгих или нестрогих решений этих задач соответственно совпадают. Это позволяет без труда переносить результаты, полученные для задачи минимизации, на задачи максимизации, и наоборот.

Один из основных вопросов при изучении задач оптимизации – вопрос установления факта существования решения. Часто он решается посредством применения теоремы Вейерштрасса и следствий из нее.

Теорема 1.1 (Вейерштрасса). Пусть X – компакт в $\mathbf{R}^n$, f – непрерывная функция на X. Тогда точка глобального минимума функции f на X (глобальное решение задачи (1.1)) существует.

Следствие 1.1. Пусть X – замкнутое множество в $\mathbf{R}^n$, f – непрерывная функция на X, причем при некотором $x^0 \in X$ множество $N(x^0) = \{x \in X \mid f(x) \le f(x^0)\}$ ограничено. Тогда точка глобального минимума функции f на X существует.

Следствие 1.2. Если функция f непрерывна на $\mathbf{R}^n$ и $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$, то функция f достигает своего глобального минимума на любом замкнутом подмножестве $\mathbf{R}^n$.

Задача (1.1) представляет собой общую постановку задачи оптимизации. Классификацию задач оптимизации можно проводить по нескольким признакам в зависимости от вида функции f и множества X.

1.2. Задача безусловной оптимизации

Задача (1.1) называется задачей *безусловной* оптимизации, если $X = \mathbf{R}^n$, т.е. если она имеет вид

$$f(x) \to \min, \qquad x \in \mathbf{R}^n. \tag{1.5}$$

При изучении любого типа задач оптимизации важное место занимает вопрос об *условиях оптимальности* или, как еще говорят, *условиях экстремума*. Различают *необходимые* условия оптимальности, т.е. условия, которым должна удовлетворять точка, являющаяся решением задачи, и *достаточные* условия оптимальности, т.е. условия, из которых следует, что данная точка является решением задачи. Интерес к условиям оптимальности объясняется тем, что они, во-первых, составляют основу качественных методов теории оптимизации, т.е. методов, направленных на изучение свойств экстремальных задач; во-вторых, используются при построении и обосновании численных методов решения этих задач; в-третьих, позволяют в простых случаях явно решить задачу.

Следующая теорема указывает необходимое условие локальной оптимальности первого порядка.

Теорема 1.2. Пусть функция f дифференцируема в точке $x^* \in \mathbf{R}^n$. Если x^* – локальное решение задачи (1.5), то

$$f'(x^*) = 0. \tag{1.6}$$

Точка x^*, удовлетворяющая условию (1.6), называется *стационарной* точкой функции f или задачи (1.5). Ясно, что стационарная точка не обязана быть решением, т.е. (1.6) не является достаточным условием оптимальности. Для выявления «посторонних» стационарных точек может использоваться необходимое условие локальной оптимальности второго порядка.

Теорема 1.3. Пусть функция f дважды дифференцируема в точке $x^* \in \mathbf{R}^n$. Если x^* – локальное решение задачи (1.5), то матрица $f''(x^*)$ неотрицательно определена, т.е.

$$\langle f''(x^*)h,\, h\rangle \ge 0 \quad \text{при всех} \quad h \in \mathbf{R}^n. \tag{1.7}$$

Достаточное условие локальной оптимальности содержит характерное усиление требований на матрицу $f''(x^*)$.

Теорема 1.4. Пусть функция f дважды дифференцируема в точке $x^* \in \mathbf{R}^n$. Предположим, что $f'(x^*) = 0$, а матрица $f''(x^*)$ положительно определена. т.е.

$$\langle f''(x^*)h, h\rangle > 0 \quad \text{при всех} \quad h \in \mathbf{R}^n,\ h \ne 0. \tag{1.8}$$

Тогда x^* – строгое локальное решение задачи (1.5).

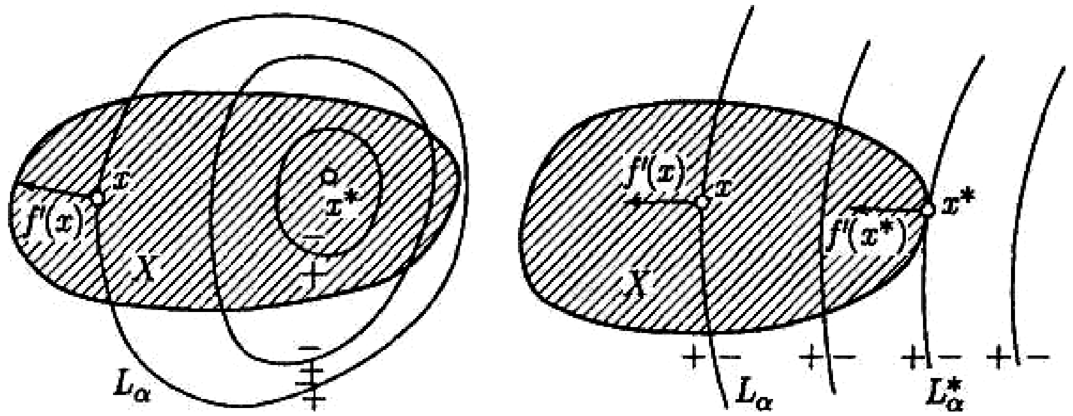
В тех случаях, когда функция f достаточно проста, теоремы 1.2-1.4 позволяют явным образом решить задачу (1.5). При этом для исследования матрицы $f''(x^*)$ на неотрицательную и положительную определенность, как правило, используется критерий Сильвестра.

1.3. Задача условной оптимизации

Задача (1.1) называется задачей *условной* оптимизации (*условной* задачей), если X – собственное подмножество пространства $\mathbf{R}^n$. Очевидно, что утверждения теорем 1.2-1.4 имеют место и для такой задачи, если ее локальное решение x^* является внутренней точкой допустимого множества X ($x^* \in \operatorname{int} X$). Однако для многих условных задач минимум достигается именно на границе, в силу чего для них эти классические результаты анализа неприменимы.

Для геометрической интерпретации двумерной задачи условной оптимизации необходимо изобразить ее допустимое множество X и несколько характерных линий уровня целевой функции f (рис. 1).

Линия уровня функции f – множество вида $L_\alpha = \{x \mid f(x) = \alpha\}$, $\alpha \in \mathbf{R}$. Чтобы отразить характер изменения функции, у данной линии уровня L_α полезно ставить знак «+» с той стороны, где f принимает значения, большие α, и знак «–» – с другой. Здесь уместно напомнить, что если функция f дифференцируема в точке x, то градиент $f'(x)$ ортогонален к проходящей через x линии уровня и направлен (если, естественно, $f'(x) \neq 0$) в сторону возрастания функции, т.е. в сторону знака «+». В геометрическом плане поиск (глобального) решения сводится к нахождению минимального числа α^* среди всех α таких, что линия уровня L_α имеет непустое пересечение с X. При этом любая точка $x^* \in L_{\alpha^*} \cap X$ является решением задачи, а само $\alpha^* = f(x^*)$ – минимальным значением функции f на X.



а *б*

Рис. 1.1. Геометрическая интерпретация задач оптимизации

1.4. Классическая задача на условный экстремум

Наряду с безусловной задачей, в курсе математического анализа изучается так называемая *классическая задача на условный экстремум*. Это задача (1.1) (или (1.4)) с допустимым множеством X, заданным системой конечного числа уравнений:

$$X = \{x \in \mathbf{R}^{\boldsymbol{n}} \mid g_i(x) = 0,\ i = 1,\dots,m\}.$$

Обычно эта задача записывается в виде

$$\begin{aligned} &f(x) \to \min, \\ &g_i(x) = 0,\ i = 1,\dots,m, \end{aligned} \tag{1.9}$$

т.е. явно указывается не само допустимое множество, а система, его определяющая.

При исследовании задачи (1.9) важную роль играет ее *функция Лагранжа*, т.е. функция

$$L(x, y_0, y) = y_0 f(x) + \sum_{i=1}^{m} y_i g_i(x),$$

где $x \in \mathbf{R}^{\boldsymbol{n}}$, $y_0 \in \mathbf{R}$, $y = (y_1,\dots,y_m) \in \mathbf{R}^{\boldsymbol{m}}$. Частные производные этой функции по координатам вектора x имеют вид

$$\frac{\partial L}{\partial x_j}(x, y_0, y) = y_0 \frac{\partial f}{\partial x_j}(x) + \sum_{i=1}^{m} y_i \frac{\partial g_i}{\partial x_j}(x),\ j = 1,\dots,n.$$

Составленный из них вектор обозначается через $L'_x(x, y_0, y)$, т.е.

$$L'_x(x, y_0, y) = y_0 f'(x) + \sum_{i=1}^{m} y_i g'_i(x). \tag{1.10}$$

Имеет место следующая теорема о необходимых условиях локальной оптимальности в задаче (1.9), известная как *правило множителей Лагранжа*.

Теорема 1.5. Пусть функции f, $g_1,\dots,g_m$ непрерывно дифференцируемы в некоторой окрестности точки $x^* \in \mathbf{R}^{\boldsymbol{n}}$. Если x^* – локальное решение задачи (1.9), то существуют число $y_0^* \in \mathbf{R}$ и вектор $y^* = (y_1^*,\dots,y_m^*) \in \mathbf{R}^{\boldsymbol{m}}$, не равные нулю одновременно и такие, что

$$L'_x(x^*, y_0^*, y^*) = 0. \tag{1.11}$$

Если при этом градиенты $g'_1(x^*),\dots,g'_m(x^*)$ линейно независимы (условие регулярности), то $y_0^* \neq 0$.

Условие (1.11) с учетом (1.10) означает, что градиенты $f'(x^*)$, $g_1'(x^*),\dots,g_m'(x^*)$ линейно зависимы. В частности, если m=1, то $f'(x^*)$ и $g_1'(x^*)$ коллинеарны.

Фигурирующие в теореме 1.5 числа y_0^*, $y_1^*,\dots,y_m^*$, называются *множителями Лагранжа*. Любая точка x^*, удовлетворяющая при некоторых y_0^* и y^*, $(y_0^*,y^*)\neq 0$, условию (1.11), а также условию *допустимости*

$$g_i(x^*)=0,\ i=1,\dots,m, \tag{1.12}$$

называется *стационарной* точкой задачи (1.9). Таким образом, стационарные точки определяются системой (1.11), (1.12), состоящей из $n+m$ уравнений с $n+m+1$ неизвестными. В случае если $y_0^*\neq 0$, то всегда можно считать $y_0^*=1$. Для достижения этого следует поделить все множители Лагранжа на y_0^*. В связи с этим при решении задач достаточно исследовать лишь два случая: $y_0^*=0$ или $y_0^*=1$, и, следовательно, неизвестных в системе фактически $n+m$. Если же выполняется указанное в теореме 1.5 условие регулярности, то все сводится к случаю $y_0^*=1$ и можно ограничиться рассмотрением функции Лагранжа вида

$$L(x,y)=f(x)+\sum_{i=1}^{m}y_i g_i(x),$$

которую называют *регулярной*.

Как и в случае безусловной задачи оптимизации, стационарные точки задачи (1.9) не обязаны быть ее решениями. Дальнейшее исследование может опираться на необходимые и достаточные условия оптимальности с привлечением вторых производных. Обозначим через

$$L''_{xx}(x,y_0,y)=y_0 f''(x)+\sum_{i=1}^{m}y_i g_i''(x)$$

матрицу вторых производных функции Лагранжа по координатам вектора x.

Теорема 1.6. Пусть функции $f, g_1,\dots,g_m$ дважды дифференцируемы в точке $x^*\in\mathbf{R}^n$ и непрерывно дифференцируемы в ее некоторой окрестности, причем градиенты $g_1'(x^*),\dots,g_m'(x^*)$ линейно независимы. Если x^* – локальное решение задачи (1.9), то

$$\langle L''_{xx}(x^*,y_0^*,y^*)h,\ h\rangle\geq 0$$

при любых y_0^*, y^*, удовлетворяющих (1.12), и всех $h\in\mathbf{R}^n$ таких, что

$$\langle g_i'(x^*),\ h\rangle=0,\quad i=1,\dots,m. \tag{1.13}$$

Теорема 1.7. Пусть функции $f, g_1,\ldots,g_m$ дважды дифференцируемы в точке $x^* \in \mathbf{R}^n$, удовлетворяющей (1.12). Предположим, что при некоторых y_0^* и y^* выполняется условие (1.11) и, кроме того,

$$\langle L''_{xx}(x^*, y_0^*, y^*)h,\, h\rangle > 0$$

при всех ненулевых $h \in \mathbf{R}^n$, удовлетворяющих (1.13). Тогда x^* – строгое локальное решение задачи (1.9).

1.5. Выпуклая задача оптимизации

Задача (1.1) называется *выпуклой*, если X – выпуклое множество, f – выпуклая функция на X.

Подобные задачи обладают рядом важных свойств.

Предположим, что x^* – локальное решение выпуклой задачи (1.1), т.е. при некотором $\varepsilon > 0$ выполняется (1.3). Для любой точки $x \in X,\ x \neq x^*$, положим $\lambda = \min(\varepsilon / \|x\|, 1)$. Тогда $\lambda x + (1-\lambda)x^* \in X \cap U_\varepsilon(x^*)$ и, следовательно,

$$f(x^*) \le f(\lambda x + (1-\lambda)x^*) \le \lambda f(x) + (1-\lambda) f(x^*).$$

Отсюда $f(x^*) \le f(x)$. т.е. x^* – глобальное решение задачи.

Таким образом, для выпуклых задач понятия локального и глобального решений не различаются.

Теорема 1.8. Если задача (1.1) выпукла, то любое ее локальное решение является также глобальным.

Укажем еще одно полезное свойство выпуклых задач.

Теорема 1.9. Пусть задача (1.1) выпукла и имеет решение. Тогда множество ее решений $X^* = \operatorname{Arg}\min\limits_{x \in X} f(x)$ выпукло. Если при этом f строго выпукла на X, то решение задачи единственно.

Доказательство. Пусть $x^1, x^2 \in X^*,\ \lambda \in [0, 1]$. Тогда $f(x^1) = f(x^2) = f^*$. При этом

$$f(\lambda x^1 + (1-\lambda)x^2) \le \lambda f(x^1) + (1-\lambda) f(x^2) = f^*. \qquad (1.14)$$

По определению X^* неравенство здесь может выполняться только как равенство. Следовательно, $\lambda x^1 + (1-\lambda)x^2 \in X^*$, т.е. X^* выпукло.

Пусть f строго выпукла. Если предположить, что в X^* существуют две различные точки x^1 и x^2, то при $\lambda \in (0, 1)$ неравенство в (1.14) должно быть строгим, что невозможно.

Следующее свойство выпуклых задач можно высказать в виде следующего общего принципа: необходимые условия оптимальности в том или ином классе задач оптимизации при соответствующих предположениях выпуклости оказываются и достаточными. Так для выпуклой задачи имеет место следующая теорема

Теорема 1.10. Пусть функция f выпукла на $\mathbf{R}^n$ и дифференцируема в точке $x^* \in \mathbf{R}^n$. Если $f'(x^*) = 0$, то x^* – точка минимума f на X, т.е. решение задачи (1.5).

Доказательство. Для любых $x \in X$ и $\lambda \in (0, 1]$ имеем

$$f(\lambda x + (1-\lambda)x^*) \le \lambda f(x) + (1-\lambda) f(x^*).$$

Преобразуя эту формулу, а затем, пользуясь дифференцируемостью f в точке x^*, получаем

$$f(x) - f(x^*) \ge \frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda} = \frac{\langle f'(x^*), \lambda(x - x^*)\rangle + o(\lambda)}{\lambda} = \frac{o(\lambda)}{\lambda}.$$

Отсюда предельным переходом при $\lambda \to 0$ следует, что $f(x) \ge f(x^*)$.

Полученные свойства выпуклых задач имеют важное значение не только в теории, но и в численных методах оптимизации. Дело в том, что большинство существующих численных методов позволяет, вообще говоря, находить лишь локальные решения, а точнее, стационарные точки задачи. Теоремы 1.8, 1.10 говорят о том, что для выпуклой задачи отыскание стационарной точки автоматически означает отыскание решения, причем глобального, если же целевая функция строго выпукла, то решение единственно.

1.6. Задача математического программирования

Важнейший класс условных задач оптимизации составляют задачи *математического программирования*. Так принято называть задачу (1.1) (или (1.4)), если ее допустимое множество имеет вид

$$X = \{x \in P \mid g_i(x) \le 0,\ i = 1,\ldots,k;\ g_i(x) = 0,\ i = k+1,\ldots,m\},$$

т.е. задается системой конечного числа неравенств и уравнений, рассматриваемых, вообще говоря, на некотором множестве $P \subset \mathbf{R}^n$. Иными словами, задачу математического программирования можно записать в виде

$$\begin{aligned} &f(x) \to \min, \\ &g_i(x) \le 0,\ i = 1,\ldots,k, \\ &g_i(x) = 0,\ i = k+1,\ldots,m, \quad x \in P. \end{aligned} \tag{1.15}$$

Условия типа $g_i(x) \le 0$ называются *ограничениями-неравенствами*; типа $g_i(x) = 0$ – *ограничениями-равенствами*. Оба этих типа условий – *функциональными ограничениями*; условие $x \in P$ носит название *прямого ограничения*.

Классическая задача на условный экстремум является частным случаем задачи математического программирования, когда ограничения-неравенства и прямое ограничение отсутствуют.

Деление ограничений задачи математического программирования на функциональные и прямые условно. Например, в задаче (1.15) можно исключить ограничения-равенства путем введения их в прямое ограничение, т.е. перейти к задаче вида

$$f(x) \to \min,$$
$$g_i(x) \le 0,\ i = 1,\dots,k, \quad x \in P',$$

где $P' = \{x \in P \mid g_i(x) = 0,\ i = k+1,\dots,m\}$. Исключение всех функциональных ограничений лишь возвращает нас к задаче вида (1.1). То или иное представление задачи определяется целями исследования. Однако, как правило, чем подробнее описаны функциональные ограничения задачи, тем детальнее удается изучить ее свойства. Обычно в качестве *P* берется множество «простой» структуры, например координатный параллелепипед

$$P = \{x \in \mathbf{R}^n \mid a_j \le x_j \le b_j,\ j = 1,\dots,n\},$$

причем здесь не исключается, что $a_j = -\infty$ или $b_j = +\infty$ при тех или иных *j*. Особенно часто используются $P = \mathbf{R}^n_+$ (когда $a_j = 0$, $b_j = +\infty$ при всех *j*) или просто $P = \mathbf{R}^n$.

1.7. Задачи выпуклого программирования

Задача (1.15) называется задачей *выпуклого программирования*. если множество *P* выпукло, функции f, $g_1,\dots,g_k$ выпуклы на *P*, функции $g_{k+1},\dots,g_m$ линейны. Очевидно, что допустимое множество задачи выпуклого программирования выпукло и, следовательно, задача выпукла.

Пример 1.1. Задача планирования производства

Пусть имеется предприятие, выпускающее определенный товар и использующее при этом *n* ресурсов. Предприятие характеризуется *технологическим* множеством $X \subset \mathbf{R}^n_+$, указывающим всевозможные наборы ресурсов, из которых может быть получен данный товар, и *производственной функцией* *f*(*x*), показывающей объем выпуска при наборе затраченных ресурсов $x = (x_1,\dots,x_n) \in X$. Например, множество X может определяться условиями

$$x_1 \geq 0, \quad a_i x_1 \leq x_i \leq b_i x_1, \quad i = 2, \ldots, n,$$

где a_i, b_i – заданные числа, $0 \leq a_i \leq b_i$, т.е. пропорции между затратами первого (базового) и остальных ресурсов должны находиться в определенных пределах. В качестве производственной функции часто фигурирует так называемая *функция Кобба-Дугласа*

$$f(x) = \gamma \prod_{i=1}^{n} x_i^{\lambda_i},$$

где $\gamma > 0, \lambda_1 \geq 0, \ldots, \lambda_n \geq 0$ – заданные числа, $\sum_{i=1}^{n} \lambda_i = 1$. Отметим, что эта функция вогнута.

Пусть $p_i > 0$ – известная цена единицы *i*-го ресурса, $p = (p_1, \ldots, p_n)$ – вектор цен на ресурсы. Тогда $\langle p, x \rangle$ – денежная стоимость данного набора ресурсов *x*. Если $\beta > 0$ – денежные средства предприятия, то можно сформулировать задачу о максимизации объема выпуска в рамках бюджетных и технологических ограничений:

$$f(x) \to \max, \quad \langle p, x \rangle \leq \beta, \quad x \in X.$$

Если задан параметр $\bar{f} > 0$ – плановое задание предприятию по выпуску товара, то можно поставить задачу о минимизации издержек производства при безусловном выполнении плана:

$$\langle p, x \rangle \to \min, \quad f(x) \geq \bar{f}, \quad x \in X.$$

1.8. Задачи линейного и квадратичного программирования

Задачей *линейного программирования* (ЛП) называется задача минимизации или максимизации линейной функции при линейных ограничениях. Так, задача (1.21) является задачей ЛП, если $f, g_1, \ldots, g_m$ – линейные функции, а *P* – полиэдр, т.е. множество, само задаваемое линейными условиями.

Задачей *квадратичного программирования* называется задача минимизации при линейных ограничениях *квадратичной функции* вида

$$f(x) = \langle Cx, x \rangle + \langle d, x \rangle,$$

где *C* – симметрическая неотрицательно определенная матрица размера $n \times n$, *d* – вектор из $\mathbf{R}^n$. При указанном условии на *C* эта функция выпукла.

Таким образом, задача квадратичного программирования – это частный случай задачи выпуклого программирования, а задача ЛП – частный случай их обеих (*C* = 0).

§2. ТЕОРИЯ НЕОБХОДИМЫХ И ДОСТАТОЧНЫХ УСЛОВИЙ ОПТИМАЛЬНОСТИ

2.1. Условия оптимальности в общей задаче минимизации

Рассматривается задача минимизации:

$$f(x) \to \min, \qquad x \in X. \tag{2.1}$$

Вектор $h \in \mathbf{R}^n$ задает *направление убывания* функции *f* в точке $x^* \in \mathbf{R}^n$, если $f(x^* + \alpha h) < f(x^*)$ при всех достаточно малых $\alpha > 0$. Множество всех таких *h* обозначается через $U(x^*, f)$.

Вектор $h \in \mathbf{R}^n$ задает *возможное направление* относительно множества *X* в точке $x^* \in X$, если $x^* + \alpha h \in X$ при всех достаточно малых $\alpha > 0$. Множество всех таких *h* обозначим через $V(x^*, X)$.

Теорема 2.1. Если x^* – локальное решение задачи (2.1), то

$$U(x^*, f) \cap V(x^*, X) = \varnothing. \tag{2.2}$$

Доказательство. Допустим, что (2.2) неверно, т.е. существует вектор $h \in \mathbf{R}^n$, для которого $f(x^* + \alpha h) < f(x^*)$ и $x^* + \alpha h \in X$ при всех достаточно малых $\alpha > 0$. Стало быть, в любой сколь угодно малой окрестности точки x^* найдется точка $x \in X$ (вида $x = x^* + \alpha h$) такая, что $f(x) < f(x^*)$, но это противоречит определению локального решения.

Теорема 2.1 означает, что из точки x^*, являющейся локальным решением, нельзя осуществить сколь угодно малый сдвиг вдоль какого бы то ни было луча так, чтобы уменьшить значение целевой функции и остаться при этом в пределах допустимого множества.

Теорема 2.2. Пусть в задаче (2.1) множество *X* выпукло, функция *f* дифференцируема в точке $x^* \in X$.

Тогда:

1) если x^* – локальное решение задачи (2.1), то

$$\langle f'(x^*), x - x^* \rangle \ge 0 \quad \text{при всех} \quad x \in X; \tag{2.3}$$

2) если функция *f* выпукла на *X* и выполняется (2.3), то x^* – (глобальное) решение задачи (2.1).

Доказательство. 1) Допустим, что (2.3) неверно, т.е. $\langle f'(x^*), x - x^* \rangle \ge 0$ при некотором $x \in X$. Положим $h = x - x^*$. Тогда $h \in U(x^*, f)$. В то же время с учетом выпуклости *X* для любого $\alpha \in [0, 1]$ имеем $x^* + \alpha h = \alpha x + (1 - \alpha) x^* \in X$, т.е. $h \in V(x^*, X)$. Получаем противоречие с теоремой 2.1.

Утверждение 2) очевидно (см. теоремы 1.8, 1.10).

Таким образом, соотношение (2.3) является необходимым условием локальной оптимальности в задаче минимизации дифференцируемой функции на выпуклом множестве; причем, для выпуклой задачи оно оказывается и достаточным условием глобальной оптимальности.

Геометрически условие (2.3) означает, что градиент $f'(x^*)$ (если он отличен от нуля) составляет нетупой угол с вектором, направленным из x^* в любую точку $x \in X$. Точка *x*, которая не удовлетворяет данному условию, не может быть решением: из нее можно совершить сдвиг, уменьшая значение функции *f* и оставаясь в пределах *X*.

Полезно конкретизировать условие (2.3) в некоторых частных случаях.

Лемма 2.1. Если $x^* \in \operatorname{int} X$, то (2.3) эквивалентно условию:

$$f'(x^*) = 0, \quad \text{т.е.} \quad \frac{\partial f}{\partial x_j}(x^*) = 0, \quad j = 1, \dots, n. \tag{2.4}$$

Доказательство. При достаточно малом $\alpha > 0$ точка $x = x^* - \alpha f'(x^*)$ принадлежит *X*. Подставляя *x* в (2.3), получаем (2.4). Напротив, из (2.4) тривиально следует (2.3).

Лемма 2.2. Пусть множество X имеет вид

$$X = \{x \in \mathbf{R}^{n} \mid a_j \le x_j \le b_j,\ j = 1, \dots, n\}, \tag{2.5}$$

где $-\infty \le a_j < b_j \le +\infty,\ j = 1, \dots, n$ (если $a_j = -\infty$ или $b_j = +\infty$, то соответствующий знак неравенства в (2.5) следует понимать как строгий). Тогда (2.3) эквивалентно условию: для любого $j = 1, \dots, n$

$$\frac{\partial f}{\partial x_j}(x^*) \quad \begin{cases} = 0, \text{ если } a_j < x_j^* < b_j, \\ \ge 0, \text{ если } x_j^* = a_j \ne -\infty, \\ \le 0, \text{ если } x_j^* = b_j \ne +\infty. \end{cases}$$

Лемма 2.3. Пусть множество X имеет вид

$$X = \{x \in \mathbf{R}^{n} \mid x_j \ge 0,\ j = 1, \dots, s\},$$

где $0 \le s \le n$ (*s* = 0 соответствует $X = \mathbf{R}^{n}$). Тогда (2.3) эквивалентно совокупности условий:

$$\frac{\partial f}{\partial x_j}(x^*) \ge 0, \quad x_j^* \frac{\partial f}{\partial x_j}(x^*) = 0, \quad j = 1, \dots, s,$$

$$\frac{\partial f}{\partial x_j}(x^*) = 0, \quad j = s+1, \ldots, n.$$

2.2 Дифференциальные условия оптимальности в задаче математического программирования

Рассматривается задача математического программирования вида

$$\begin{aligned} & f(x) \to \min, \\ & g_i(x) \le 0,\ i = 1, \ldots, k, \\ & g_i(x) = 0,\ i = k+1, \ldots, m, \quad x \in P. \end{aligned} \tag{2.6}$$

Обозначим через X допустимое множество задачи (2.6), т.е.

$$X = \{x \in P \mid \ g_i(x) \le 0,\ i = 1, \ldots, k;\ g_i(x) = 0,\ i = k+1, \ldots, m\}.$$

Кроме того, свяжем с этой задачей множество

$$Q = \{y = (y_1, \ldots, y_m) \in \mathbf{R}^{m} \mid y_i \ge 0,\ i = 1, \ldots, k\},$$

состоящее из всех m-мерных векторов, у которых первые k координат неотрицательны. В частности, $Q = \mathbf{R}^{m}$, если ограничения-неравенства отсутствуют ($k = 0$); $Q = \mathbf{R}^{m}_{+}$, если ограничения-равенства отсутствуют ($k = m$).

Введем *функцию Лагранжа* задачи (2.6)

$$L(x, y_0, y) = y_0 f(x) + \sum_{i=1}^{m} y_i g_i(x),$$

где $x \in P$, $y_0 \ge 0$, $y = (y_1, \ldots, y_m) \in Q$. Данная функция имеет такой же вид, как и в случае классической задачи на условный экстремум.

Обобщением правила множителей Лагранжа в классической задаче на условный экстремум является следующая теорема, занимающая центральное место в теории дифференциальных условий оптимальности.

Теорема 2.3 (принцип Лагранжа). Пусть в задаче (2.6) множество P выпукло, функции $f, g_1, \ldots, g_k$ дифференцируемы в точке $x^* \in X$, функции $g_{k+1}, \ldots, g_m$ непрерывно дифференцируемы в некоторой окрестности x^*. Если x^* – локальное решение задачи (2.6), то существуют число $y_0^* \ge 0$ и вектор $y^* = (y_1^*, \ldots, y_m^*) \in Q$, не равные нулю одновременно и такие, что

$$\langle L_x'(x^*, y_0^*, y^*), x - x^* \rangle \ge 0 \quad \text{при всех} \quad x \in P, \tag{2.7}$$

$$y_i^* g_i(x^*) = 0,\ i = 1, \ldots, k. \tag{2.8}$$

Любая точка $x^* \in X$, удовлетворяющая условиям (2.7), (2.8) при некоторых $y_0^* \geq 0$, $y^* \in Q$, $(y_0^*, y^*) \neq 0$, называется *стационарной* точкой задачи (2.6). Принцип Лагранжа утверждает, что при указанных предположениях любое локальное решение задачи (2.6) является стационарной точкой. Обратное гарантируется лишь при тех или иных дополнительных предположениях о задаче (2.6).

Числа $y_0^*, y_1^*, \ldots, y_m^*$ называются *множителями Лагранжа*. Согласно определению Q множители $y_1^*, \ldots, y_k^*$, соответствующие ограничениям-неравенствам, неотрицательны, а множители $y_{k+1}^*, \ldots, y_m^*$ – соответствующие ограничениям-равенствам, могут иметь любой знак.

Отметим, что множители Лагранжа определены с точностью до положительной константы, т.е. если пара (y_0^*, y^*) удовлетворяет условиям (2.7), (2.8), то при любом $\lambda > 0$ пара $(\lambda y_0^*, \lambda y^*)$ обладает тем же свойством. В частности, можно взять $\lambda = 1/y_0^*$, если $y_0^* > 0$. Это позволяет без ограничения общности рассматривать в теореме 2.1 лишь два случая: $y_0^* = 0$ или $y_0^* = 1$. Любое дополнительное предположение о задаче (2.1), обеспечивающее в рамках данной теоремы случай $y_0^* = 1$, принято называть *условием регулярности*. При этом саму задачу называют *регулярной*. Для такой задачи достаточно рассматривать лишь функцию Лагранжа вида

$$L(x, y) = L(x, 1, y) = f(x) + \sum_{i=1}^{m} y_i g_i(x), \tag{2.9}$$

которую также называют *регулярной*.

Условие (2.8) называют *условием дополняющей нежесткости*.

Для любой точки $x^* \in X$ введем в рассмотрение множества

$$I(x^*) = \{i \mid g_i(x) = 0, 1 \leq i \leq k\},$$
$$S(x^*) = \{i \mid g_i(x) = 0, 1 \leq i \leq m\}.$$

Ограничения-неравенства с индексами $i \in I(x^*)$ называются *активными* в точке x^*, а остальные – *пассивными*.

Условие (2.8) означает, что множители Лагранжа, соответствующие пассивным ограничениям-неравенствам, должны обращаться в нуль, т.е. $y_i^* = 0$ при всех $i \in \{1, \ldots, k\} \setminus I(x^*)$. Таким образом, из условий (2.7), (2.8) следует

$$\langle y_0^* f'(x^*) + \sum_{i \in S(x^*)} y_i^* g_i'(x^*),\ x - x^* \rangle \geq 0 \quad \text{при всех} \quad x \in P. \tag{2.10}$$

Напротив, имея (2.10), всегда можно прийти к (2.7), (2.8), положив $y_i^* = 0$ при $i \in \{1, \ldots, k\} \setminus I(x^*)$.

Опираясь на леммы 2.1-2.3, конкретизируем условие (2.7) в некоторых частных случаях.

Лемма 2.4. 1) Если $x^* \in \operatorname{int} P$, то (2.7) эквивалентно условию:

$$L'_x(x^*, y_0^*, y^*) = 0, \quad \text{т.е.} \quad \frac{\partial L}{\partial x_j}(x^*, y_0^*, y^*) = 0, \quad j = 1, \dots, n;$$

2) если P имеет вид

$$P = \{x \in \mathbf{R}^{\boldsymbol{n}} \mid a_j \le x_j \le b_j,\ j = 1,\dots,n\},$$

где $-\infty \le a_j < b_j \le +\infty$, $j = 1,\dots,n$, то (2.7) эквивалентно условию: для любого $j = 1,\dots,n$

$$\frac{\partial L}{\partial x_j}(x^*, y_0^*, y^*) \begin{cases} = 0, \text{ если } a_j < x_j^* < b_j, \\ \ge 0, \text{ если } x_j^* = a_j \ne -\infty, \\ \le 0, \text{ если } x_j^* = b_j \ne +\infty; \end{cases}$$

3) если P имеет вид

$$P = \{x \in \mathbf{R}^{\boldsymbol{n}} \mid x_j \ge 0,\ j = 1,\dots,s\},$$

где $0 \le s \le n$, то (2.7) эквивалентно совокупности условий:

$$\frac{\partial L}{\partial x_j}(x^*, y_0^*, y^*) \ge 0, \quad x_j^* \frac{\partial L}{\partial x_j}(x^*, y_0^*, y^*) = 0, \quad j = 1,\dots,s,$$

$$\frac{\partial L}{\partial x_j}(x^*, y_0^*, y^*) = 0, \quad j = s+1,\dots,n.$$

Пример 2.1. Рассмотрим задачу:

$$x_1^2 + 4x_1 x_2 + x_2^2 \to \min, \ x_1^2 + x_2^2 - 1 = 0.$$

Составим функцию Лагранжа

$$L = y_0(x_1^2 + 4x_1 x_2 + x_2^2) + y_1(x_1^2 + x_2^2 - 1).$$

Необходимое условие экстремума:

$$y_0(2x_1 + 4x_2) + 2y_1 x_1 = 0, \ y_0(4x_1 + 2x_2) + 2y_1 x_2 = 0.$$

1) Пусть $y_0 = 0$, тогда $y_1 \ne 0$, а значит $x_1 = x_2 = 0$, что противоречит ограничению (точка не является допустимой).

2) Пусть $y_0 \ne 0$ ($y_0 = 1$), тогда из системы

$$(1 + y_1)x_1 + 2x_2 = 0, \ 2x_1 + (1 + y_1)x_2 = 0,$$

Учитывая, что она должна иметь ненулевое решение, получаем, что или

$y_1 = -3,\ x_1 = x_2 = \pm\sqrt{2}/2$, или $y_1 = 1,\ x_1 = \pm\sqrt{2}/2,\ x_2 = \mp\sqrt{2}/2$. Применяя теорему Вейерштрасса, легко видеть, что решением задачи являются точки $(\sqrt{2}/2, -\sqrt{2}/2),\ (-\sqrt{2}/2, \sqrt{2}/2)$.

Пример 2.2. Рассмотрим задачу:

$$x_1^2 \to \min,\ P = \mathbf{R}^3,\ x_1^2 + x_2^2 - 1 = 0,\ x_1^3 + x_2^3 + x_3^3 - 1 = 0.$$

Точка (0, 1 , 0) является решением данной задачи. Для функции Лагранжа $L = y_0 x_1^2 + y_1(x_1^2 + x_2^2 - 1) + y_2(x_1^3 + x_2^3 + x_3^3 - 1)$ необходимые условия экстремума в этой точке имеют вид

$$\frac{\partial L}{\partial x_1} = 2y_0x_1 + 2y_1x_1 + 3y_2x_1^2 = 0,$$

$$\frac{\partial L}{\partial x_2} = 2y_1x_2 + 3y_2x_2^2 = 2y_1 + 3y_2 = 0,$$

$$\frac{\partial L}{\partial x_3} = 2y_1x_3 + 3y_2x_3^2 = 0.$$

Откуда видно, что y_0 может быть любым, $y_1,\ y_2$ – любые, удовлетворяющие условию $2y_1 + 3y_2 = 0$.

Пример 2.3. Рассмотрим задачу:

$$x_n \to \min,\ P = \mathbf{R}^n,\ \sum_{i=1}^{n} x_i^2 = 0.$$

Точка x = (0, … , 0) является единственной допустимой, а, следовательно, решением данной задачи. Для функции Лагранжа $L = y_0x_n + y_1\sum_{i=1}^{n} x_i^2$ необходимые условия экстремума имеют вид

$$\frac{\partial L}{\partial x_i} = 2y_1x_i = 0,\ i = 1,\dots,n-1,\ \frac{\partial L}{\partial x_n} = y_0 + 2y_1x_n = 0.$$

Откуда видно, что данные условия могут удовлетворяться в точке (0, … ,0) только при $y_0 = 0$, т.е. задача не является регулярной.

Следующая теорема показывает, что для регулярной задачи выпуклого программирования соотношения (2.7), (2.8) являются не только необходимым, но и достаточным условием оптимальности.

Теорема 2.4. Пусть в задаче (2.6) множество P выпукло, функции $f, g_1,\dots,g_k$ выпуклы на P и дифференцируемы в точке $x^* \in X$, функции $g_{k+1},\dots,g_m$ линейны. Если при $y_0^* = 1$ и некотором $y^* \in Q$ выполняются условия (2.7), (2.8), то x^* – (глобальное) решение задачи (2.6).

Доказательство. В данном случае функция $L(x, y^*)$, очевидно, выпукла по x на P. Тогда, согласно утверждению 2) теоремы 2.2, из условия (2.7) следует, что x^* – точка минимума функции $L(x, y^*)$ на P, т.е.

$$L(x^*, y^*) \le L(x, y^*) \quad \text{при всех} \quad x \in P.$$

С учетом этого, а также формулы (2.8) для любого $x \in X$ имеем

$$f(x^*) = f(x^*) + \sum_{i=1}^{m} y_i^* g_i(x^*) = L(x^*, y^*) \le L(x, y^*) = f(x) + \sum_{i=1}^{m} y_i^* g_i(x) \le f(x).$$

В связи с данной теоремой вопрос об эффективных условиях регулярности в задаче выпуклого программирования приобретает особо важное значение.

2.3. Условия регулярности. Дифференциальная форма теоремы Каруша-Куна-Таккера

Простейшим примером условия регулярности служит требование линейной независимости градиентов $g_1'(x^*)$, ... , $g_m'(x^*)$ в классической задаче на условный экстремум. Вообще условием регулярности в задаче (2.6) при $x^* \in \operatorname{int} P$ выступает линейная независимость градиентов $g_i'(x^*)$ $(i \in S(x^*))$. В самом деле, при $x^* \in \operatorname{int} P$ формулу (2.10) можно записать в виде

$$y_0^* f'(x^*) + \sum_{i \in S(x^*)} y_i^* g_i'(x^*) = 0,$$

где, как утверждается теоремой 2.3, числа y_0^*, y_i^* $(i \in S(x^*))$ не равны нулю одновременно. Ясно, что случай $y_0^* = 0$ здесь невозможен по определению линейной независимости.

Между тем такое условие регулярности труднопроверяемо. Более удобные удается получить в рамках задачи с выпуклыми ограничениями-неравенствами и линейными ограничениями-равенствами.

Теорема 2.5. Пусть в задаче (2.6) множество P выпукло, функции f, $g_1,\dots,g_k$ дифференцируемы, в точке $x^* \in X$, функции $g_1,\dots,g_k$ выпуклы на P, функции $g_{k+1},\dots,g_m$ линейны.

Предположим, что дополнительно выполняется хотя бы одно из следующих условий:

1) ограничения-равенства отсутствуют ($k = m$) и существует точка $\bar{x} \in P$ такая, что $g_i(\bar{x}) < 0$ при всех $i = 1,\dots,m$;

2) множество P – полиэдр, функции $g_1,\dots,g_k$ линейны.

Если x^* – локальное решение задачи (2.6), то существует вектор $y^* \in Q$ такой, что при $y_0^* = 1$ выполняются условия (2.7), (2.8).

Фигурирующее в теореме 2.5 условие 1) называется *условием Слейтера*. Это наиболее простое и часто используемое условие регулярности. Условие 2) называют *условием линейности*. Оно автоматически выполняется для задач линейного и квадратичного программирования.

Пример 2.4. Рассмотрим задачу:

$$x_1 x_2 .. x_n \to \max,\ x_1 + x_2 + \ldots + x_n = 1,\ x_i \geq 0,\ i = 1,\ldots,n.$$

Приведем задачу к виду

$$-x_1 x_2 .. x_n \to \min,\ x_1 + x_2 + \ldots + x_n - 1 = 0,\ -x_i \leq 0,\ i = 1,\ldots,n.$$

Составим регулярную функцию Лагранжа (выполняется условие 2) теоремы регулярности)

$$L = -x_1 x_2 .. x_n + y_1(x_1 + x_2 + \ldots + x_n - 1) - y_2 x_1 - \ldots - y_{n+1} x_n = 0.$$

Необходимые условия экстремума:

$$\begin{aligned}
&-x_2 x_3 .. x_n + y_1 - y_2 = 0,\\
&-x_1 x_3 .. x_n + y_1 - y_3 = 0,\\
&\qquad\qquad \ldots\\
&-x_1 x_2 .. x_{n-1} + y_1 - y_{n+1} = 0,\\
&-y_2 x_1 = 0, \ldots, -y_{n+1} x_n = 0.
\end{aligned}$$

Легко видеть, что никакая точка, имеющая хотя бы одну нулевую координату, не является решением поставленной задачи, а тогда уз условий дополняющей нежесткости имеем $y_2 = \ldots = y_{n+1} = 0$. Значит $y_1 = x_2 x_3 .. x_n$.

С другой стороны, сложив первые n уравнений из системы необходимых условий, предварительно умножив соответственно на x_i, получаем $y_1 = n x_1 x_1 .. x_n$. Таким образом, легко видеть, что $x_1 = \ldots = x_n = 1/n$ – единственная стационарная точка, подозрительная на решение задачи.

Из теорем 2.4 и 2.5 непосредственно следует один из важнейших фактов теории выпуклого программирования.

Теорема 2.6 (теорема Каруша-Куна-Таккера в дифференциальной форме). Пусть в дополнение к условиям теоремы 2.5 функция f выпукла на P. Тогда точка x^* является решением задачи (2.6) в том и только том случае, если существует вектор $y^* \in Q$ такой, что при $y_0^* = 1$ выполняются условия (2.7), (2.8).

2.4. Условия оптимальности второго порядка

Пусть

$$L''_{xx}(x, y_0, y) = y_0 f''(x) + \sum_{i=1}^{m} y_i g_i''(x)$$

– матрица вторых производных функции Лагранжа по координатам вектора x.

Для точки $x^* \in P$ введем множество

$$V(x^*) = \{h \in \mathbf{R}^n \mid h = \lambda(x - x^*),\ \lambda > 0,\ x \in P\}.$$

Ясно, что $V(x^*) = \mathbf{R}^n$, если $x^* \in \operatorname{int} P$.

Обозначим через $H(x^*)$ множество всех векторов $h \in \mathbf{R}^n$ таких, что

$$\langle f'(x^*), h\rangle \le 0,$$

$$\langle g_i'(x^*), h\rangle \le 0, \qquad i \in I(x^*), \tag{2.11}$$

$$\langle g_i'(x^*), h\rangle = 0, \qquad i = k+1, \ldots, m. \tag{2.12}$$

Приведем теорему о достаточных условиях оптимальности в общей (необязательно выпуклой и регулярной) задаче математического программирования.

Теорема 2.7. Пусть в задаче (2.6) функций $f, g_1, \ldots, g_m$ дважды дифференцируемы в точке $x^* \in X$. Предположим, что существуют число $y_0^* \ge 0$ и вектор $y^* \in Q$ такие, что выполняются условия (2.7), (2.8) и, кроме того,

$$\langle L''_{xx}(x^*, y_0^*, y^*)h, h\rangle > 0, \tag{2.13}$$

при всех ненулевых $h \in \overline{V(x^*)} \cap H(x^*)$ (черта означает замыкание). Тогда x^* – строгое локальное решение задачи (2.6), т.е. $f(x^*) < f(x)$ при всех $x \in X$, достаточно близких к x^*, но отличных от x^*.

Ясно, что для классической задачи на условный экстремум это следствие переходит в теорему 1.7.

Замечание. При любых $y_0^* \ge 0$ *и* $y^* \in Q$, удовлетворяющих (2.7), (2.8), из условия $h \in \overline{V(x^*)} \cap H(x^*)$ следует

$$\langle L'_x(x^*, y_0^*, y^*), h\rangle = 0,$$

$$y_0^* \langle f'(x^*), h\rangle = 0,$$

$$y_0^* \langle g_i'(x^*), h\rangle = 0, \quad i \in I(x^*). \tag{2.14}$$

Из приведенного замечания и теоремы 2.7 следует, что формулировка теоремы допускает некоторые модификации.

Следствие 2.1. Пусть в задаче (2.6) функции $f, g_1,\dots,g_m$ дважды дифференцируемы в точке $x^* \in X$. Предположим, что существуют $y_0^* \ge 0$ и $y^* \in Q$ такие, что выполняются условия (2.7), (2.8) и (2.13) при всех ненулевых $h \in \overline{V(x^*)}$, удовлетворяющих (2.11), (2.12), (2.14). Тогда x^* – строгое локальное решение задачи (2.6).

Следствие 2.2. Пусть в задаче (2.6) функции $f, g_1,\dots,g_m$ дифференцируемы в точке $x^* \in X$. Если $\overline{V(x^*)} \cap H(x^*) = \{0\}$, то x^* – строгое локальное решение задачи (2.6).

Заметим, что для задачи выпуклого программирования теорема 2.7 и ее следствия указывают достаточные условия *единственности* (глобального) решения.

Приведем теперь теорему о необходимом условии оптимальности второго порядка, ограничившись случаем $x^* \in \operatorname{int} P$.

Теорема 2.8. Пусть в задаче (2.6) множество P выпукло, функции $f, g_1,\dots,g_m$ дважды дифференцируемы в точке $x^* \in \operatorname{int} P \cap X$. Пусть, кроме того, функции g_i $(i \in S(x^*))$ непрерывно дифференцируемы в некоторой окрестности x^*, а градиенты $g'(x^*)_i$ $(i \in S(x^*))$ линейно независимы. Если x^* – локальное решение задачи (2.6), то

$$\langle L''_{xx}(x^*, y_0^*, y^*)h,\, h\rangle \ge 0,$$

при любых $y_0^* \ge 0, y^* \in Q$, удовлетворяющих условиям $L'_x(x^*, y_0^*, y^*) = 0$ и (2.8), и всех $h \in H(x^*)$.

Нетрудно понять, что для классической задачи на условный экстремум это утверждение переходит в теорему 1.6 .

2.5. Отыскание решений простейших задач

В некоторых случаях изложенная теория позволяет в явном виде найти решение задачи математического программирования. Последовательность действий состоит из следующих этапов:

1) составление функции Лагранжа;
2) составление системы, характеризующей стационарные точки;
3) решение системы;
4) исследование стационарных точек в целях отбора среди них решений.

На этапе 1) задачу необходимо предварительно привести именно к виду (2.6). Так, если дана задача максимизации функции *f*, то ее следует рассматривать как задачу минимизации функции $-f$; если имеется ограничение вида $g_i(x) \geq 0,$ то его надо заменить на ограничение $-g_i(x) \leq 0,$ и т.п. Заметим также, что для конкретной задачи можно составить, вообще говоря, много функций Лагранжа в зависимости от того, какие ограничения считать функциональными, а какие – прямыми, т.е. включенными в множество *P*. Обычно *P* выбирается таким образом, чтобы можно было применить одно из утверждений леммы 2.4. Кроме того, на этапе 1) проверяется, удовлетворяет ли задача какому-либо условию регулярности, и если да, то сразу строится функция Лагранжа вида (2.9).

На этапе 2) выписываются в конкретной форме условия (2.7), (2.8), а также условие допустимости $x^* \in X$ и условия на множители Лагранжа, указанные в теореме 2.3. Так, если множество *P* имеет вид $P = \{x \in \mathbf{R}^n \mid x_j \geq 0,\ j = 1, \ldots, s\},$ то полную систему для определения стационарных точек можно записать в виде

$$x_j \geq 0, \quad \frac{\partial L}{\partial x_j}(x, y_0, y) \geq 0, \quad x_j \frac{\partial L}{\partial x_j}(x, y_0, y) = 0, \quad j = 1, \ldots, s,$$

$$\frac{\partial L}{\partial x_j}(x, y_0, y) = 0, \quad j = s+1, \ldots, n,$$

$$y_i \geq 0, \quad g_i(x) \leq 0, \quad y_i g_i(x) = 0, \quad i = 1, \ldots, k,$$

$$g_i(x) = 0, \quad i = k+1, \ldots, m,$$

$$y_0 \geq 0, \quad (y_0,\ y) \neq 0,$$

причем для регулярных задач последняя строка заменяется условием $y_0 = 1$.

Этап 3) – наиболее трудный, и преодолеть его удается лишь в редких случаях.

Этап 4) для регулярной задачи выпуклого программирования сводится к ссылке на теорему 2.6, а точнее – на теорему 2.4. В общем случае этот этап также непрост. Иногда удается воспользоваться довольно громоздкими условиями оптимальности второго порядка. В некоторых случаях проще провести непосредственное исследование поведения целевой функции в стационарной точке. На данном этапе могут быть полезными разного рода специальные соображения. Так, если известно, что задача на минимум имеет глобальное решение, то им, очевидно, является та стационарная точка, в которой целевая функция принимает наименьшее значение.

2.6. Задачи

1. Найти промежутки выпуклости и вогнутости следующих функций:

1.1. $f(x)=x+\sin x$.

1.2. $f(x)=e^{-x^2}$.

1.3. $f(x)=\sqrt{1+x^2}$.

1.4. $f(x)=\ln(1+x^2)$.

1.5. $f(x)=x\sin(\ln x)\quad (x>0)$.

1.6. $f(x)=x^x\quad (x>0)$.

2. Проверить, что функция является выпуклой

2.1. $f(x)=\sqrt{1+x_1^2+x_2^2}$ на $\mathbf{R}^2$.

2.2. $f(x)=\ln(\sum\limits_{i=1}^{n} e^{x_i})$ на $\mathbf{R}^n$.

3. Найти экстремумы следующих функций:

3.1. $f(x)=2x^2-x^4$.

3.2. $f(x)=\sqrt{2x-x^2}$.

3.3. $f(x)=xe^{-x}$.

3.4. $f(x)=\sqrt{x}\ln x$.

3.5. $f(x)=\cos x+\frac{1}{2}\cos 2x$.

3.6. $f(x)=e^x\sin x$.

3.7. $f(x)=\frac{2x}{1+x^2}$.

3.8. $f(x)=\frac{\ln^2 x}{x}$.

3.9. $f(x)=x\sqrt[3]{x-1}$.

4. Исследовать на экстремум следующие функции нескольких переменных:

4.1. $f(x_1,x_2)=x_1^2+(x_2-1)^2$.

4.2. $f(x_1,x_2)=x_1^2-x_1x_2+x_2^2-2x_1+x_2$.

4.3. $f(x_1,x_2)=x_1^3+x_2^3-3x_1x_2$.

4.4. $f(x_1,x_2)=1-\sqrt{x_1^2+x_2^2}$.

4.5. $f(x_1,x_2)=x_1x_2+\frac{50}{x_1}+\frac{20}{x_2}\quad (x_1>0,\ x_2>0)$

4.6. $f(x_1,x_2)=e^{x_1^2-x_2}(5-2x_1+x_2)$.

4.7. $f(x_1,x_2)=x_1x_2\ln(x_1^2+x_2^2)$.

4.8. $f(x_1,x_2)=(x_1^2+x_2^2)e^{-(x_1^2+x_2^2)}$.

4.9. $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + 12x_1 x_2 + 2x_3$.

4.10. $f(x_1, x_2, x_3) = \sin x_1 + \sin x_2 + \sin x_3 - \sin(x_1 + x_2 + x_3)$

$(0 \le x_1 \le \pi;\ 0 \le x_2 \le \pi;\ 0 \le x_3 \le \pi)$

5. Применяя геометрические построения, найти решения следующих задач математического программирования

5.1. $2x_1 - x_2 \to \max,\ x_1^2 + x_2^2 \le 1,\ x_1 - x_2 \le 0,\ x_1 + x_2 \le -1.$

5.2. $(x_1 - 1)^2 + (x_2 + 1)^2 \to \min,\ x_1^2 + x_2^2 \ge 9,\ x_1 \ge 2,\ x_2 \ge 0.$

5.3. $\ln x_1 - x_2 \to \max,\ x_1 - x_2 \le 2,\ x_1 - x_2^2 \le 4,\ x_1 \ge 0.$

5.4. $x_1 - \sqrt{x_2} \to \min,\ x_1 + x_2 \ge 1,\ x_1^2 + x_2^2 = 1.$

5.5. $x_1 - x_2 \to \max,\ 2x_1 + x_2 \le 2,\ 3x_1 - x_2 \ge -3,\ x_1 - 2x_2 \le 2.$

5.6. $x_2 \to \min,\ x_1^2 + x_2^2 \le 1,\ -x_1 + x_2^2 \le 0,\ x_1 + x_2 \ge 0.$

5.7. $x_1^2 + (x_2 - 1)^2 \to \min,\ x_1^2 + 4x_2^2 \le 4,\ 2x_1^2 + x_2 \ge 2,\ x_1 \ge 2x_2.$

5.8. $x_1 \to \max,\ x_1^2 + x_2^2 \le 1,\ (x_1 - 1)^2 + x_2^2 \ge 1,\ x_1 + x_2 \le 1.$

5.9. $10(x_1 - 3{,}5)^2 + 20(x_2 - 4)^2 \to \min,\ x_1 + x_2 \le 6,\ x_1 - x_2 \le 1,$

$2x_1 + x_2 \ge 6,\ 0{,}5x_1 - x_2 \ge -4,\ x_1 \ge 1,\ x_2 \ge 1.$

5.10. $25(x_1 - 2)^2 + (x_2 - 2)^2 \to \max,\ x_1 + x_2 \ge 2,\ x_1 - x_2 \ge -2,$

$x_1 + x_2 \le 6,\ x_1 - 3x_2 \le 2,\ x_1 \ge 0,\ x_2 \ge 0.$

5.11. $-\ln x_1 + x_2 \to \min,\ x_1 - x_2 \le 2,\ x_1 + x_2^2 \le 4,\ x_1 > 0.$

5.12. $2x_1 - x_2 \to \max,\ x_1 - x_2 \le 0,\ x_1^2 + x_2^2 \le 1,\ x_1 + x_2 \le 1.$

5.13. $x_1 + \sqrt{x_2} \to \min,\ x_1 + x_2 \ge 1,\ x_1^2 + x_2^2 = 1.$

5.14. $\sin(x_1 + x_2) - \sin x_1 - \sin x_2 \to \text{extr},\ x_1 + x_2 \le 2\pi,\ x_1 \ge 0,\ x_2 \ge 0.$

6. Найти решения следующих задач:

6.1. $f(x_1, x_2) = x_1 \to \text{extr},\ P = \mathbf{R}^2,\ x_1^3 - x_2^3 = 0.$

6.2. $f(x_1, x_2) = 4x_1 + 3x_2 \to \text{extr},\ P = \mathbf{R}^2,\ x_1^2 + x_2^2 = 1.$

6.3. $f(x_1, x_2) = x_1^2 + x_2^2 \to \text{extr},\ P = \mathbf{R}^2,\ 3x_1 + 4x_2 = 1.$

6.4. $f(x_1, x_2) = e^{x_1 x_2} \to \text{extr},\ P = \mathbf{R}^2,\ x_1 + x_2 = 1.$

6.5. $f(x_1, x_2) = 5x_1^2 + 4x_1 x_2 + x_2^2 \to \text{extr},\ P = \mathbf{R}^2,\ x_1 + x_2 = 1.$

6.6. $f(x_1, x_2) = 3x_1^2 + 4x_1 x_2 + x_2^2 \to \text{extr},\ P = \mathbf{R}^2,\ x_1 + x_2 = 1.$

6.7. $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 \to \text{extr},\ P = \mathbf{R}^3,$
$2x_1 - x_2 + x_3 \le 5,\ x_1 + x_2 + x_3 = 3.$

6.8. $f(x_1, x_2) = 6x_1 - 2x_1^2 + 2x_1 x_2 - 2x_2^2 \to \text{extr},$
$P = \{x \in \mathbf{R}^2 \mid x_1 \ge 0,\ x_2 \ge 0\},\ 3x_1 + 4x_2 \le 6,\ -x_1 + 4x_2^2 \le 2.$

6.9. $f(x_1, x_2) = 3x_1 x_2 - 2x_2^3 \to \text{extr},$
$P = \{x \in \mathbf{R}^2 \mid x_1 \ge 0,\ x_2 \ge 0\},\ 2x_1 + 5x_2 \ge 20,\ x_1 - 2x_2 = 5.$

6.10. $f(x_1, x_2) = x_1 + 2x_2 \to \text{extr},$
$P = \{x \in \mathbf{R}^2 \mid x_1 \ge 0,\ x_2 \ge 0\},\ 3x_1^2 + x_2^2 \le 1,\ x_1 - 8x_2 \le -1.$

6.11. $f(x_1, x_2, x_3) = x_1 x_2^2 x_3^3 \to \text{extr},$
$P = \mathbf{R}^3,\ x_1 + x_2 + x_3 = 1.$

6.12. $f(x_1,x_2,x_3)=x_1x_2x_3 \to \text{extr}$,

$P=\mathbf{R}^3,\ x_1^2+x_2^2+x_3^2=1,\ x_1+x_2+x_3=0.$

6.13. $f(x_1,x_2,x_3)=x_1x_2x_3 \to \text{extr}$,

$P=\mathbf{R}^3,\ x_1^2+x_2^2+x_3^2=1,\ x_1+x_2+x_3=1.$

6.14. $f(x_1,x_2,x_3)=x_1x_2+x_2x_3 \to \text{extr}$,

$P=\mathbf{R}^3,\ x_1^2+x_2^2=2,\ x_2+x_3=2.$

6.15. $\sum\limits_{i=1}^{n} x_i^2 \to \text{extr},\ P=\mathbf{R}^n,\ \sum\limits_{i=1}^{n} i\,x_i=1.$

6.16. $f(x_1,x_2,x_3)=x_1x_2x_3 \to \text{extr},\ P=\mathbf{R}^3,\ x_1^2+x_2^2+x_3^2\le 1.$

6.17. $\sum\limits_{i=1}^{n} x_i^2 \to \text{extr},\ P=\mathbf{R}^n,\ \sum\limits_{i=1}^{n} x_i^4\le 1.$

6.18. $\sum\limits_{i=1}^{n} x_i^4 \to \text{extr},\ P=\mathbf{R}^n,\ \sum\limits_{i=1}^{n} x_i^2\le 1.$

6.19. $f(x_1,x_2,x_3)=2x_1^2+2x_1+4x_2-3x_3 \to \text{extr}$,

$8x_1-3x_2+3x_3\le 40,\ -2x_1+x_2-x_3=-3,\ x_2\ge 0.$

6.20. $f(x_1,x_2,x_3)=3x_2^2-11x_1-3x_2-x_3 \to \text{extr}$,

$x_1-7x_2+3x_3+7\le 0,\ 5x_1+2x_2-x_3\le 2,\ x_3\ge 0.$

6.21. $f(x_1,x_2)=e^{x_1-x_2} \to \text{extr},\ x_1+x_2\le 1,\ x_1\ge 0,\ x_2\ge 0.$

6.22. $f(x_1,x_2)=x_1 \to \text{extr},\ P=\{x\in\mathbf{R}^2 \mid x_1x_2\ge 1,\ x_1\ge 0,\ x_2\ge 0\},$

$g(x_1,x_2)=x_1+x_2-2\le 0.$

6.23. $\frac{1}{2}ax_1^2+\frac{1}{2}bx_2^2 \to \max,\ x_1^3+x_2^3\le 1,\ x_1^2+x_2^2\ge 1.$

6.24. $f(x_1,x_2)=x_1 \to \inf$, $g_1(x_1,x_2)=-x_1 \le 0$,

$g_2(x_1,x_2)=x_1^2-x_2 \le 0$, $g_3(x_1,x_2)=x_2-2x_1^2 \le 0$.

6.25. $f(x_1,x_2)=x_1 \to \inf$, $x_1^2+x_2^2 \le 1$, $x_1^2-x_2 \le 0$, $x_1+x_2 \le 0$.

6.26. $f(x_1,x_2)=x_1 \to \inf$, $x_1^2+x_2^2 \le 1$, $x_1^3+x_2^3=1$.

6.27. $f(x_1,x_2)=2x_1^{-2}+4x_1^5x_2^{-2} \to \inf$, $x_1>0$, $x_2>0$, $x_1^{-4}x_2^2 \le 1$.

6.28. $f(x_1,x_2,x_3)=x_1+x_2^{-1}x_3^{-0,5} \to \inf$, $x_1>0$, $x_2>0$, $x_3>0$,

$x_1^{-1}x_2+x_1^{-1}x_3 \le 1$

7. Найти точки экстремума функций

$$f(x)=x_1+x_2; \quad f(x)=|x_1|+|x_2-1|; \quad f(x)=x_1^2+2x_2^2$$

на множестве X, если

а) $X=\{x\in \mathbf{R}^2 \mid 0\le x_1 \le 1,\ 0\le x_2 \le 1\}$;

б) $X=\{x\in \mathbf{R}^2 \mid x_1 \ge 0,\ x_2 \ge 0,\ ax_1+bx_2=1\}$ (a, b – неотрицательные числа);

в) $X=\{x\in \mathbf{R}^2 \mid x_1-x_2^2 \ge 0,\ x_1^2+x_2^2 \le 1\}$;

г) $X=\{x\in \mathbf{R}^2 \mid x_1^4-4x_1^3+4x_1^2+1 \le x_2 \le 1\}$.

8. Разделить число 8 на две части так, чтобы произведение их произведения на разность было максимально (задача Тартальи).
9. Найти прямоугольный треугольник наибольшей площади, если сумма длин его катетов равна заданному числу (задача Ферма).
10. На стороне *BC* треугольника *ABC* найти точку *E* так, чтобы параллелограмм *ADEF*, у которого точки *D* и *F* лежат соответственно на сторонах *AB* и *AC*, имел наибольшую площадь (задача Евклида).
11. Найти закон преломления света на плоской границе двух однородных сред, пользуясь принципом Ферма, согласно которому свет из одной точки в другую попадает за кратчайшее время.
12. Вписать в круг прямоугольник максимальной площади.
13. Вписать в круг треугольник максимальной площади.
14. Среди цилиндров, вписанных в шар единичного радиуса, найти цилиндр с максимальным объемом (задача Кеплера).
15. Вписать в единичный шар конус наибольшего объема.

16. Среди конусов, вписанных в шар единичного радиуса, найти конус с максимальной боковой поверхностью.
17. Вписать в единичный шар пространства $\mathbf{R}^n$ цилиндр наибольшего объема (обобщенная задача Кеплера).
18. Вписать в шар пространства $\mathbf{R}^n$ прямоугольный параллелепипед наибольшего объема.
19. Среди треугольников данного периметра найти треугольник наибольшей площади.
20. Среди всех треугольников, имеющих заданный периметр, найти треугольник наибольшей площади (задача Зенодора).
21. Вписать в круг n-угольник наибольшей площади.
22. Вписать в круг треугольник с максимальной суммой квадратов сторон.
23. Найти наибольшую площадь четырёхугольника с заданными сторонами.
24. На данной прямой найти точку C, чтобы сумма расстояний от C до точек A и B была минимальной (задача Герона).
25. Среди всех тетраэдров с данными основанием и высотой найти тетраэдр с наименьшей боковой поверхностью.
26. Среди всех тетраэдров с заданными основанием и площадью боковой поверхности найти тетраэдр наибольшего объема.
27. Среди всех тетраэдров, имеющих заданную площадь поверхности, найти тетраэдр наибольшего объема.
28. Найти расстояние от точки в пространстве $\mathbf{R}^n$ до прямой.
29. Найти расстояние от точки до эллипса. Сколько нормалей можно провести из точки к эллипсу (задача Аполлония).
30. Решить задачу Аполлония для параболы.
31. Решить задачу Аполлония для гиперболы.

§3. ЭЛЕМЕНТЫ ТЕОРИИ ДВОЙСТВЕННОСТИ

Рассматривается задача вида

$$\begin{aligned}
&f(x) \to \min,\\
&g_i(x) \le 0,\ i = 1,\dots,k,\\
&g_i(x) = 0,\ i = k+1,\dots,m, \quad x \in P.
\end{aligned} \tag{3.1}$$

Вектор $y^* \in Q$ называется *вектором Куна-Таккера* задачи (3.1), если

$$f^* \le f(x) + \sum_{i=1}^{m} y_i^* g_i(x) = L(x, y^*) \text{ при всех } x \in P.$$

Теорема 3.1 (о существовании вектора Куна-Таккера). Пусть в задаче (3.1) множество P выпукло, функции $g_{k+1},\dots,g_m$ линейны. Предположим, что дополнительно выполняется хотя бы одно из следующих условий:

1) ограничения-равенства отсутствуют (k=m), и существует точка $\bar{x} \in P$ такая, что $g_i(\bar{x}) < 0$ при всех i =1,...,m;

2) множество P – полиэдр, функции $f, g_1,\dots,g_k$ линейны, множество X непусто.

Тогда вектор Куна-Таккера существует.

Любой задаче математического программирования можно поставить в соответствие так называемую двойственную задачу оптимизации. Между этими задачами обнаруживаются тесные связи, изучение которых составляет предмет теории двойственности.

Двойственной к задаче (3.1) называется задача

$$\varphi(y) \to \max,\ y \in Y, \tag{3.2}$$

где

$$\varphi(y) = \inf_{x \in P} L(x, y) = \inf_{x \in P} (f(x) + \sum_{i=1}^{m} y_i g_i(x)),$$
$$Y = \{y \in Q \mid \varphi(y) > -\infty\}.$$

При этом задача (3.1) называется *прямой*.

Предполагая, что $Y \ne \varnothing$, обозначим через $\varphi^* = \sup_{y \in Y} \varphi(y)$ значение задачи (3.2).

Двойственную задачу (3.2) можно записывать просто в виде

$$\varphi(y) \to \max, \quad y \in Q,$$

допуская тем самым бесконечные значения функции $\varphi(y)$.

Теорема 3.2. В задаче (11.2) множество Y выпукло, функция φ вогнута на Y.

Прямая задача (3.1), очевидно, представима в форме

$$\varphi(x) \to \min, \quad x \in P,$$

где

$$\psi(x) = \sup_{y \in Q} L(x, y) = \begin{cases} f(x), \; \text{если} \;\; x \in X, \\ +\infty, \; \text{если} \;\; x \in P \setminus X. \end{cases}$$

Формально положим $f^* = +\infty$, если $X=\varnothing$, т.е. $\sup\limits_{y \in Q} L(x, y) = +\infty$ при всех $x \in P$, $\varphi^* = -\infty$, если $Y=\varnothing$, т.е. $\inf\limits_{x \in P} L(x, y) = -\infty$ при всех $y \in Q$. Тогда во всяком случае можем записать:

$$f^* = \inf_{x \in P} \sup_{y \in Q} L(x, y), \quad \varphi^* = \sup_{y \in Q} \inf_{x \in P} L(x, y).$$

Таким образом, прямая и двойственная задачи определяются симметрично относительно функции Лагранжа $L(x, y)$ прямой задачи: чтобы получить двойственную, достаточно переставить операции $\inf\limits_{x}$ и $\sup\limits_{y}$ над данной функцией. Отметив этот факт, тем не менее будем использовать в дальнейшем более удобные несимметричные формы записи (3.1) и (3.2) для прямой и двойственной задач.

Следующие теоремы указывают на важные взаимосвязи между прямой и двойственной задачами.

Теорема 3.3 (теорема двойственности). Пусть выполняются предположения теоремы (3.1). Если значение прямой задачи (3.1) конечно ($f^* > -\infty$), в частности, если она имеет решение, то множество решений двойственной задачи (3.2) непусто и совпадает с множеством векторов Куна-Таккера задачи (3.1). При этом справедливо соотношение двойственности

$$f^* = \varphi^*.$$

Теорема 3.4 (теорема Куна-Таккера в форме двойственности). Пусть выполнены предположения теоремы 3.2. Тогда точка $x^* \in X$ является решением задачи (3.1) в том и только том случае, если существует вектор $y^* \in Q$ такой, что справедливо соотношение двойственности $f(x^*) = \varphi(y^*)$, равносильное условиям

$$L(x^*, y^*) = \min_{x \in P} L(x^*, y^*), \quad y_i^* g_i(x^*) = 0, \;\; i = 1, \ldots, k. \tag{3.3}$$

Множество векторов $y^* \in Q$ совпадает с множеством решений двойственной задачи (3.2) или же с множеством векторов Куна-Таккера прямой задачи (3.1).

В случае, когда дополнительно к сделанным предположениям функции $f, g_1,...,g_k$ дифференцируемы в точке x^* условия (3.3) равносильны условиям (2.7), (2.8) при $y_0^* = 1$. Таким образом, теорема 3.4 выступает обобщением теоремы 2.6 на случай недифференцируемых функций. В связи с этим, понятие вектора Куна-Таккера служит обобщением вектора множителей Лагранжа, т.е. вектора $y^* \in Q$, удовлетворяющего условиям (2.7), (2.8) при $y_0^* = 1$.

§4. Решение конечномерных оптимизационных задач с использованием пакета MATLAB

В состав MATLAB входит Optimization Toolbox, предназначенный для решения линейных и нелинейных оптимизационных задач. Optimization Toolbox позволяет решать ряд оптимизационных задач, в которых минимизируемая функция нелинейна, и к линейным ограничениям добавляются нелинейные.

Пусть рассматривается следующая задача: требуется отыскать

$$\min f(x)$$

среди всех векторов *x,* удовлетворяющих системе линейных неравенств

$$Ax \le b,$$

системе линейных уравнений

$$A_{eq} x = b_{eq},$$

двусторонним покомпонентным ограничениям

$$lb \le x \le ub,$$

нелинейной системе неравенствам и равенств

$$c(x) \le 0;\ c_{eq}(x) = 0.$$

Такая постановка включает в себя и задачи математического программирования. Решение поставленной задачи производится при помощи функции `fmincon`, при этом нелинейные ограничения $c(x) \le 0$ и $c_{eq} = 0$ задаются в файл-функции.

Обращение к `fmincon` в достаточно общем случае выглядит следующим образом:

x = fmincon (fun, *x*0, A, b, Aeq, beq, lb, ub, nonlcon, options, P1, P2, ...)

Указание второго дополнительного выходного аргумента позволяет получить значение функции в точке минимума, а третьего – информацию о результате. Если третий аргумент больше нуля, то результат найден с требуемой точностью, ноль – достигнуто максимальное число итераций или количество вызовов исследуемой функции в процессе решения, меньше нуля – решение не найдено. Первый входной аргумент fun является указателем на файл-функцию (или ее именем), вычисляющую минимизируемую функцию $f(x)$, которая может зависеть от нескольких параметров. Значения параметров, в случае их наличия, передаются в последних аргументах P1, P2, ... начиная с 11-ой позиции в списке входных аргументов. Входным аргументом файл-функции fun должен быть вектор, длина которого совпадает с числом переменных, т.е.

компонент вектора x. Неиспользуемые векторы и матрицы ограничений заменяются в списке входных аргументов `fmincon` квадратными скобками (пустым массивом). Начальное приближение к решению указывается в $x0$. Список входных аргументов `fmincon` содержит управляющую структуру options, предназначенную для задания опций вычислительных алгоритмов. Нелинейные ограничения (неравенства и равенства) программируются в файл-функции, указатель на нее (или ее имя) указывается в аргументе nonlcon. Входным аргументом nonlcon является вектор x, соответствующий искомому вектору x, а двумя выходными аргументами – векторы c и ceq левых частей нелинейных ограничений c и c_{eq}. Последние идущие подряд входные аргументы функции `fmincon` могут быть опущены, если они не используются. Например, при отсутствии нелинейных ограничений и параметров применяется следующий вызов `fmincon`:

x = fmincon(fun, *x*0, A, b, Aeq, beq, lb, ub)

Поскольку в данном случае управляющая структура options не задана, то вычисления будут производиться с принятыми по умолчанию опциями, узнать о которых можно при помощи функции `optimset`. Для этого достаточно вызвать ее от строки с именем вычислительной функции Optimization Toolbox, например, выполнение команды

```
>> optimset ('fmincon')
```

приводит к выводу информации о всех настройках алгоритма функции `fmincon`, в том числе и точности 10^{-6}по аргументу и функции. Подробная информация о назначении каждого параметра приведена в справочной системе MATLAB (см. разд. **Optimization Toolbox: Function Reference: Optimization Parameters,** где приведена таблица со всеми параметрами и описаны функции Toolbox).

Пример 1.5.

Рассмотрим следующую задачу :

$$x_1^2 + (x_2 - 1)^2 \to \min,$$

$$x_1^2 + 4x_2^2 \le 4,$$

$$2x_1^2 + x_2 \ge 2,$$

$$x_1 \ge 2x_2.$$

Для решения данной задачи в пакете MATLAB зададим в файлах-функциях целевую функцию и ограничения.

Пример файл-функции минимизируемой функции

```
function f = celevaya(x)
% Целевая функция
f = x(1)^2+(x(2)-1)^2;
```

Пример файл-функции нелинейных ограничений

```
function [c, ceq] = ogranicheniya(x)
% Задание нелинейных ограничений
c(1) = x(1)^2+4*x(2)^2-4;
c(2) = 2-2*x(1)^2-x(2);
% Правая часть ограничений-равенств является пустым массивом,
% поскольку данных ограничений нет
ceq = [];
```

Обращение к fmincon с тремя выходными аргументами позволяет найти не только вектор решения, но и значение целевой функции, а положительное значение flag свидетельствует о том, что решение успешно найдено:

```
clear all  % очищаем память
format long  % формат отображения чисел с 14 знаками
% Задание матрицы и вектора правой части системы неравенств
A = [-1 2];
b = 0;
[x, f, flag]= fmincon(@celevaya, [1 0], A, b, [], [], [], [], @ogranicheniya)
```

```
Warning: Large-scale (trust region) method does not currently solve this type of problem,
 switching to medium-scale (line search).
> In fmincon at 260
Optimization terminated: first-order optimality measure less
 than options.TolFun and maximum constraint violation is less
 than options.TolCon.
Active inequalities (to within options.TolCon = 1e-006):
  lower      upper     ineqlin   ineqnonlin
                          1           2

x =   0.88278221870525   0.44139110935262

f =   1.09134833837245

flag =  1
```

ПРИЛОЖЕНИЕ

Некоторые обозначения, понятия и результаты линейной алгебры, математического анализа и выпуклого анализа

$x \in X$ – элемент x принадлежит множеству X.

$x \notin X$ – элемент x не принадлежит множеству X.

$X \subset Y$ – множество X содержится в множестве Y (это не исключает случая $X = Y$).

$X \cup Y, \bigcup\limits_{i \in I} X_i$ – объединение множеств.

$X \cap Y, \bigcap\limits_{i \in I} X_i$ – пересечение множеств.

$X \setminus Y$ – (теоретико-множественная) разность множеств.

$X \times Y$ – декартово произведение множеств.

$\varnothing$ – пустое множество.

$\{x | P\}$, $\{x \in X | P\}$ – множество элементов и подмножество элементов множества X, обладающих свойством P.

$\{x, y, z\}$ – множество, состоящее из элементов x, y, z.

$\mathbf{R}$ – числовая прямая, множество действительных чисел.

$\mathbf{R}_+$ – множество неотрицательных чисел.

$\mathbf{R}^n$ – n-мерное координатное пространство.

$x = (x_1, \ldots, x_n)$, $x^0 = (x_1^0, \ldots, x_n^0)$, $y^1 = (y_1^1, \ldots, y_n^1)$ – стандартное обозначение элементов из $\mathbf{R}^n$: для обозначения координат данного элемента всегда используется тот же символ с индексом внизу; для обозначения различных элементов чаще всего применяется верхний индекс. Элементы из $\mathbf{R}^n$ называются также точками или векторами. За исключением специально оговоренных мест для нас несущественно, как записывается вектор: в виде строки или в виде столбца.

$0 = (0, \ldots, 0)$ – нулевой элемент в $\mathbf{R}^n$.

e^j – j-й единичный орт в $\mathbf{R}^n$.

$x + y = (x_1 + y_1, \ldots, x_n + y_n)$ – сумма элементов x и y из $\mathbf{R}^n$.

$\lambda x = (\lambda x_1, \ldots, \lambda x_n)$ – произведение элемента x на число λ.

$X \pm Y = \{z \in \mathbf{R}^n \mid z = x \pm y,\ x \in X,\ y \in Y\}$ – (алгебраические) сумма и разность множеств X и Y из $\mathbf{R}^n$.

$\langle x, y \rangle = \sum\limits_{j=1}^{n} x_j y_j$ – скалярное произведение элементов x и y; эти элементы называются ортогональными, если $\langle x, y \rangle = 0$.

$\|x\| = \sqrt{\langle x, y \rangle}$ – норма элемента x.

Для любых элементов x и y справедливо неравенство Коши-Буняковского: $|\langle x,y\rangle| \le ||x||\cdot||y||$.

Для элементов x и y из $\mathbf{R}^n$ пишем: $x \ge y$ и $x > y$, если соответственно $x_j \ge y_j$ и $x_j > y_j$ при всех $j = 1, ..., n$.

$\mathbf{R}_+^n = \{ x \in \mathbf{R}^n \mid x \ge 0 \}$ – неотрицательный ортант в $\mathbf{R}^n$.

Если дана матрица A размера $m \times n$ (с m строками и n столбцами), то, как правило, a_i – ее i-я строка, a^j – j-й столбец, a_{ij} – элемент на пересечении i-й строки и j-го столбца.

A^T – матрица, транспонированная к матрице A.

A^{-1} – матрица, обратная к квадратной матрице A.

$\det A$ – определитель квадратной матрицы A.

rang A – ранг матрицы A, т.е. максимальное число ее линейно независимых строк или столбцов.

E – единичная матрица (на диагонали стоят единицы, остальные элементы – нули).

Если даны матрица A размера $m \times n$, векторы $x \in \mathbf{R}^n$ и $y \in \mathbf{R}^m$, то:

Ax – вектор из $\mathbf{R}^m$ с координатами $(Ax)_i = \langle a_i, x\rangle = \sum\limits_{j=1}^{n} a_{ij}x_j$, $i = 1,\ldots,n$;

yA – вектор из $\mathbf{R}^n$ с координатами $(yA)_j = \langle y, a^j\rangle = \sum\limits_{i=1}^{m} y_i a_{ij}$, $j = 1,\ldots,m$.

Ясно, что $yA = A^T y$ и $\langle y, Ax\rangle = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} y_i a_{ij} x_j = \langle yA, x\rangle$.

AB – произведение матрицы A размера $m \times n$ и матрицы B размера $n \times s$, т.е. матрица размера $m \times s$ с элементами $\langle a_i, b^j\rangle = \sum\limits_{k=1}^{n} a_{ik}b_{kj}$, $i = 1,\ldots,n$, $j = 1,\ldots,m$.

Квадратная матрица A называется *симметрической*, если $A^T = A$.

Симметрическая матрица A размера $n \times n$ называется:

неотрицательно определенной, если $\langle Ax, x\rangle \ge 0$ при всех $x \in \mathbf{R}^n$;

положительно определенной, если $\langle Ax, x\rangle > 0$ при всех ненулевых $x \in \mathbf{R}^n$.

Имеет место следующий **критерий Сильвестра**: симметрическая матрица A неотрицательно (положительно) определена в том и только том случае, если все ее главные (угловые) миноры неотрицательны (положительны). При этом главным минором матрицы A называется определитель матрицы, построенной из элементов матрицы A, стоящих на пересечении строк и столбцов с одинаковыми номерами Если это несколько первых номеров, то главный минор называется угловым.

$U_\varepsilon(a) = \{x \in \mathbf{R}^n \mid ||x - a|| < \varepsilon\}$ – шар радиуса $\varepsilon > 0$ с центром в точке $a \in \mathbf{R}^n$.

$\{x^k\}$ – последовательность точек $x^1,\ x^2, \ldots$.

int $X = \{x \in R^n \mid U_\varepsilon(x) \subset X$ при некотором $\varepsilon > 0\}$ – *внутренность* множества $X \subset \mathbf{R}^n$; элементы из int X называются внутренними точками множества X; множество X называется открытым, если X = int X.

$\overline{X} = \{x \in R^n \mid U_\varepsilon(x) \cap X \neq \varnothing$ при любом $\varepsilon > 0\}$ – *замыкание* множества $X \subset \mathbf{R}^n$; элементы из $\overline{X}$ называются предельными точками множества X; множество X называется замкнутым, если $X = \overline{X}$.

$\partial X = \overline{X} \setminus \operatorname{int} X$ – *граница* множества X; элементы из ∂X называются граничными точками множества X.

Окрестностью точки x (множества X) называется любое множество, содержащее x (соответственно, X) в своей внутренности.

Множество X называется *ограниченным*, если существует число $R > 0$ такое, что $||x|| \le R$ при всех $x \in X$.

Множество $X \subset \mathbf{R}^n$ называется компактом, если оно замкнуто и ограничено.

Непустое множество $X \subset \mathbf{R}^n$ называется *выпуклым*, если $\lambda x^1 + (1-\lambda)x^2 \in X$ при всех $x^1, x^2 \in X$, $\lambda \in [0,1]$, т.е. если X вместе с любыми своими двумя точками x^1 и x^2 содержит соединяющий их отрезок.

Пустое множество считается выпуклым.

Теорема. *Пусть I – любое, конечное или бесконечное, множество индексов, X_i $(i \in I)$ – выпуклые множества. Тогда их пересечение $X = \bigcap\limits_{i \in I} X_i$ выпукло.*

Теорема. *Пусть $X_1, \ldots, X_m$ – выпуклые множества, $\alpha_1,\ldots,\alpha_m$ – любые числа. Тогда множество*

$$\sum_{i=1}^{m} \alpha_i X_i = \{x \mid x = \sum_{i=1}^{m} \alpha_i x^i,\ x^i \in X_i,\ i = 1,\ldots,m\},$$

называемое линейной комбинацией множеств $X_1, \ldots, X_m$, выпукло.

Множество $X \subset \mathbf{R}^n$ называется:

1) *конусом*, если $\lambda x \in X$ при всех $x \in X$, $\lambda \ge 0$, т.е. если X вместе с любой своей точкой x содержит проходящий через нее луч с началом в нуле;

2) *выпуклым конусом*, если $\lambda_1 x^1 + \lambda_2 x^2 \in X$ при всех $x^1, x^2 \in X$, $\lambda_1 \ge 0, \lambda_2 \ge 0$, т.е. если X одновременно является выпуклым множеством и конусом.

Множество $X \subset \mathbf{R}^n$ называется *аффинным*, если $\lambda x^1 + (1-\lambda)x^2 \in X$ при всех $x^1, x^2 \in X$, $\lambda \in \mathbf{R}$, т.е. если X вместе с любыми своими двумя точками x^1 и x^2 содержит проходящую через них прямую.

Пусть $x^1, \ldots, x^m$ – точки из $\mathbf{R}^n$, тогда линейная комбинация $\sum\limits_{i=1}^{m} \lambda_i x^i$

называется:

1) *выпуклой*, если $\lambda_i \geq 0$, $i = 1, ..., m$, $\sum_{i=1}^{m} \lambda_i = 1$;

2) *неотрицательной*, если $\lambda_i \geq 0$, $i = 1, ..., m$;

3) *аффинной*, если $\sum_{i=1}^{m} \lambda_i = 1$.

Теорема. *Выпуклое множество* (*выпуклый конус*, *аффинное множество*) *содержит всевозможные выпуклые* (*неотрицательные*, *аффинные*) *комбинации своих точек*.

Пусть X – произвольное множество в $\mathbf{R}^n$. Пересечение всех выпуклых множеств (выпуклых конусов, аффинных множеств) из $\mathbf{R}^n$, содержащих X, называется его *выпуклой* (*конической*, *аффинной*) *оболочкой* и обозначается через conv X (cone X, aff X).

Точка x множества $X \subset \mathbf{R}^n$ называется его *относительно внутренней точкой*, если $U_\varepsilon(x) \cap \operatorname{aff} X \subset X$ при некотором $\varepsilon > 0$. Совокупность всех относительно внутренних точек множества X называется его *относительной внутренностью* и обозначается ri X.

Для любого $X \subset \mathbf{R}^n$ его выпуклая оболочка conv X непуста (так как X содержится, по крайней мере, в $\mathbf{R}^n$, являющемся выпуклым множеством) и выпукла. При этом если множество Y выпукло и содержит X, то, по определению, conv X $\in$ Y. Другими словами, conv X есть наименьшее (по включению) выпуклое множество, содержащее X. Ясно, что X выпукло в том и только том случае, если conv X = X.

Аналогичные замечания можно высказать о соответствии понятий выпуклого конуса и конической оболочки, аффинного множества и аффинной оболочки.

Теорема. *Выпуклая* (*коническая*, *аффинная*) *оболочка произвольного множества X совпадает с множеством, всевозможных выпуклых* (*неотрицательных*, *аффинных*) *комбинаций точек из X*.

Теорема. *Если точка x является неотрицательной комбинацией точек* x^1, ..., x^m, *не равных нулю одновременно, то x можно представить в виде неотрицательной комбинации линейно независимой подсистемы этих точек*.

Следствие. *Пусть X – произвольное множество в* $\mathbf{R}^n$. *Тогда любую точку из* cone *X можно представить в виде неотрицательной комбинации не более n точек из X*.

Теорема (Каратеодори). *Пусть X – произвольное множество в* $\mathbf{R}^n$. *Тогда любую точку из* conv *X можно представить в виде выпуклой комбинации не более n*+1 *точек из X*.

Выпуклая (коническая) оболочка множества, состоящего из конечного числа точек, называется *выпуклым многогранником* (*многогранным конусом*), натянутым на эти точки.

Выпуклый многогранник имеет вид

$$X = \mathrm{conv}\{x^1,\dots,x^m\} = \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x^i,\ \lambda_i \ge 0,\ i = 1,\dots,m,\ \sum_{i=1}^{m} \lambda_i = 1\},$$

а многогранный конус – вид

$$X = \mathrm{cone}\{x^1,\dots,x^m\} = \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x^i,\ \lambda_i \ge 0,\ i = 1,\dots,m\}.$$

Точка x выпуклого множества X называется *крайней* (иногда говорят – *угловой*, или *экстремальной*), если ее нельзя представить в виде $x = \lambda x^1 + (1-\lambda)x^2$, где $x^1, x^2 \in X,\ x^1 \ne x^2,\ 0 < \lambda < 1$.

Совокупность всех крайних точек множества X обозначим через $E(X)$.

Говоря геометрическим языком, точка x – крайняя в X, если ее нельзя поместить внутрь отрезка, концы которого лежат в X.

Теорема. Пусть X – выпуклый компакт в $\mathbf{R}^n$. Тогда $X = \mathrm{conv}E(X)$, т.е. X совпадает с выпуклой оболочкой множества своих крайних точек.

Множество $X \subset \mathbf{R}^n$ называется *полиэдром*, если оно представляет собой множество решений некоторой системы конечного числа линейных неравенств, т.е. является пересечением конечного числа полупространств:

$$X = \{x \in \mathbf{R}^n \mid Ax \le b\} = \{x \in \mathbf{R}^n \mid \langle a_i, x\rangle \le b_i, i = 1,\dots,m\}, \qquad (\text{П.}1)$$

где A – матрица размера $\mathrm{m}\times\mathrm{n}$ со строками $a_1,\dots,a_m$, $b = (b_1,\dots,b_m) \in \mathbf{R}^m$.

Полиэдр вида (П.1) при $b = 0$ называется *однородным*.

Множество $X \subset \mathbf{R}^n$ называется *многогранным*, если оно представляет собой сумму выпуклого многогранника и многогранного конуса, т.е.

$$X = \mathrm{conv}\{x^1,\dots,x^r\} + \mathrm{cone}\{x^{r+1},\dots,x^s\} =$$

$$= \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{s} \lambda_i x^i,\ \lambda_i \ge 0,\ i = 1,\dots,s,\ \sum_{i=1}^{r} \lambda_i = 1\},$$

где $x^1,\dots,x^s$ – некоторые точки в $\mathbf{R}^n$.

Теорема. *Любой полиэдр имеет не более конечного числа крайних точек*.

Теорема. 1) *Любой ограниченный полиэдр является выпуклым многогранником*; 2) *любой однородный полиэдр – многогранным конусом*; 3) *любой полиэдр – многогранным множеством*.

Теорема. 1) *Любой выпуклый многогранник является ограниченным полиэдром*; 2) *любой многогранный конус – однородным полиэдром*; 3) *любое многогранное множество – полиэдром*.

$f: X \to Y$ – *функция* (отображение) с областью определения X и областью значений Y. Эта функция называется числовой, если $Y \subset \mathbf{R}$, и функцией числового аргумента, если $X \subset \mathbf{R}$.

Говорят, что $a \in \mathbf{R}^m$ является *пределом* {предельным значением) функции $f: X \to Y$ ($X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$) в точке $x^0 \in X$, если для любого числа $\varepsilon > 0$ существует число $\delta > 0$ такое, что $f(x) \in U_\varepsilon(a)$ для всех $x \in U_\delta(x^0) \cap X$;

при этом пишут: $a = \lim\limits_{x \to x^0} f(x)$ или $f(x) \to a$ при $x \to x^0$.

Функция $f: X \to Y$ называется *непрерывной в точке* $x^0 \in X$, если $\lim\limits_{x \to x^0} f(x) = f(x^0)$, и непрерывной на множестве $A \subset X$, если это равенство справедливо для всех $x^0 \in A$.

$f(A),\ f^{-1}(B)$ – *образ* множества $A \subset X$ и *прообраз* множества $B \subset Y$ при отображении $f: X \to Y$.

$\inf\limits_{x \in X} f(x)$, $\sup\limits_{x \in X} f(x)$ – *точные нижняя* и *верхняя грани* числовой функции *f* на множестве *X*.

$\min\limits_{x \in X} f(x)$, $\max\limits_{x \in X} f(x)$ – *минимальное* и *максимальное* значения числовой функции *f* на множестве *X*, т.е. величины $\min\limits_{x \in X} f(x)$, $\max\limits_{x \in X} f(x)$ в предположении, что они достигаются на некоторых элементах из *X*.

Ниже, говоря о дифференциальных характеристиках функции *f* в точке x^*, подразумеваем, что $x^* \in \mathbf{R}^{\boldsymbol{n}}$ и f – числовая функция, определенная в некоторой окрестности x^*.

$\dfrac{\partial f}{\partial x_j}(x^*) = \lim\limits_{\alpha \to 0} \dfrac{f(x^* + \alpha e^j) - f(x^*)}{\alpha}$ – *частная производная* функции f в точке x^* по аргументу x_j.

$f'(x^*) = \left(\dfrac{\partial f}{\partial x_1}(x^*), \ldots, \dfrac{\partial f}{\partial x_n}(x^*) \right)$ – *градиент* (вектор частных производных) функции *f* в точке x^*.

Функция f называется *дифференцируемой в точке* x^*, если градиент $f'(x^*)$ существует и при всех достаточно малых $h \in \mathbf{R}^{\boldsymbol{n}}$ справедлива формула

$$f(x^* + h) = f(x^*) + \langle f'(x^*), h \rangle + o(\| h \|).$$

Здесь, как и всюду далее, $o(\| h \|)$ – некоторая числовая функция числового аргумента, удовлетворяющая условию $\lim\limits_{\alpha \to 0} \dfrac{o(\alpha)}{\alpha} = 0$.

Функция f называется *непрерывно дифференцируемой в точке* x^*, если градиент $f'(x)$ существует в некоторой окрестности точки x^* и непрерывен в самой точке x^*.

Говорят, что функция f *дифференцируема* (*непрерывно дифференцируема*) *на множестве* $X \subset \mathbf{R}^{\boldsymbol{n}}$, если она дифференцируема (непрерывно дифференцируема) в каждой точке из *X*. При этом, как следует из предыдущего, функция f фактически должна быть определена в некоторой окрестности множества *X*.

Для любого вектора $h \in \mathbf{R}^n$ используем обозначение

$$f'(x^*, h) = \lim_{\alpha \to 0+} \frac{f(x^* + \alpha e^j) - f(x^*)}{\alpha};$$

если $\|h\| = 1$, то величина $f'(x^*, h)$ называется *производной* функции f в точке $x^* \in \mathbf{R}^n$ *по направлению вектора* h. Функция f называется дифференцируемой в точке x^* по направлению вектора h, если величина $f'(x^*, h)$ существует и конечна.

Если функция f дифференцируема в точке x^*, то f дифференцируема в x^* по направлению любого вектора h, причем

$$f'(x^*, h) = \langle f'(x^*), h \rangle .$$

$\dfrac{\partial f}{\partial x_i \partial x_j}(x^*)$ – вторая частная производная функции f в точке $x^* \in \mathbf{R}^n$ по аргументам x_i, x_j (частная производная функции $\dfrac{\partial f}{\partial x_i}(x)$ в точке x^* по аргументу x_j).

$f''(x^*) = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x^*) \right)_{i,j=1,\dots,n}$ – *гессиан* (матрица вторых частных производных) функции f в точке x^*.

Функция f называется дважды дифференцируемой в точке x^*, если матрица $f''(x^*)$ существует, является симметрической матрицей и при всех достаточно малых $h \in \mathbf{R}^n$ справедлива формула

$$f(x^* + h) = f(x^*) + \langle f'(x^*), h \rangle + \frac{1}{2} \langle f''(x^*) h, h \rangle + o(\| h \|^2).$$

Функция f называется дважды непрерывно дифференцируемой в точке $x^* \in \mathbf{R}^n$, если матрица $f''(x)$ существует в некоторой окрестности точки x^* и непрерывна в самой точке x^*.

Если функция f дважды непрерывно дифференцируема в точке x^*, то она дважды дифференцируема в этой точке.

Если функция f дважды дифференцируема на выпуклом множестве $X \subset \mathbf{R}^n$ (т.е. в каждой точке из X), то для любых точек $x, x^* \in X$ справедлива **формула Тейлора** (с остаточным членом в формуле Лагранжа)

$$f(x) - f(x^*) = \langle f'(x^*), h \rangle + \frac{1}{2} \langle f''(x^* + \alpha h) h, h \rangle,$$

где $h = x - x^*$, а α – некоторое число из интервала (0, 1).

Функция f, определенная на выпуклом множестве $X \subset \mathbf{R}^n$, называется *выпуклой*, если

$$f(\lambda x^1+(1-\lambda)x^2)\le \lambda f(x^1)+(1-\lambda)f(x^2) \qquad (П.2)$$

при всех $x^1, x^2 \in X$, $\lambda\in[0,1]$.

Если при всех $x^1, x^2 \in X$, $x^1 \ne x^2$, $\lambda\in(0,1)$ неравенство (П.2) выполняется как строгое, то f называется *строго выпуклой* на *X*.

Функция f, определенная на выпуклом множестве $X\subset \mathbf{R}^n$, называется *сильно выпуклой* с константой $\theta>0$ на *X*, если

$$f(\lambda x^1+(1-\lambda)x^2)\le \lambda f(x^1)+(1-\lambda)f(x^2)-\theta\| x^1-x^2\|^2 \qquad (П.3)$$

при всех $x^1, x^2 \in X$, $\lambda\in[0,1]$.

Простейшим примером сильно выпуклой функции является функция $f(x)=\|x\|^2$ на $\mathbf{R}^n$. Легко проверить, что для нее неравенство (П.3) выполняется как равенство при $\theta=1$. Сильно выпуклая функция, очевидно, строго выпукла. Обратное неверно. Например, функция $f(x)=x^4$ (геометрически «похожая» на сильно выпуклую функцию $f(x)=x^2$) строго, но не сильно выпукла на $\mathbf{R}$.

Функция f, определенная на выпуклом множестве $X\subset \mathbf{R}^n$, называется (*строго*, *сильно*) *вогнутой* на *X*, если функция $-f$ (строго, сильно) выпукла на *X*.

Функцию вида

$$f(x)=\langle a,\, x\rangle+b,$$

где $a\in\mathbf{R}^n$, $b\in\mathbf{R}$, будем называть *линейной*. Линейная функция одновременно выпукла и вогнута на $\mathbf{R}^n$, но не строго.

Функция вида

$$f(x)=\frac{1}{2}\langle Ax,x\rangle+\langle b,x\rangle+c,$$

где A – симметрическая матрица размера $n\times n$, $b\in\mathbf{R}^n$, $c\in\mathbf{R}$, называется квадратичной. Для нее

$$f'(x)=Ax+b,\ \ f''(x)=A.$$

Теорема (неравенство Йенсена). *Пусть f – выпуклая функция на выпуклом множестве X. Тогда*

$$f\left(\sum_{i=1}^m \lambda_i x^i\right)\le \sum_{i=1}^m \lambda_i f(x^i)$$

при всех $m=1,\,2,\ldots;\, x^i\in X,\ \lambda_i\ge 0,\ i=1,\ldots,m,\ \sum_{i=1}^m \lambda_i=1.$

Теорема. *Пусть* $f_1,\ldots,f_m$ *– выпуклые функции на выпуклом множестве X,* $\alpha_1,\ldots,\alpha_m$ *– неотрицательные числа. Тогда функция* $f(x)=\sum_{i=1}^m \alpha_i f_i(x)$ *выпукла* *на X.*

Теорема. *Пусть* $g_1,\dots,g_m$ *– выпуклые функции на выпуклом множестве* $X \subset \mathbf{R}^n$, $g=(g_1,\dots,g_m)$ *– образованная из них вектор-функция,* φ *– монотонно неубывающая выпуклая функция на выпуклом множестве* $U \subset \mathbf{R}^m$, *причем* $g(X) \subset U$. *Тогда функция* $f(x)=\varphi(g(x))$ *выпукла на X.*

Теорема. *Пусть* φ *– выпуклая функция на выпуклом множестве* $U \subset \mathbf{R}^m$, *A – матрица размера* $m \times n$, $b \in \mathbf{R}^m$, *причем множество* $X=\{x \in \mathbf{R}^n \mid Ax+b \in X\}$ *непусто. Тогда функция* $f(x)=\varphi(Ax+b)$ *выпукла на X.*

Теорема. *Пусть f – дифференцируемая функция на выпуклом множестве* $X \subset \mathbf{R}^n$. *Тогда f сильно выпукла с константой* $\theta \ge 0$ *на X в том и только том случае, если*

$$f(x)-f(x^*) \ge \langle f'(x^*), x-x^* \rangle + \theta \| x-x^* \|^2 \text{ при всех } x,\ x^* \in X.$$

Теорема. *Пусть f – выпуклая функция на выпуклом множестве* $X \subset \mathbf{R}^n$, *дифференцируемая в точке* $x^* \in X$. *Тогда*

$$f(x)-f(x^*) \ge \langle f'(x^*), x-x^* \rangle \text{ при всех } x \in X.$$

Теорема. *Пусть f – непрерывно дифференцируемая функция на выпуклом множестве* $X \subset \mathbf{R}^n$. *Тогда f сильно выпукла с константой* $\theta \ge 0$ *на X в том и только том случае, если*

$$\langle f'(x)-f'(x^*), x-x^* \rangle \ge 2\theta \| x-x^* \|^2 \text{ при всех } x,\ x^* \in X.$$

Теорема (критерий выпуклости в терминах вторых производных). *Пусть f – дважды непрерывно дифференцируемая функция на выпуклом множестве* $X \subset \mathbf{R}^n$, *причем* $\operatorname{int} X \ne \varnothing$. *Тогда f сильно выпукла с константой* $\theta \ge 0$ *на X в том и только том случае, если*

$$\langle f''(x^*)h, h \rangle \ge 2\theta \| h \|^2 \quad \textit{при всех} \quad x^* \in X, \quad h \in \mathbf{R}^n.$$

При $\theta = 0$ данное условие говорит о том, что в любой точке $x^* \in X$ матрица вторых производных $f''(x^*)$ неотрицательно определена. При $\theta > 0$ оно означает несколько большее, чем просто положительная определенность этих матриц (заметим, что θ не зависит от x^*).

Следствие. *Пусть A – симметрическая матрица размера* $n \times n$, $b \in \mathbf{R}^n$. *Тогда квадратичная функция* $f(x)=\langle Ax, x \rangle + \langle b, x \rangle$ *выпукла (сильно выпукла) на* $\mathbf{R}^n$ *в том и только том случае, если матрица A неотрицательно (положительно) определена.*

Теорема. *Пусть f – выпуклая функция на выпуклом множестве* $X \subset \mathbf{R}^n$. *Тогда f непрерывна в любой точке* $x^* \in \operatorname{ri} X$.

Литература

1. Алексеев, В. М. Сборник задач по оптимизации. Теория. Примеры. Задачи: Учеб. пособие / В. М. Алексеев, Э. М. Галеев, В. М. Тихомиров. – Москва: ФИЗМАТЛИТ, 2011. – 256 с.

2. Аттетков, А. В. Методы оптимизации: учебное пособие / А. В. Аттетков, В. С. Зарубин, А. Н. Канатников. – Москва: РИОР: ИНФРА-М, 2019. – 270 с.

3. Ашманов, С. А. Теория оптимизации в задачах и упражнениях: учебное пособие / С. А. Ашманов, А. В. Тимохов. – Санкт-Петербург: Лань, 2012. – 448 с.

4. Балдин, К. В. Методы оптимальных решений: учебник / К. В. Балдин, В. Н. Башлыков, А. В. Рукосуев; под общ. ред. К. В. Балдина. – Москва: ФЛИНТА, 2020. – 323 с.

5. Пантелеев, А. В. Методы оптимизации. Практический курс: учебное пособие / А. В. Пантелеев, Т. А. Летова. – Москва: Логос, 2020. – 424 с.

6. Сдвижков, О. А. Практикум по методам оптимизации: учебное пособие / О. А. Сдвижков. – Москва: Вузовский учебник: ИНФРА-М, 2020. – 231 с.

7. Сухарев, А. Г. Методы оптимизации: учебник и практикум для бакалавриата и магистратуры / А. Г. Сухарев, А. В. Тимохов, В. В. Федоров. – М. : Издательство Юрайт, 2014 – 367 с.

Юрий Алексеевич **Кузнецов**
Алексей Валерьевич **Семенов**

**МЕТОДЫ ОПТИМИЗАЦИИ:
ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ**

Учебно-методическое пособие

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный
университет им. Н. И. Лобачевского»
603950, Нижний Новгород, пр. Гагарина, 23.