ЛЕКЦИИ

Лекция 1

Раздел I. ТЕОРИЯ ОПТИМИЗАЦИИ

1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ ОПТИМИЗАЦИИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Постановка задачи поиска минимума функций содержит:

- *целевую функцию* $f(x)$, где $x = (x_1, ..., x_n)^T$, определенную на n-мерном евклидовом пространстве R^n. Ее значения характеризуют степень достижения цели, во имя которой поставлена или решается задача;
- *множество допустимых решений* $X \subseteq R^n$, среди элементов которого осуществляется поиск.

Требуется найти такой вектор x^* из множества допустимых решений, которому соответствует минимальное значение целевой функции на этом множестве:

$$f(x^*) = \min_{x \in X} f(x). \tag{1.1}$$

З а м е ч а н и я.

1. Задача поиска максимума функции $f(x)$ сводится к задаче поиска минимума путем замены знака перед функцией на противоположный (рис. 1):

$$f(x^*) = \max_{x \in X} f(x) = -\min_{x \in X} \left[-f(x)\right].$$

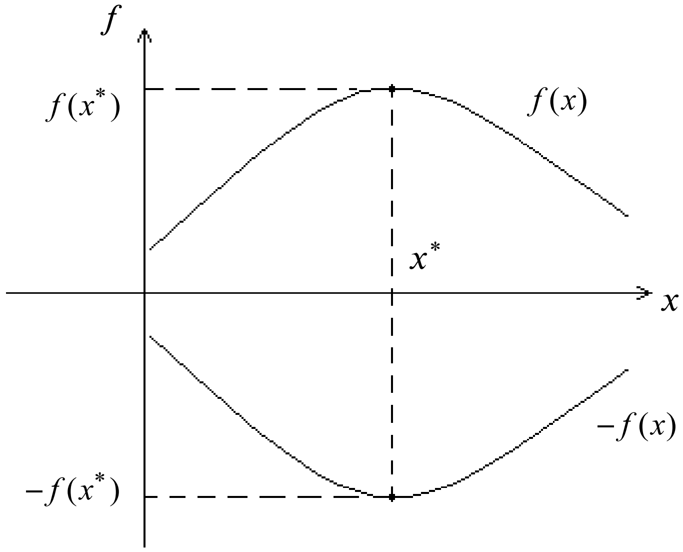


Рис. 1

2. Задача поиска минимума и максимума целевой функции $f(x)$ называется задачей поиска *экстремума*: $f(x^*) = \underset{x \in X}{\mathrm{extr}} f(x)$.

3. Если множество допустимых решений X задается ограничениями (условиями), накладываемыми на вектор x, то решается задача поиска *условного экстремума*. Если $X = R^n$, т.е. ограничения (условия) на вектор x отсутствуют, решается задача поиска *безусловного экстремума*.

4. Решением задачи поиска экстремума является пара $(x^*, f(x^*))$, включающая точку x^* и значение целевой функции в ней.

5. Множество точек минимума (максимума) целевой функции $f(x)$ на множестве X обозначим X^*. Оно может содержать конечное число точек (в том числе одну), бесконечное число точек или быть пустым.

Определение 1.1. Точка $x^* \in X$ называется точкой *глобального* (*абсолютного*) *минимума* функции $f(x)$ на множестве X, если функция достигает в этой точке своего наименьшего значения, т.е.

$$f(x^*) \le f(x) \quad \forall x \in X .$$

Определение 1.2. Точка $x^* \in X$ называется точкой *локального* (*относительного*) *минимума* функции $f(x)$ на множестве допустимых решений X, если существует $\varepsilon > 0$, такое, что если $x \in X$ и $\|x - x^*\| < \varepsilon$, то $f(x^*) \le f(x)$. Здесь $\|x\| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$ – евклидова норма вектора x.

З а м е ч а н и я.

1. В определении 1.1 точка x^* сравнивается по величине функции со всеми точками из множества допустимых решений X, а в определении 1.2 – только с принадлежащими ее ε-окрестности (рис. 2).

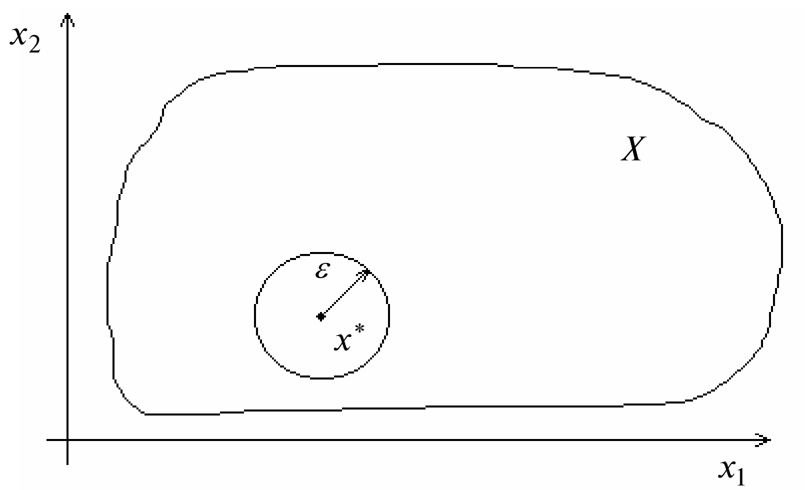


Рис. 2

2. Если в определениях 1.1 и 1.2 знак неравенства $\le$ заменить на $\ge$, то получим определения *глобального* (*абсолютного*) и *локального* (*относительного*) *максимумов*.

3. Глобальный экстремум всегда является одновременно локальным, но не наоборот.

Определение 1.3. *Поверхностью уровня* функции $f(x)$ называется множество точек, в которых функция принимает постоянное значение, т.е. $f(x) = \text{const}$. Если $n = 2$, поверхность уровня изображается *линией уровня* на плоскости R^2.

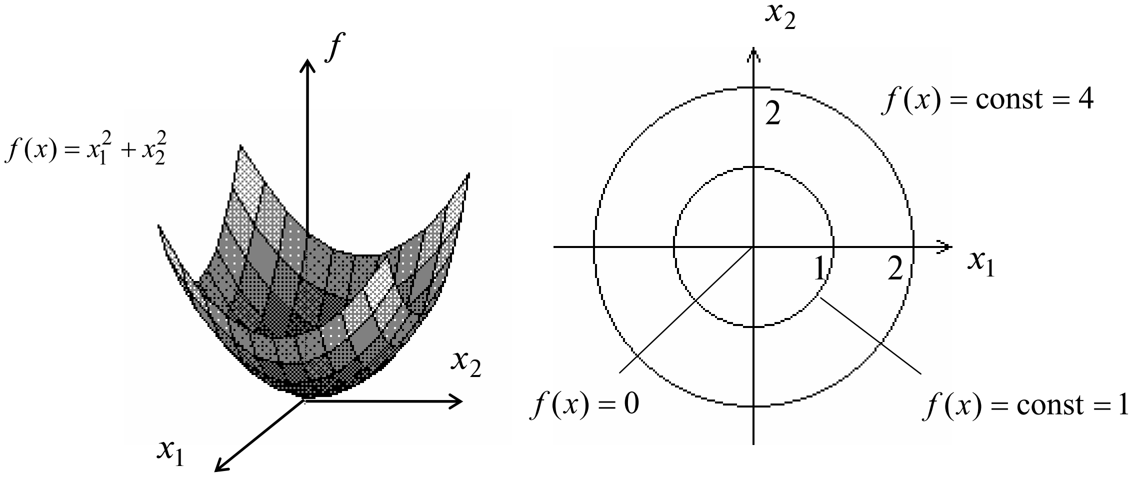
Пример. Построить линии уровня функций:

а) $f(x) = x_1^2 + x_2^2$; б) $f(x) = \dfrac{x_1^2}{4} + x_2^2$;

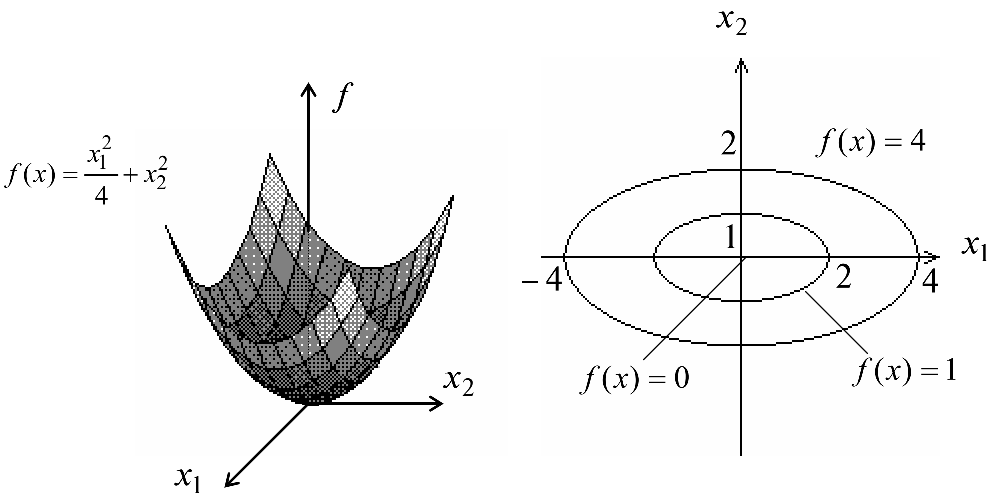
☐ Уравнения линий уровня имеют следующий вид:

а) $f(x) = x_1^2 + x_2^2 = \text{const} = r^2$ – уравнение окружностей с центром в точке $(0,0)^T$ и радиусом, равным r (рис. 3, *а*);

б) $f(x) = \dfrac{x_1^2}{4} + x_2^2 = \text{const}$ – уравнение эллипса. Если $\text{const} = 1$, то $a = 2$ и $b = 1$ – большая и малая полуоси (рис. 3, *б*) . ■



а)



б)

Рис. 3

Определение 1.4. *Градиентом* $\nabla f(x)$ непрерывно дифференцируемой функции $f(x)$ в точке x называется вектор-столбец, элементами которого являются частные производные первого порядка, вычисленные в данной точке:

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix}.$$

Градиент функции направлен по нормали к поверхности уровня (см. определение 1.3), т.е. перпендикулярно к касательной плоскости, проведенной в точке x, в сторону наибольшего возрастания функции в данной точке.

Определение 1.5. *Матрицей Гессе* $H(x)$ дважды непрерывно дифференцируемой в точке x функции $f(x)$ называется матрица частных производных второго порядка, вычисленных в данной точке:

$$H(x) = \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix},$$

где $h_{ij} = \dfrac{\partial^2 f(x)}{\partial x_i \, \partial x_j}$, $i, j = 1, \ldots, n$.

З а м е ч а н и я.

1. Матрица Гессе является симметрической размеров $(n \times n)$.

2. Вместе с градиентом можно определить вектор *антиградиента*, равный по модулю вектору градиента, но противоположный по направлению. Он указывает в сторону наибольшего убывания функции в данной точке.

3. С помощью градиента и матрицы Гессе, используя разложение в ряд Тейлора, приращение функции $f(x)$ в точке x может быть записано в форме

$$\Delta f(x) = f(x + \Delta x) - f(x) = \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T H(x) \Delta x + o(\|\Delta x\|^2), \tag{1.2}$$

где $o(\|\Delta x\|^2)$ – сумма всех членов разложения, имеющих порядок выше второго; $\Delta x^T H(x) \Delta x$ – квадратичная форма.

Пример. Для функции $f(x) = x_1^2 + x_2^4$ вычислить градиент и найти матрицу Гессе в точках $x^0 = (0,0)^T$, $x^1 = (1,1)^T$.

☐ Согласно определениям 1.4 и 1.5 имеем:

$$\nabla f(x) = \left(2x_1,\, 4x_2^3\right)^T,\ H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}; \quad \nabla f(x^0) = (0,\, 0)^T,\ H(x^0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix};$$

$$\nabla f(x^1) = (2,\, 4)^T,\ H(x^1) = \begin{pmatrix} 2 & 0 \\ 0 & 12 \end{pmatrix}. \ ■$$

Определение 1.6. Квадратичная форма $\Delta x^T H(x) \Delta x$ (а также соответствующая матрица Гессе $H(x)$) называется:

- *положительно определенной* $\left(H(x) > 0\right)$, если для любого ненулевого Δx выполняется неравенство $\Delta x^T H(x) \Delta x > 0$;
- *отрицательно определенной* $\left(H(x) < 0\right)$, если для любого ненулевого Δx выполняется неравенство $\Delta x^T H(x) \Delta x < 0$;
- *положительно полуопределенной* $\left(H(x) \ge 0\right)$, если для любого Δx выполняется неравенство $\Delta x^T H(x) \Delta x \ge 0$ и имеется отличный от нуля вектор Δx, для которого $\Delta x^T H(x) \Delta x = 0$;
- *отрицательно полуопределенной* $\left(H(x) \le 0\right)$, если для любого Δx выполняется неравенство $\Delta x^T H(x) \Delta x \le 0$ и имеется отличный от нуля вектор Δx, для которого $\Delta x^T H(x) \Delta x = 0$;
- *неопределенной* ($H(x) \gtrless 0$) , если существуют такие векторы Δx, $\Delta \tilde{x}$, что выполняются неравенства $\Delta x^T H(x) \Delta x > 0$, $\Delta \tilde{x}^T H(x) \Delta \tilde{x} < 0$;
- *тождественно равной нулю* $\left(H(x) \equiv 0\right)$, если для любого Δx выполняется равенство $\Delta x^T H(x) \Delta x = 0$.

2. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ БЕЗУСЛОВНОГО ЭКСТРЕМУМА

Постановка задачи

Дана дважды непрерывно дифференцируемая функция $f(x)$, определенная на множестве $X = R^n$.

Требуется исследовать функцию $f(x)$ на экстремум, т.е. определить точки $x^* \in R^n$ ее локальных минимумов и максимумов на R^n:

$$f(x^*) = \min_{x \in R^n} f(x); \qquad f(x^*) = \max_{x \in R^n} f(x). \tag{2.1}$$

Стратегия решения задачи

Находятся точки x^* локальных экстремумов с помощью необходимых условий первого и второго порядка (порядок условий определяется порядком используемых производных), а также достаточных условий безусловного локального экстремума. Вычисляются значения $f(x^*)$ функции в найденных точках локальных экстремумов.

Утверждение 2.1 (необходимые условия экстремума первого порядка).

Пусть $x^* \in R^n$ *есть точка локального минимума (максимума) функции* $f(x)$ *на множестве* R^n *и* $f(x)$ *дифференцируема в точке* x^*. *Тогда градиент функции* $f(x)$ *в точке* x^* *равен нулю, т.е.*

$$\nabla f(x^*) = 0 \tag{2.2}$$

или

$$\frac{\partial f(x^*)}{\partial x_i} = 0, \quad i = 1, \ldots, n. \tag{2.3}$$

Определение 2.1. Точки x^*, удовлетворяющие условию (2.2) или (2.3), называются *стационарными.*

Утверждение 2.2 (необходимые условия экстремума второго порядка).

Пусть точка x^* *есть точка локального минимума (максимума) функции* $f(x)$ *на множестве* R^n *и функция* $f(x)$ *дважды дифференцируема в этой точке. Тогда матрица Гессе* $H(x^*)$ *функции* $f(x)$, *вычисленная в точке* x^*, *является положительно полуопределенной (отрицательно полуопределенной), т.е.*

$$H(x^*) \ge 0, \tag{2.4}$$

$$(H(x^*) \le 0). \tag{2.5}$$

Утверждение 2.3 (достаточные условия экстремума).

Пусть функция $f(x)$ *в точке* $x^* \in R^n$ *дважды дифференцируема, ее градиент равен нулю, а матрица Гессе является положительно определенной (отрицательно определенной), т.е.*

$$\nabla f(x^*) = 0 \ \text{ и } \ H(x^*) > 0, \tag{2.6}$$

$$(H(x^*) < 0). \tag{2.7}$$

Тогда точка x^* *есть точка локального минимума (максимума) функции* $f(x)$ *на множестве* R^n.

Определение 2.2. Рассмотрим определитель матрицы Гессе $H(x^*)$, вычисленной в стационарной точке

$$\det H(x^*) = \begin{vmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{vmatrix}.$$

1. Определители $\Delta_1 = h_{11}$, $\Delta_2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}$,..., $\Delta_n = \begin{vmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nn} \end{vmatrix}$ называются *угловыми минорами*.

2. Определители m-го порядка ($m \le n$), получающиеся из определителя матрицы $H(x^*)$ вычеркиванием каких-либо ($n - m$) строк и ($n - m$) столбцов с одними и теми же номерами, называются *главными минорами.*

Для проверки выполнения достаточных условий экстремума и необходимых условий второго порядка используются два способа.

Первый способ (с помощью угловых и главных миноров – табл. 1).

- **Критерий проверки достаточных условий экстремума** (критерий Сильвестра).

Для того чтобы матрица Гессе $H(x^*)$ была положительно определенной ($H(x^*) > 0$) и точка x^* являлась точкой локального минимума, необходимо и достаточно, чтобы знаки угловых миноров были строго положительны:

$$\Delta_1 > 0, \quad \Delta_2 > 0, ..., \quad \Delta_n > 0. \tag{2.8}$$

Для того чтобы матрица Гессе $H(x^*)$ была отрицательно определенной ($H(x^*) < 0$) и точка x^* являлась точкой локального максимума, необходимо и достаточно, чтобы знаки угловых миноров чередовались, начиная с отрицательного:

$$\Delta_1 < 0, \ \Delta_2 > 0, \ \Delta_3 < 0, ..., \ (-1)^n \Delta_n > 0. \tag{2.9}$$

- **Критерий проверки необходимых условий экстремума второго порядка.**

1. *Для того чтобы матрица Гессе $H(x^*)$ была положительно полуопределенной ($H(x^*) \ge 0$) и точка x^* может быть являлась точкой локального минимума, необходимо и достаточно, чтобы все главные миноры определителя матрицы Гессе были неотрицательны.*

2. *Для того чтобы матрица Гессе $H(x^*)$ была отрицательно полуопределенной ($H(x^*) \le 0$) и точка x^* может быть являлась точкой локального максимума, необходимо и достаточно, чтобы все главные миноры четного порядка были неотрицательны, а все главные миноры нечетного порядка – неположительны.*

Таблица 1

№ п/п	$\nabla f(x^*)$	$H(x^*)$	Первый способ	Тип стационарной точки x^*
1	0	>0	$\Delta_1 > 0, \Delta_2 > 0, ..., \Delta_n > 0$	Локальный минимум
2	0	<0	$\Delta_1 < 0, \Delta_2 > 0, ..., (-1)^n \Delta_n > 0$	Локальный максимум
3	0	≥ 0	Все главные миноры определителя матрицы $H(x^*)$ неотрицательны	Может быть локальный минимум, требуется дополнительное исследование
4	0	≤ 0	Все главные миноры четного порядка неотрицательны, а нечетного порядка неположительны	Может быть локальный максимум, требуется дополнительное исследование
5	0	$=0$	Матрица Гессе состоит из нулевых элементов	Требуется дополнительное исследование
6	0	$\gtrless 0$	Не выполняются условия п. 1–5	Нет экстремума

Второй способ (с помощью собственных значений матрицы Гессе – табл. 2).

Определение 2.3. *Собственные значения* λ_i, $i = 1,...,n$, матрицы $H(x^*)$ размеров $(n \times n)$ находятся как корни характеристического уравнения (алгебраического уравнения n-й степени):

$$\left| H(x^*) - \lambda E \right| = \begin{vmatrix} h_{11} - \lambda & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} - \lambda & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} - \lambda \end{vmatrix} = 0. \qquad (2.10)$$

З а м е ч а н и е. Собственные значения вещественной симметрической матрицы $H(x^*)$ вещественные.

Таблица 2

№ п/п	Второй способ	Тип стационарной точки x^*
1	$\lambda_1 > 0, ..., \lambda_n > 0$	Локальный минимум
2	$\lambda_1 < 0, ..., \lambda_n < 0$	Локальный максимум
3	$\lambda_1 \geq 0, ..., \lambda_n \geq 0$	Может быть локальный минимум, требуется дополнительное исследование
4	$\lambda_1 \leq 0, ..., \lambda_n \leq 0$	Может быть локальный максимум, требуется дополнительное исследование
5	$\lambda_1 = 0, ..., \lambda_n = 0$	Требуется дополнительное исследование
6	λ_i имеют разные знаки	Нет экстремума

Алгоритм решения задачи

Шаг 1. Записать необходимые условия экстремума первого порядка в виде (2.3) и найти стационарные точки x^* в результате решения системы n в общем случае нелинейных алгебраических уравнений с n неизвестными. Для численного решения системы могут использоваться методы простой итерации, Зейделя, Ньютона.

Шаг 2. В найденных стационарных точках x^* проверить выполнение достаточных, а если они не выполняются, то необходимых условий второго порядка с помощью одного из двух способов (см. табл. 1 и 2).

Шаг 3. Вычислить значения $f(x^*)$ в точках экстремума.

Описанный алгоритм отображен на рис. 1, где показана последовательность действий в случаях выполнения и невыполнения соответствующих условий экстремума при применении первого способа.

Замечания.

1. Продолжение исследований, которое требуется в ряде случаев, разобранных в табл. 1 и 2, при решении практических задач, как правило, не проводится, за исключением небольшого числа модельных примеров.

2. Часто на практике, особенно при применении численных методов поиска экстремума, рассматриваемых в последующих разделах, требуется проверить, выполняются ли необходимые и достаточные условия экстремума в некоторой точке. Такой анализ необходим, так как многие численные методы позволяют найти лишь стационарную точку, тип которой требует уточнения.

Рис. 1