11. ЧИСЕЛЬНІ МЕТОДИ ОПТИМІЗАЦІЇ. ПОСТАНОВКА ТА КЛАСИФІКАЦІЯ ЗАДАЧ ОПТИМІЗАЦІЇ, ЗАГАЛЬНІ ПОНЯТТЯ

11.1. Постановка та класифікація задач оптимізації

У загальному вигляді задача оптимізації формулюється таким чином.

Нехай X – деяка підмножина простору R^n (тобто $X \subset R^n$), $f(x)$ – деяка дійсна функція на R^n (тобто $f : R^n \to R^1$). Необхідно знайти таку точку $x^* \in X$, у якій функція $f(x)$ приймає мінімальне значення. Множина X при цьому називається **допустимою множиною**, а функція $f(x)$ – **цільовою функцією**. Стисло загальна задача оптимізації записується у вигляді

$$f(x) \to \min_{x \in X} . \qquad (11.1)$$

Зазначимо, що пошук точки максимуму деякої функції $\phi(x)$ еквівалентний пошуку точки мінімуму функції $f(x) = -\phi(x)$ (рис 9.1), тобто все одно як формулювати задачу оптимізації: через мінімум або через максимум.

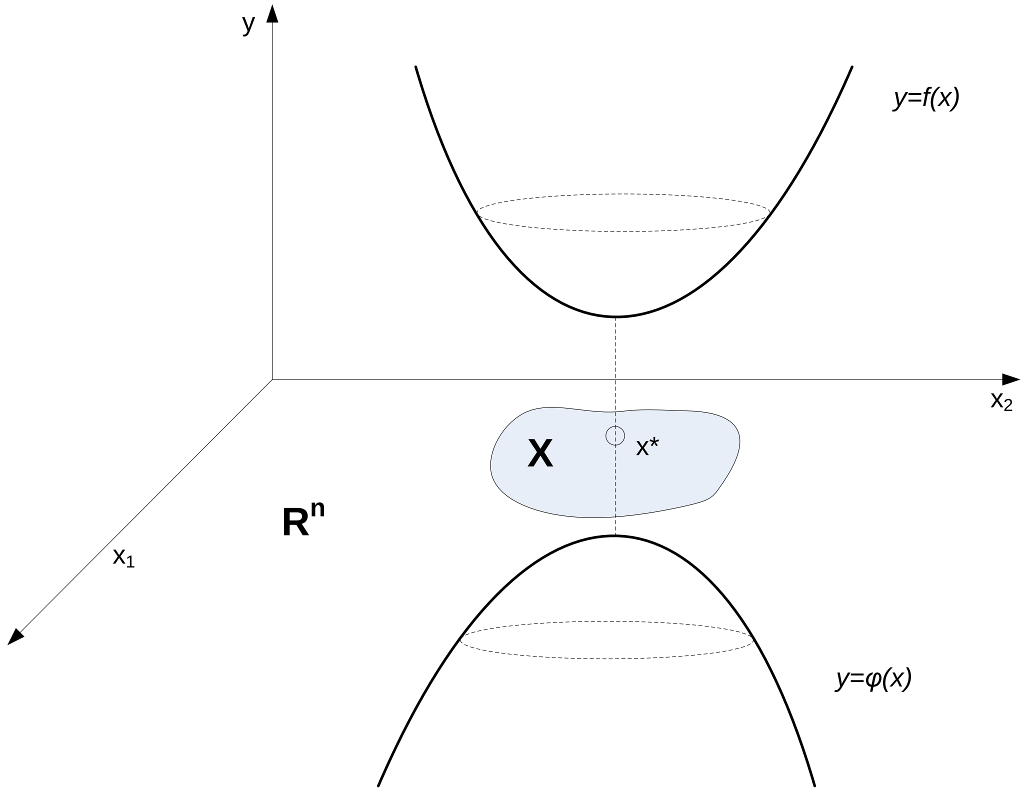

Рис. 11.1. **Точка екстремуму**

Визначення. Точка $x^* \in X$ називається **точкою глобального мінімуму** функції $f(x)$ на множині X, якщо для всіх $x \in X$, $x \neq x^*$, виконується нерівність $f(x) \geq f(x^*)$. При цьому точка x^* є розв'язком задачі (11.1), оскільки вона задовольняє умовам задачі.

Визначення. Точка $x^* \in X$ називається **точкою строгого глобального мінімуму** функції $f(x)$ на множині X, якщо для всіх $x \in X$, $x \neq x^*$, виконується нерівність $f(x) > f(x^*)$.

Визначення. Точка $x^* \in X$ називається **точкою локального мінімуму** функції $f(x)$ на множині X, якщо існує деякий малий окіл $V(x^*) \subset X$ точки x^* така, що для всіх $x \in V(x^*)$, $x \neq x^*$, виконується нерівність $f(x) \geq f(x^*)$. При цьому точка x^* може не бути розв'язком задачі (11.1).

Визначення. Точка $x^* \in X$ називається **точкою строгого локального мінімуму** функції $f(x)$ на множині X, якщо існує деякий малий окіл $V(x^*) \subset X$ точки x^* такий, що для всіх $x \in V(x^*)$, $x \neq x^*$, виконується нерівність $f(x) > f(x^*)$.

Приклади глобальних і локальних, строгих та нестрогих точок мінімуму наведені на рисунках 9.2 і 9.3.

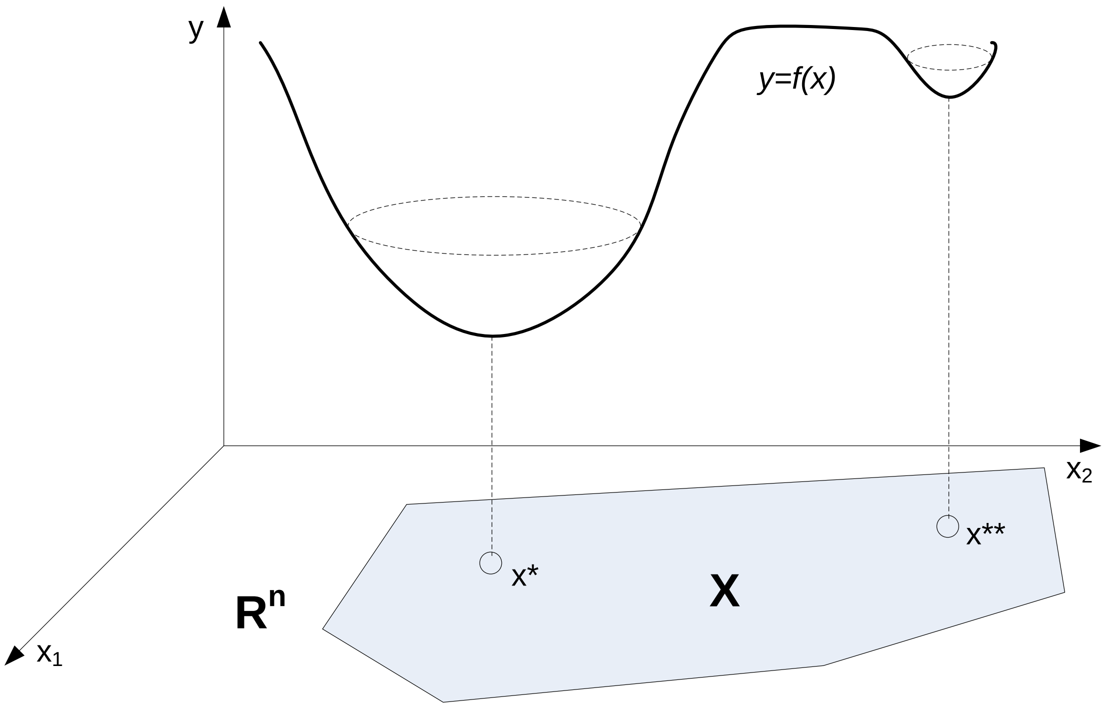

Рис. 11.2. **Точки мінімуму: x^* – глобальний мінімум, x^{**} – локальний мінімум**

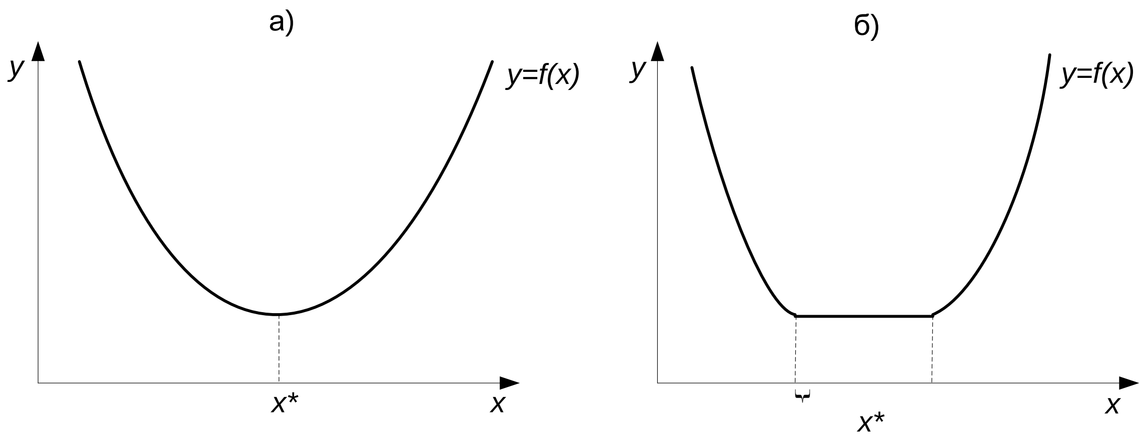

Рис. 11.3. **Точки мінімуму: а) – строгий мінімум, б) – нестрогий мінімум**

Визначення. Якщо множина X описується лінійними і нелінійними обмеженнями у вигляді рівностей і нерівностей, наприклад

$$X = \{ x \in R^n : \ g_i(x) = 0, i = \overline{1, m_1}, \ h_i(x) \leq 0, i = \overline{m_1 + 1, m} \}, \quad (11.2)$$

де $g_i(x), i=\overline{1,m_1}$, $h_i(x), i=\overline{m_1+1,m}$, – деякі дійсні функції від x, то задача (11.1) – (11.2) називається задачею **математичного програмування**. При цьому умови $g_i(x)=0$, $i=\overline{1,m_1}$ називаються **обмеженнями типу рівності**, а $h_i(x)\le 0$, $i=\overline{m_1+1,m}$ – **обмеженнями типу нерівності.**

Зазначимо, що запис (11.2) означає, що всі умови на змінні x в задачі (11.1) явно сформульовані. В противному випадку (тобто множина X не записується у вигляді (11.2)) задача (11.1) достатньо складна.

Нехай деяке підприємство випускає продукцію n видів, при цьому $c_1, c_2, \ldots, c_n$ – ціна реалізації продукції кожного виду. Позначимо через $x_1, x_2, \ldots, x_n$ кількість продукції кожного виду. Тоді $\sum_{j=1}^{n} c_j x_j$ – загальна виручка від реалізації всієї продукції. Якщо $Costs$ – витрати на виробництво і реалізацію продукції, то $\sum_{j=1}^{n} c_j x_j - Costs$ – прибуток підприємства. Природно розглядати функцію $f(x)=\sum_{j=1}^{n} c_j x_j - Costs$ як цільову і добиватися її максимуму. Позначимо через $b_1, b_2, \ldots, b_m$ запаси матеріалів, з яких виготовляється продукція, а через a_{ij} – кількість i-го матеріалу, необхідного для виготовлення однієї одиниці j-ї продукції. Тоді $\sum_{j=1}^{n} a_{ij} x_j$ – витрата i-го матеріалу необхідного для виготовлення всієї продукції і, отже, повинні виконуватися обмеження

$$X=\{x:\ \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i=\overline{1,m},\ x_j \ge 0,\ j=\overline{1,n}\}.$$

Таким чином, оптимальне планування виробництва продукції підприємством зводиться до розв'язання задачі математичного програмування.

Класифікація задач математичного програмування

У математичному програмуванні можна виділити два основні напрями.

Перший напрям – **власне математичне програмування**. До нього відносяться детерміновані задачі, тобто коли вся початкова інформація повністю визначена. Так у задачі (11.1) – (11.2) всі функції $f(x)$, $g_i(x)$, $i=\overline{1,m_1}$, $h_i(x)$, $i=\overline{m_1+1,m}$, повністю визначені.

Другий напрям – **стохастичне програмування.** До нього відносяться задачі, в яких початкова інформація містить невизначеності, або коли деякі параметри задачі носять випадковий характер з відомими імовірнісними характеристиками. Так, планування виробничої діяльності часто проводиться в умовах неповної інформації про реальну ситуацію, в якій виконуватиметься план. Або скажімо, коли екстремальна задача моделює роботу автоматичних пристроїв, яка супроводжується випадковими перешкодами. Зазначимо, що одна з головних труднощів стохастичного програмування полягає в самій постановці задач через складності аналізу початкової інформації.

У подальшому викладі розглядатимуться чисельні методи тільки для розв'язання задач математичного програмування (11.1) – (11.2), що відносяться до першого напряму.

Визначення. Задача математичного програмування (11.1) – (11.2) називається задачею **безумовної оптимізації**, якщо $X=R^n$, тобто умови на змінні x відсутні. В противному випадку, тобто коли $X\subset R^n$, говорять про **задачу умовної оптимізації.**

Визначення. Задача математичного програмування (11.1) – (11.2) називається **задачею лінійного програмування**, якщо всі функції $f(x)$, $g_i(x)$, $i=\overline{1,m_1}$, $h_i(x)$, $i=\overline{m_1+1,m}$, лінійні відносно x. Якщо хоча б одна з цих функцій є нелінійною, то задача (11.1) – (11.2) називається задачею **нелінійного програмування**. Так, задача планування випуску продукції, описана в прикладі вище, є задачею лінійного програмування.

Визначення. Задача математичного програмування (11.1) – (11.2) називається задачею **цілочисельного програмування**, якщо кожен елемент x_j $(j=\overline{1,n})$ вектора $x\in X$ може приймати значення тільки на множині цілих чисел.

Визначення. Задача цілочисельного програмування називається задачею **булевого програмування**, якщо кожен елемент x_j $(j=\overline{1,n})$ вектора $x\in X$ може приймати значення тільки на множині чисел $\{0,1\}$.

Визначення. Задача математичного програмування (11.1) – (11.2) називається задачею **одновимірної оптимізації**, якщо $X \subset R^1$. В цьому випадку, множина X є відрізком на дійсній осі, тобто $X = [a, b]$.

Схематично класифікація задач математичного програмування в найзагальнішому вигляді наведена на рис. 9.4.

Рис. 11.4. **Класифікація задач математичного програмування**

11.2. Необхідні та достатні умови точки оптимуму в задачах безумовної оптимізації

Нехай $f(x)$ – дійсна функція декількох змінних, яка має в кожній точці $x \in R^n$ усі частинні похідні по x_i ($i = \overline{1,n}$) (у таких випадках говорять, що функція $f(x)$ диференційована на R^n).

Нехай $\breve{x}$ – деяка точка простору R^n ($\breve{x} \in R^n$). Позначимо через $f'(\breve{x})$ вектор $\begin{pmatrix} \dfrac{\partial f(\breve{x})}{\partial x_1} \\ \cdots \\ \dfrac{\partial f(\breve{x})}{\partial x_n} \end{pmatrix}$. При цьому вектор $f'(\breve{x})$ називають **градієнтом функції** $f(x)$ у точці $\breve{x}$.

Наприклад, розглянемо функцію $f(x) = x_1^2 + 5x_1x_2 + x_2^4$, тобто $x \in R^2$. Тоді для всіх $x \in R^2$ $f'(x) = \begin{pmatrix} 2x_1 + 5x_2 \\ 5x_1 + 4x_2^3 \end{pmatrix}$. Тому в точці $\breve{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ градієнт $f'(\breve{x}) = \begin{pmatrix} 12 \\ 37 \end{pmatrix}$, в точці $\breve{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ градієнт $f'(\breve{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Зазначимо, що градієнт $f'(\breve{x})$, обчислений у точці $\breve{x}$, має важливу властивість, а саме він вказує напрямок з точки $\breve{x}$, у якій вона зростає найшвидше. Відповідно вектор $-f'(\breve{x})$, що називається **анти-градієнтом**, вказуватиме напрямок з точки, в якому вона убуває найшвидше [28].

Розглянемо задачу безумовної оптимізації, тобто необхідно знайти точку $x^* \in R^n$, у якій функція $f(x)$ приймає мінімальне значення. Стисло задача безумовної оптимізації записується у вигляді

$$f(x) \to \min_{x \in R^n}. \qquad (11.3)$$

Справедливе таке твердження [24; 28; 34].

Якщо x^* є точкою локального мінімуму (максимуму) функції $f(x)$, то градієнт функції $f(x)$ в точці x^* дорівнює нулю, тобто

$$f'(x^*) = \begin{pmatrix} \dfrac{\partial f(x^*)}{\partial x_1} \\ \cdots \\ \dfrac{\partial f(x^*)}{\partial x_n} \end{pmatrix} = 0 \text{ або } \frac{\partial f(x^*)}{\partial x_i} = 0 \quad \forall\, i = \overline{1,n}. \qquad (11.4)$$

Умова (11.4) називається **необхідною умовою мінімуму першого порядку**.

Припустимо тепер, що функція $f(x)$ має в кожній точці $x \in R^n$ всі другі частинні похідні по x_i ($i = \overline{1,n}$) (у таких випадках говорять, що функція $f(x)$ двічі диференційована на R^n).

Для деякої точки $\breve{x} \in R^n$ позначимо через $f''(\breve{x})$ матрицю $\begin{pmatrix} \dfrac{\partial f^2(\breve{x})}{\partial x_1^2} & \dots & \dfrac{\partial f^2(\breve{x})}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \dfrac{\partial f^2(\breve{x})}{\partial x_n \partial x_1} & \dots & \dfrac{\partial f^2(\breve{x})}{\partial x_n^2} \end{pmatrix}$ розмірності $n \times n$. При цьому матрицю $f''(\breve{x})$ називають **матрицею Гессе функції** $f(x)$ в точці $\breve{x}$ або **матрицею других похідних**.

Наприклад, для вже розглянутої вище функції $f(x) = x_1^2 + 5x_1x_2 + x_2^4$ для всіх $x \in R^2$ $f''(x) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}$. Тому в точці $\breve{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ матриця других похідних $f''(\breve{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 48 \end{pmatrix}$, в точці $\breve{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ – $f''(\breve{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.

Зазначимо, що якщо функція $f(x)$ двічі неперервно-диференційована на R^n, то матриця $f''(\breve{x})$ є симетричною, оскільки для всіх $\breve{x} \in R^n$, $\dfrac{\partial f^2(\breve{x})}{\partial x_i \partial x_j} = \dfrac{\partial f^2(\breve{x})}{\partial x_j \partial x_i}$, $i = \overline{1,n}$, $j = \overline{1,n}$.

Визначення. Кажуть, що симетрична матриця A розмірності $n \times n$ є **невід'ємно визначеною**, якщо для всіх $u \in R^n$ $\langle Au, u \rangle \geq 0$. При цьому записують скорочено $A \geq 0$.

Визначення. Кажуть, що симетрична матриця A розмірності $n \times n$ є **додатньо визначеною**, якщо для всіх $u \in R^n$ $\langle Au, u \rangle > 0$. При цьому записують скорочено $A > 0$.

Справедливе таке твердження [24; 28; 34].

Якщо x^* є точкою локального мінімуму функції $f(x)$, то

$$\begin{aligned} f'(x^*) &= 0, \\ f''(x^*) &\geq 0, \end{aligned} \tag{11.5}$$

тобто градієнт функції $f(x)$ в точці x^* дорівнює нулю, а матриця других похідних $f''(x^*)$ – невід'ємно визначена.

Умова (11.5) називається **необхідною умовою мінімуму другого порядку**.

Таким чином, для знаходження локальної точки мінімуму функції $f(x)$ з (11.3) можна визначити точку x^*, в якій виконуються умови (11.5). Однак така точка може виявитися точкою нестрогого мінімуму, тобто існуватиме множина таких точок, а значить, виникає проблема вибору. У таких випадках говорять, що задача (11.3) **некоректна**.

Справедливе таке твердження [24; 28; 34].

Якщо в деякій точці $x^* \in R^n$ виконуються умови

$$\begin{aligned} &f'(x^*) = 0, \\ &\left\langle f''(x^*)u, u \right\rangle \geq m\|u\|^2, \forall\, u \in R^n, \end{aligned} \tag{11.6}$$

де $m > 0$ (тобто градієнт функції $f(x)$ в точці x^* дорівнює нулю, а матриця других похідних $f''(x^*)$ – додатньо визначена), то точка x^* є точкою строгого локального мінімуму функції $f(x)$.

Умова (11.6) називається **достатньою умовою мінімуму другого порядку**.

Для подальшого викладення матеріалу нам знадобиться поняття "яружної" функції.

Визначення. Нехай A – невироджена матриця розмірності $n \times n$. **Числом обумовленості** матриці A називається величина $\mu(A) \equiv \left\|A^{-1}\right\| \times \|A\|$.

Визначення. Розглянемо двічі неперервно-диференційовану функцію $f(x)$ таку, що матриця других похідних $f''(x)$ невироджена для всіх $x \in V(x^*)$, де x^* – точка локального мінімуму функції $f(x)$, $V(x^*)$ – деякий окіл точки x^*. Тоді якщо для матриць $f''(x)$, при $x \in V(x^*)$, число обумовленості дуже велике, то говорять, що функція $f(x)$ **погано обумовлена** або "**яружна**" в околі $V(x^*)$.

Визначення. Нехай $c \in R^1$. Тоді множина $S_c = \left\{x \in R^n : f(x) = c\right\}$ називається **лінією рівня функції** $f(x)$, відповідною значенню c.

Для "яружної" функції $f(x)$ двох змінних ($x \in R^2$) лінії рівня мають вигляд дуже сильно витягнутих еліпсів.

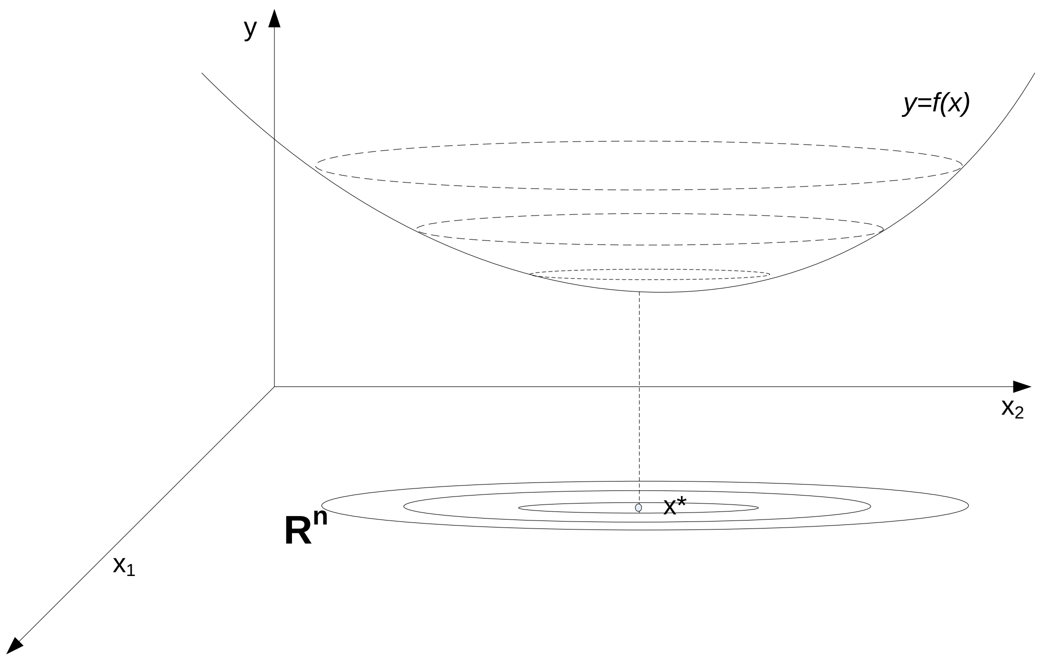

Рис. 11.5. **Лінії рівня "яружної" функції** $f(x)$

Лінії рівня витягнуті настільки сильно, що функція має вид яру, тобто уздовж яру функція убуває дуже повільно, а упоперек – швидко.

11.3. Необхідні та достатні умови точки оптимуму в задачах умовної оптимізації

Розглянемо для простоти викладення необхідні та достатні умови точки мінімуму для задачі умовної оптимізації (11.1) – (11.2) у випадку обмежень тільки типу рівності, тобто для задачі

$$f(x) \to \min_{x \in X}; \qquad (11.7)$$

$$X = \{x \in R^n : \ g_i(x) = 0, i = \overline{1,m}\}. \qquad (11.8)$$

Визначення. Функція двох векторних аргументів

$$L(x,y) \equiv f(x) + \sum_{i=1}^{m} y_i g_i(x), \ x \in R^n, \ y \in R^m \qquad (11.9)$$

називається **функцією Лагранжа** для задачі (11.7) – (11.8). Елементи y_i вектора y, $i = \overline{1,m}$ при цьому називаються **множниками Лагранжа**.

Необхідна умова оптимальності для задачі (11.7) – (11.8) формулюється таким чином [28].

Якщо $x^* \in R^n$ – розв'язок задачі (11.7) – (11.8), то існує вектор $y^* \in R^m$ такий, що точка $\begin{pmatrix} x^* \\ y^* \end{pmatrix} \in R^{n+m}$ є сідловою точкою функції Лагранжа (11.9), тобто

$$L(x,y^*) \geq L(x^*,y^*) \ \forall\, x \in V(x^*); \tag{11.10}$$

$$L(x^*,y) \leq L(x^*,y^*) \ \forall\, y \in V(y^*). \tag{11.11}$$

Це також означає, що x^* є точкою локального мінімуму по аргументу x функції $L(x,y^*)$, а y^* є точкою локального максимуму по аргументу y функції $L(x^*,y)$.

Для функції $\varphi(x,y) = x^2 - y^2$ ($x \in R^1$, $y \in R^1$) сідловою буде точка $\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in R^2$

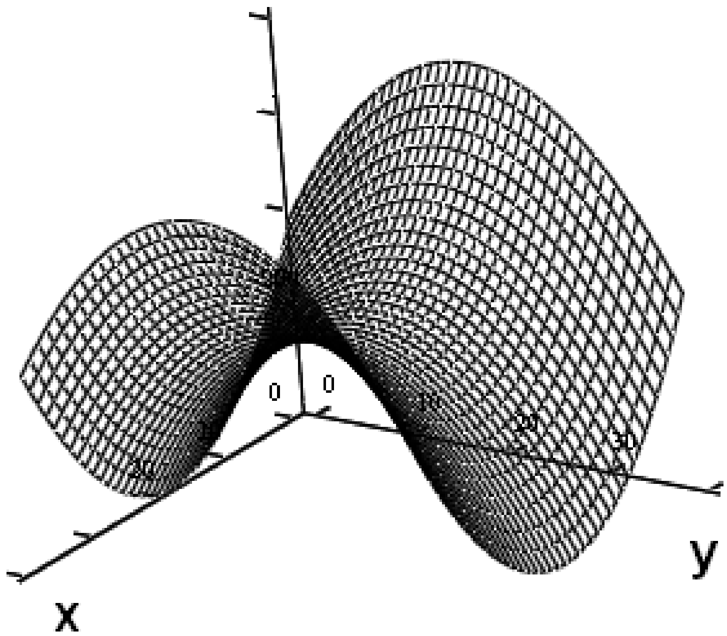

Рис. 11.6. **Сідлова точка** $\phi(x,y)$

З (11.9) згідно з необхідною умовою безумовного мінімуму (11.4) витікає, що

$$\frac{\partial L(x^*, y^*)}{\partial x} = f'(x^*) + \sum_{i=1}^{m} y_i^* g_i'(x^*) = 0, \quad (11.12)$$

а з (11.10) згідно з необхідною умові безумовного максимуму (11.4), що

$$\frac{\partial L(x^*, y^*)}{\partial y_i} = g_i(x^*) = 0, \ \forall i = \overline{1, n}. \quad (11.13)$$

Умова (11.13) якраз і підтверджує, що точка $x^* \in X$, а умову (11.12) називають **необхідною умовою мінімуму** для задачі (11.7) – (11.8).

11.4. Висновки

1. Класифікація задач оптимізації базується на виді цільової функції та допустимій множині розв’язків.
2. Для кожного класу задач оптимізації застосовуються свої чисельні методи розв'язку.
3. Необхідні і достатні умови мінімуму є важливим поняттям теорії оптимізації.

11.5. Контрольні запитання та завдання

1. Дайте визначення поняттям "цільова функція" і "допустима множина" в задачах оптимізації.
2. Сформулюйте постановку задачі безумовної оптимізації.
3. Сформулюйте необхідні умови мінімуму другого порядку в задачі безумовної оптимізації.
4. Сформулюйте достатні умови мінімуму другого порядку в задачі безумовної оптимізації.
5. Сформулюйте постановку задачі умовної оптимізації.
6. Сформулюйте постановку задачі лінійного програмування.
7. Сформулюйте постановку задачі нелінійного програмування.
8. Що називається функцією Лагранжа для задачі нелінійного програмування?
9. Сформулюйте необхідні умови умовного мінімуму 1-го порядку для задачі нелінійного програмування з обмеженнями типу рівнянь.