ЗМІСТОВИЙ МОДУЛЬ 6

Нелінійне програмування (НП).

Лекція 28. *Нелінійне програмування (НП). Огляд методів та моделей. Загальні питання нелінійного програмування. – 2год. [1].*

Завдання для самостійної роботи. *Нелінійне програмування, класифікація задач. Основні методи розв'язування задач НП, порівняння з методами розв'язування задач ЛП. Гіперповерхня рівня функції. Основні локальні властивості функцій багатьох змінних. Неперервність, диференційовність функції. Градієнт, гессіан, похідна за напрямком. Геометричний зміст градієнта. Глобальні та локальні екстремуми функції. -4год. [1-3,6,7].*

Нелінійне програмування

Постановки задач

В загальному випадку задачу математичного програмування можна формулювати так:

знайти мінімум (або максимум) функції

$$f(\boldsymbol{x}),\ \boldsymbol{x} \in E^n, \qquad (6.1)$$

за виконання умов

$$f_i(\boldsymbol{x})\ R_i\ 0,\ \ i=1,\ldots,m, \qquad (6.2)$$

де $R_i \in \{\ \le\ ,\ =\ ,\ \ge\ \}$.

У тому випадку, коли хоча б одна з функцій $f(\boldsymbol{x})$, $f_i(\boldsymbol{x})$, $i=1,\ldots,m$, є нелінійною, задачу (6.1)–(6.2) називають *задачею нелінійного програмування* (*ЗНЛП*).

Як і в *ЛП* функцію $f(\boldsymbol{x})$ називають *цільовою функцією*, область

$$D = \{\boldsymbol{x} \in E^n : f_i(\boldsymbol{x})\ R_i\ 0,\ \ i=1,\ldots,m\}$$

допустимою областю, довільний елемент $\boldsymbol{x} \in D$ – *допустимим вектором* (*точкою, розв'язком*) задачі нелінійного програмування.

Допустимий розв'язок $\boldsymbol{x}^* \in D$ такий, що

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in D} f(\boldsymbol{x}) \quad (\text{або}\ \ \boldsymbol{x}^* = \arg\max_{\boldsymbol{x}\in D} f(\boldsymbol{x})),$$

називається *оптимальним розв'язком* задачі (6.1)–(6.2).

Функції $f_i(\boldsymbol{x})$, $i=1,\ldots,m$, називають *функціями умов* задачі *НЛП*.

Зауважимо, що задача максимізації функції $f(\boldsymbol{x})$ еквівалентна задачі мінімізації $-f(\boldsymbol{x})$, обмеження $f_i(\boldsymbol{x}) \ge 0$ еквівалентне обмеженню $-f_i(\boldsymbol{x}) \le 0$, рівність $f_i(\boldsymbol{x}) = 0$ еквівалентна системі двох нерівностей $f_i(\boldsymbol{x}) \le 0$ і $-f_i(\boldsymbol{x}) \le 0$. Тому при формулюванні і розв'язуванні *ЗНЛП* можна обмежитись лише випадком мінімізації функції $f(\boldsymbol{x})$ за умов $f_i(\boldsymbol{x}) \le 0,\ i=1,\ldots,m,\ \boldsymbol{x} \in E^n$. Цю задачу ми часто будемо записувати у вигляді

$$\min \{f(\boldsymbol{x}) : \boldsymbol{x} \in E^n,\ f_i(\boldsymbol{x}) \le 0,\ i=1,\ldots,m\}\ . \qquad (6.3)$$

Зауважимо також, що обмеження-нерівності $f_i(\boldsymbol{x}) \le 0$ завжди можна перетворити в обмеження рівності шляхом введення невід'ємних змінних y_i^2

$$f_i(\boldsymbol{x}) + y_i^2 = 0.$$

В свою чергу систему умов $f_i(\boldsymbol{x})=0,\ i=1,\ldots,m,$ можна записати у вигляді однієї умови $F(\boldsymbol{x}) = f_1^2(\boldsymbol{x}) + \ldots + f_n^2(\boldsymbol{x}) = 0.$ Здається, на перший погляд, що задача

$$\min \{f(\boldsymbol{x}) : \boldsymbol{x} \in E^n,\ F(\boldsymbol{x}) = 0\} \qquad (6.4)$$

простіша за задачу (6.1)–(6.2), але це не так. При зведенні конкретної задачі типу (6.1)–(6.2) до задачі (6.4) втрачається специфіка *ЗНЛП*, що ускладнює пошук її розв'язку.

Необхідно також мати на увазі, що нелінійність легко дозволяє враховувати умову цілочисельності змінних задачі. Наприклад, для того, щоб змінна x_j набувала значень 0 і 1, достатньо, щоб вона задовольняла систему нерівностей: $x_j \ge 0$, $x_j \le 1$, $x_j\,(1 - x_j) \le 0$.

І останнє зауваження. Дуже часто обмеження *ЗНЛП* включають умову

$$\boldsymbol{x} \in X, \text{ де } X \subset E^n,$$

і, як правило,

$$X = \{\boldsymbol{x} \in E^n : x_j \ge 0,\ j=1,\ldots,n\}.$$

Тому поряд із задачею (6.3) ми також будемо розглядати задачу *НЛП* у вигляді:

$$\min \{f(\boldsymbol{x}) : \boldsymbol{x} \in E^n,\ f_i(\boldsymbol{x}) \le 0,\ i=1,\ldots,m,\ \boldsymbol{x} \in X\}. \qquad (6.5)$$

Розглянемо тепер класифікацію задач *НЛП*.

1. Класичні задачі оптимізації

Ці задачі полягають у знаходженні екстремуму функції $f(\boldsymbol{x}),\ \boldsymbol{x} \in E^n,$ при обмеженнях-рівностях

$$f_i(\boldsymbol{x}) = 0,\ i=1,\ldots,m,\ m \le n.$$

Їх ще називають *задачами відшукання умовного екстремуму*.

Якщо $m=0,$ то маємо *класичну задачу відшукання безумовного екстремуму* функції $f(\boldsymbol{x}),\ \boldsymbol{x} \in E^n.$

Для класичних задач оптимізації суттєвою є вимога гладкості (існування і неперервність у функцій $f(\boldsymbol{x})$ і $f_i(\boldsymbol{x})$ частинних похідних принаймні до 2-го порядку включно).

Класичні задачі оптимізації, хоча б принципово, можуть бути розв'язані класичними методами з використанням апарату диференціального числення. Однак труднощі обчислювального характеру, які виникають при цьому настільки значні, що для розв'язування практичних задач цього типу необхідно застосовувати інші методи.

2. Задачі з нелінійною цільовою функцією і лінійними обмеженнями

Ці задачі мають вигляд

$$f(\boldsymbol{x}) \to \min,$$

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} a_{ij} x_j - b_i \le 0,\ i = 1,\dots,m,$$

$$x_j \ge 0,\ j = 1,\dots,n.$$

Характерною ознакою таких задач є те, що їх допустима множина є многогранною множиною.

3. Задачі квадратичного програмування

В цих задачах потрібно мінімізувати квадратичну функцію

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_i x_j \to \min$$

при лінійних обмеженнях

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} a_{ij} x_j - b_i \le 0,\ i = 1,\dots,m,$$

$$x_j \ge 0,\ j = 1,\dots,n,$$

і за умови, що $f(\boldsymbol{x})$ є опуклою донизу функцією.

Задачі квадратичного програмування (*ЗКП*) можна віднести як до попереднього класу, так і до класу задач опуклого програмування. Але їх виділяють в окремий клас як із-за специфіки цільової функції, так і через специфіку умов-обмежень.

4. Задачі опуклого програмування

ЗНЛП (6.3) або (6.5), в яких цільова функція $f(\boldsymbol{x})$ є опуклою донизу, а допустима множина - опуклою, відносять до класу задач опуклого програмування (*ЗОП*). Методи розв'язування цих задач є найбільш розробленими у нелінійному програмування.

5. Задачі сепарабельного програмування

Для цих задач характерною ознакою є те, що і цільова функція $f(\boldsymbol{x})$ і функції умов, які ми позначимо для цього випадку через $g_i(\boldsymbol{x})$, є адитивними функціями, тобто їх можна подати у вигляді

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} f_j(x_j),$$

$$g_i(\boldsymbol{x}) = \sum_{j=1}^{n} g_{ij}(x_j),\ i = 1,\dots,m.$$

Специфіка цих задач визначає спеціальний клас методів їх розв'язування, які застосовуються і є ефективними тільки для таких задач.

Коротко зупинимось на *особливостях методів*, за допомогою яких розв'язуються задачі нелінійного програмування.

Згадаємо *ЗЛП*. Симплекс-метод дозволяє за скінченне число кроків установити, чи існує розв'язок *ЗЛП*, і знайти його у випадку існування.

Для розв'язування задач *НЛП* доводиться застосовувати, як правило, методи, що дозволяють знаходити лише наближені розв'язки або вимагають нескінченного числа кроків для досягнення точного розв'язку. Окрім цього, майже завжди ці методи дають лише локальні оптимуми. Прикладом таких методів може бути група *градієнтних методів*.

У деяких випадках при розв'язуванні *ЗНЛП* застосовують симплекс-метод, але в основному як допоміжний, для розв'язування допоміжних *ЗЛП*, що виникають в процесі розв'язування *ЗНЛП*.

Загальні питання нелінійного програмування

З геометричної точки зору функція багатьох змінних $z = f(x_1, \dots, x_n)$ визначає n-вимірну поверхню (*гіперповерхню*) в просторі E^{n+1}, кожній точці якої відповідає впорядкований набір дійсних чисел (z, $x_1, \dots, x_n$).

Поверхнею (*гіперповерхнею*) *рівня* функції $z = f(x_1, \dots, x_n)$ (у просторі E^3- *лінією рівня*) називається множина точок

$$\{\boldsymbol{x} \in E^n : f(\boldsymbol{x}) = z^0 = \text{const}\}.$$

Приклад 6.6. Розглянемо функцію $z = x_1^2 + x_2^2$.

Ця функція задає в 3-вимірному просторі параболоїд обертання. Лінії рівня функції z суть концентричні кола з центром у початку координат (див. рис. 6.6).

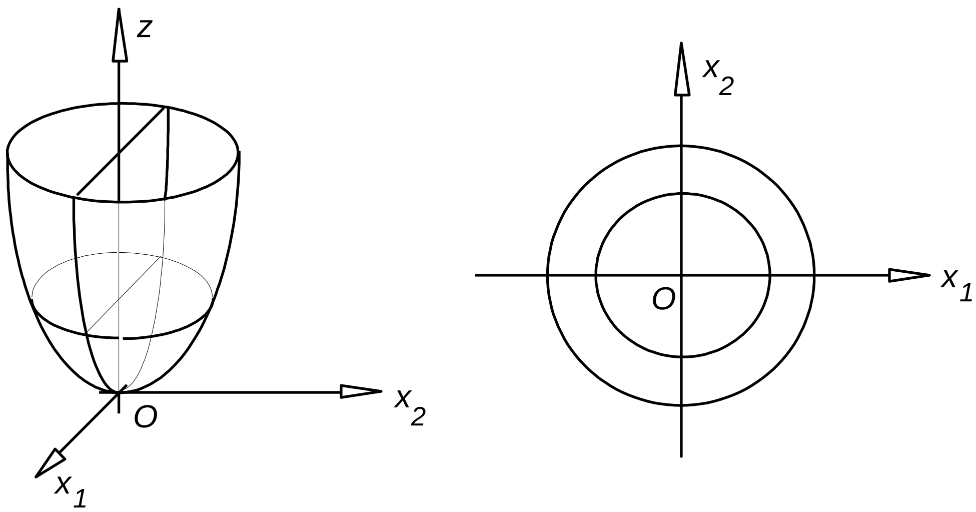


Рис. 6.6

Складність задачі *НЛП* визначається в першу чергу не числом обмежень, а властивостями цільової функції і функцій обмежень. Ці властивості насамперед поділяють на *локальні* та *глобальні*.

До локальних властивостей функції відносять, в першу чергу, ступінь її *гладкості*, тобто властивості, зв'язані з неперервністю функції, існуванням та неперервністю її частинних похідних різних порядків і похідних за напрямком.

Говорять що функція $z = f(\boldsymbol{x})$ *неперервна* в точці $\boldsymbol{x}^0$, якщо $f(\boldsymbol{x}^0)$ існує і $\forall \varepsilon > 0$ $\exists \delta > 0$ ($\delta = \delta(\varepsilon)$) таке, що для всіх $\boldsymbol{x}$ таких, що $\|\boldsymbol{x} - \boldsymbol{x}^0\| < \delta$ буде

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}^0)| < \varepsilon .$$

Тут і надалі використовується евклідова норма вектора.

Якщо функція $z = f(\boldsymbol{x})$ неперервна в кожній внутрішній точці деякої області D, то говорять, що $f(\boldsymbol{x})$ *неперервна на* D. Клас функцій неперервних на D позначають черед C_D (або просто C, якщо $D = E^n$).

Частинні похідні функції $z = f(\boldsymbol{x})$ в точці $\boldsymbol{x}^0$ визначаються як границі (якщо вони звичайно існують)

$$\frac{\partial f(\boldsymbol{x}^0)}{\partial x_j} = \lim_{h \to 0} \frac{f(\boldsymbol{x}^0 + h\,\mathbf{e}_j) - f(\boldsymbol{x}^0)}{h} ,$$

де $\boldsymbol{e}_j = (0,\ldots,0,1,0,\ldots,0)^T$ (1 стоїть на j-у місці) - *n-вимірний* одиничний вектор-стовпець, транспонований у вектор-рядок.

Функція $z = f(\boldsymbol{x})$ називається *диференційовною* в точці $\boldsymbol{x}^0$, якщо вона визначена в деякому околі точки $\boldsymbol{x}^0$ і для довільних $\boldsymbol{x}$ із цього околу $f(\boldsymbol{x})$ можна подати у вигляді

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^0) + \nabla^T f(\boldsymbol{x}^0)\,(\boldsymbol{x} - \boldsymbol{x}^0) + o(\|\boldsymbol{x} - \boldsymbol{x}^0\|).$$

Вектор

$$\nabla f(\boldsymbol{x}^0) = \left(\frac{\partial f(\boldsymbol{x}^0)}{\partial \boldsymbol{x}_1}, \frac{\partial f(\boldsymbol{x}^0)}{\partial \boldsymbol{x}_2}, \ldots, \frac{\partial f(\boldsymbol{x}^0)}{\partial \boldsymbol{x}_n} \right)^T$$

називається *градієнтом* функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^0$.

Отже, якщо функція $f(\boldsymbol{x})$ диференційовна в точці $\boldsymbol{x}^0$, то вона буде диференційовною по кожній змінній в точці $\boldsymbol{x}^0$, але не навпаки.

Якщо ж функція $f(\boldsymbol{x})$ неперервно диференційовна по кожній змінній в точці $\boldsymbol{x}^0$, то вона буде і диференційовною в цій точці.

Якщо функція $f(\boldsymbol{x})$ має неперервні частинні похідні у всіх внутрішніх точках деякої множини D, то говорять, що вона *належить класу* C_D^1 (якщо $D = E^n$, то говорять, що $f(\boldsymbol{x}) \in C^1$). У цьому випадку функція $f(\boldsymbol{x})$ буде диференційовною, а отже матиме градієнт в кожній внутрішній точці множини D.

Звичайним чином вводяться *частинні похідні більш високих порядків*. Наприклад, частинна похідна другого порядку $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\,\partial x_j}$ визначається як частинна похідна $\frac{\partial}{\partial x_j}$ від функції $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$.

Зауважимо, що, якщо $f(\boldsymbol{x}) \in C^2$, то

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\,\partial x_j} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j\,\partial x_i} .$$

Поряд з градієнтом ми також будемо використовувати *гессіан* функції $f(\boldsymbol{x})$, який визначається як матриця других частинних похідних функції $f(\boldsymbol{x}) \in C^2$

$$H_f(\boldsymbol{x}) = \left\| \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \, \partial x_j} \right\|, \quad i, j = 1,...,n.$$

Наведемо без доведення два формулювання **теореми Тейлора**, які надалі нам будуть потрібні.

Якщо $f(\boldsymbol{x})$ *належить класу* C_X^1 *на відкритій опуклій множині* $X \subset E^n$, *то для всіх точок* $\boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ $(\boldsymbol{x}^2 = \boldsymbol{x}^1 + \boldsymbol{y})$ *існує* $\theta \in (0, 1)$ *таке, що*

$$f(\boldsymbol{x}^2) = f(\boldsymbol{x}^1) + \nabla^T f(\theta \boldsymbol{x}^1 + (1 - \theta)\, \boldsymbol{x}^2)\, \boldsymbol{y}.$$

Якщо $f(\boldsymbol{x})$ *належить класу* C_X^2 *на відкритій опуклій множині* $X \subset E^n$, *то для всіх точок* $\boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ $(\boldsymbol{x}^2 = \boldsymbol{x}^1 + \boldsymbol{y})$ *існує* $\theta \in (0, 1)$ *таке, що*

$$f(\boldsymbol{x}^2) = f(\boldsymbol{x}^1) + \nabla^T f(\boldsymbol{x}^1)\, \boldsymbol{y} + (1/2)\, \boldsymbol{y}^T H_f(\theta \boldsymbol{x}^1 + (1 - \theta)\, \boldsymbol{x}^2)\, \boldsymbol{y}.$$

Похідною за напрямком $\boldsymbol{r} = (r_1, \ldots, r_n)^T$ $(\|\boldsymbol{r}\| = 1)$ функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^0$ називається границя (звичайно, якщо вона існує)

$$D_r f(\boldsymbol{x}^0) = \lim_{h \to +0} \frac{f(\boldsymbol{x}^0 + h\boldsymbol{r}) - f(\boldsymbol{x}^0)}{h},$$

Нехай $f(\boldsymbol{x}) \in C^1$. Покладемо $\boldsymbol{x}^1 = \boldsymbol{x}^0$, $\boldsymbol{x}^2 = \boldsymbol{x}^0 + \boldsymbol{y} = \boldsymbol{x}^0 + h\boldsymbol{r}$ в першій формулі Тейлора. Отримаємо

$$\frac{f(\boldsymbol{x}^0 + h\boldsymbol{r}) - f(\boldsymbol{x}^0)}{h} = \nabla^T f(\theta \boldsymbol{x}^0 + (1 - \theta)(\boldsymbol{x}^0 + h\boldsymbol{r}))\, \boldsymbol{r}.$$

Після переходу до границі при $h \to +0$ в останній рівності будемо мати

$$D_r f(\boldsymbol{x}^0) = \nabla^T f(\boldsymbol{x}^0)\boldsymbol{r} = (\nabla f(\boldsymbol{x}^0), \boldsymbol{r}) = |\nabla f(\boldsymbol{x}^0)|\; |\boldsymbol{r}| \cos(\nabla f(\boldsymbol{x}^0), \boldsymbol{r}).$$

З фізичної точки зору частинні похідні функції $f(\boldsymbol{x}) \in C^1$ характеризують швидкість зміни функції в напрямках координатних осей, а похідна за напрямком – швидкість зміни функції в заданому напрямку.

Із останньої формули випливає, що при $\|\boldsymbol{r}\| = 1$ похідна за напрямком буде максимальною, якщо напрямок $\boldsymbol{r}$ співпадає з напрямком градієнта $\nabla f(\boldsymbol{x}^0)$. При цьому $D_r f(\boldsymbol{x}^0) = |\nabla f(\boldsymbol{x}^0)|$. З фізичної точки зору це означає, що *градієнт визначає напрямок максимального зростання функції* в точці $\boldsymbol{x}^0$.

Вкажемо *геометричний зміст градієнта* (див. рис. 6.7). Нехай маємо функцію

$$z = f(x_1, x_2).$$

Розглянемо лінію рівня

$$f(x_1, x_2) = z^0.$$

З'ясуємо, як провести дотичну до лінії рівня в точці $\boldsymbol{x}^0 = (x_1^0, x_2^0)$, яка лежить на цій лінії.

Лінія рівня $x_2 = x_2(x_1)$ задана неявно рівнянням

$$F(x_1, x_2) = f(x_1, x_2) - z^0 = 0.$$

Рівняння дотичної до лінії рівня в точці $\boldsymbol{x}^0$ має вигляд

$$x_2 - x_2^0 = \frac{dx_2}{dx_1}(x_1^0)(x_1 - x_1^0).$$

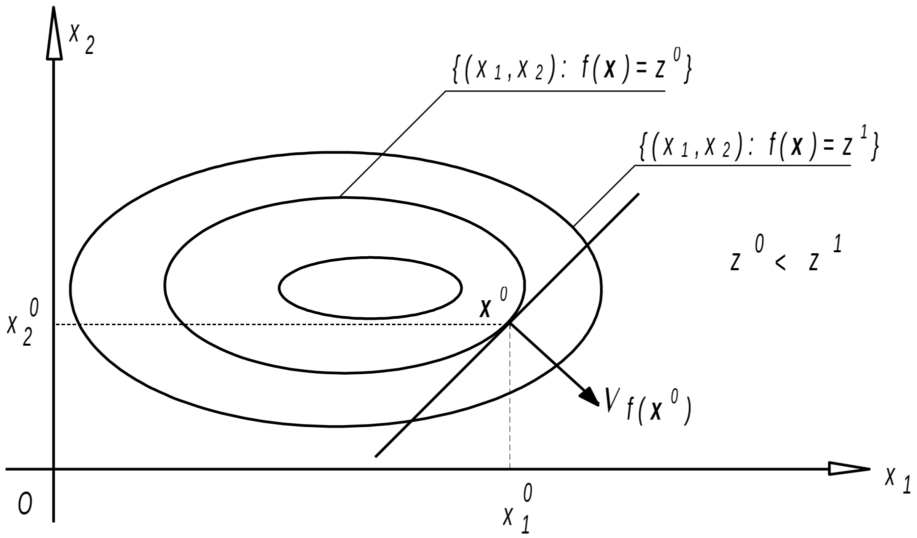


Рис. 6.7

Оскільки

$$\frac{dx_2}{dx_1} = -\frac{F'_{x_1}}{F'_{x_2}} = -\frac{f'_{x_1}}{f'_{x_2}},$$

то з рівняння дотичної будемо мати

$$\frac{\partial f(x^0)}{\partial x_2}(x_2 - x_2^0) = -\frac{\partial f(x^0)}{\partial x_1}(x_1 - x_1^0),$$

або

$$(\nabla f(\boldsymbol{x}^0), \boldsymbol{x} - \boldsymbol{x}^0) = 0.$$

З останньої рівності випливає, що з геометричної точки зору *градієнт є вектором нормалі до лінії* (*поверхні при* $n>3$) *рівня в точці* $\boldsymbol{x}^0$*, який спрямований в сторону зростання функції* $f(\boldsymbol{x})$.

До *глобальних властивостей* функцій багатьох змінних, в першу чергу, відносять наявність у них *екстремумів*.

Означення 6.1. *Говорять, що функція* $f(\boldsymbol{x})$*, яка визначена на замкненій множині* $X \subset E^n$*, досягає на ній абсолютного (глобального) мінімуму в точці* $\boldsymbol{x}^* \in X$*, якщо* $\forall\, \boldsymbol{x} \in X$ *виконується* $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$.

Означення 6.2. *Говорять, що функція* $f(\boldsymbol{x})$, *яка визначена на множині* $X \subset E^n$, *має в точці* $\boldsymbol{x}^* \in X$ *відносний (локальний) мінімум, якщо існує такий окіл* $S(\boldsymbol{x}^*)$ *точки* $\boldsymbol{x}^*$*, що* $\forall\, \boldsymbol{x} \in S(\boldsymbol{x}^*) \cap X$ *виконується* $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$.

Означення 6.3. *Задачу НЛП назвемо одноекстремальною, якщо її довільний відносний мінімум буде і абсолютним.*

Лекція 29. *Геометрична інтерпретація задач НП. Класичні задачі оптимізації. – 2год. [1].*

Завдання для самостійної роботи. *Геометрична інтерпретація задачі НП у порівнянні з геометричною інтерпретацією задачі ЛП. Основні відмінності задач НП від задач ЛП (на прикладах). Класичні задачі оптимізації. Унімодальні функції. Методи одновимірної оптимізації.*

Геометрична інтерпретація задачі нелінійного програмування

Розглянемо на прикладах відмінності *ЗНЛП* від *ЗЛП* і проаналізуємо труднощі, які породжуються нелінійністю. Використаємо для цього геометричну інтерпретацію *ЗЛП* і *ЗНЛП*.

Приклад 6.1. *Розглянемо задачу*

$$z = 0{,}5\,x_1 + 2\,x_2 \rightarrow \max,$$

$$x_1 + x_2 \le 6,$$

$$x_1 - x_2 \le 1,$$

$$x_1 - 2\,x_2 \ge -8,$$

$$x_1 \ge 0,\ x_2 \ge 0.$$

Розв'язуючи її графічно (див. рис. 6.1), знайдемо

$$\mathbf{x}^* = (4/3, 14/3),\ z^* = 10.$$

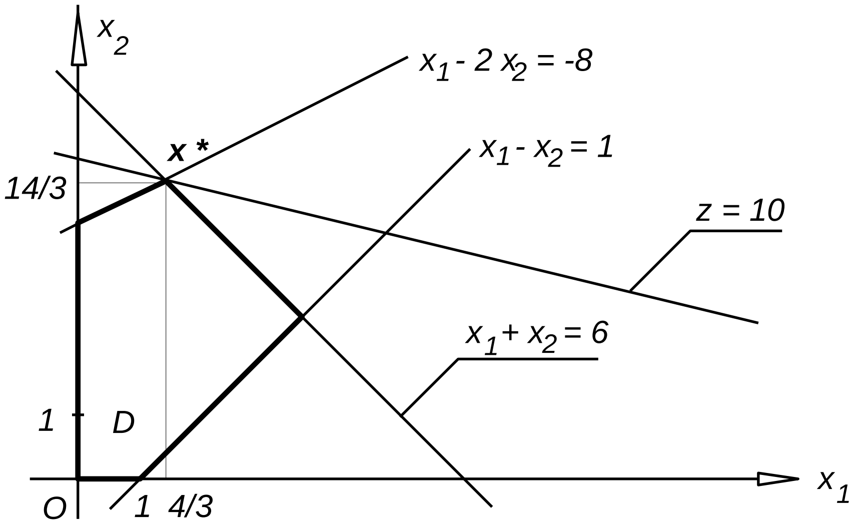


Рис. 6.1

Ще раз підкреслимо найважливіші властивості *ЗЛП*:

1. *Допустима множина ЗЛП є опуклою многогранною множиною, що має скінченне число вершин.*

2. *Поверхнею рівня цільової функції ЗЛП є гіперплощина. Гіперплощини, що відповідають різним значенням сталої рівня, паралельні між собою.*

3. *Локальний мінімум (максимум) в ЗЛП також є і глобальним.*

4. *Якщо цільова функція обмежена знизу (зверху) в задачі мінімізації (максимізації) на допустимій множині ЗЛП, то принаймні одна вершина допустимої множини є оптимальним розв'язком ЗЛП.*

Для задач *НЛП* вказані властивості, як правило, не мають місця.

Розглянемо приклади.

Приклад 6.2. *Мінімізувати функцію*

$$z = 10\,(x_1^2 \quad 3.5) + 20\,(x_2^2 \quad 4) \to \min$$

на допустимій множині прикладу 6.1 (*див. рис. 6.2*).

Ця задача є задачею квадратичного програмування. З її геометричної інтерпретації випливає, що оптимальний розв'язок $\boldsymbol{x}^*$ є точкою дотику лінії рівня цільової функції до прямої $x_1 + x_2 = 6$. Положення цієї точки через лінію рівня визначити неможливо, оскільки невідоме відповідне значення сталої рівня z^*. Отже, поки що маємо лише одне рівняння для визначення координат точки $\boldsymbol{x}^*$

$$x_1^* + x_2^* = 6. \qquad (6.6)$$

Для того, щоб отримати друге рівняння, зауважимо, що кутовий коефіцієнт дотичної до лінії рівня в точці $\boldsymbol{x}^*$ рівний кутовому коефіцієнту прямої $x_1 + x_2 = 6$.

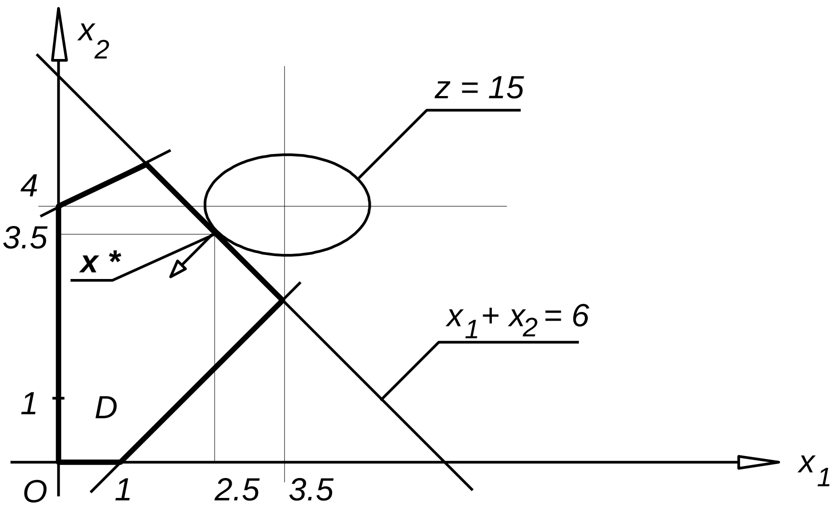


Рис. 6.2

Тоді, вважаючи x_2 неявною функцією від x_1, яка визначається співвідношенням

$$F(x_1, x_2) = 10\,(x_1^2 \quad 3{,}5) + 20\,(x_2^2 \quad 4) \,\square\, z = 0,$$

z параметром, що визначає сталу рівня, отримаємо за правилом диференціювання неявної функції

$$F'_{x_1} + F'_{x_2}\frac{dx_2}{dx_1} = 0,$$

звідки

$$\frac{d\,x_2}{d\,x_1} = -\frac{F'_{x_1}}{F'_{x_2}} = -\frac{20\,(x_1 - 3{,}5)}{40\,(x_2 - 4)} = -\frac{(x_1 - 3{,}5)}{2\,(x_2 - 4)}\,.$$

З іншого боку $\dfrac{dx_2}{dx_1} = -1$, як кутовий коефіцієнт прямої $x_2 = \ \ x_1 + 6$. Остаточно, в точці $\boldsymbol{x}^*$ будемо мати рівняння

$$-\frac{x_1^* - 3{,}5}{2(x_2^* - 4)} = -1 . \qquad (6.7)$$

Розв'язавши систему (6.6) (6.7), отримаємо $\boldsymbol{x}^* = (2{,}5;\, 3{,}5)$, і далі легко знаходимо $z^* = 15$.

Зауважимо, що знайдений оптимальний розв'язок щойно розглянутої задачі *НЛП* не є вершиною її допустимої області, хоча і досягається на її межі.

Приклад 6.3. *За умов прикладу 6.1 знайти мінімум функції*

$$z = 10\,(x_1 \quad 2)^2 + 20\,(x_2 \quad 3)^2$$

(*див. рис. 6.3*).

Функція z є опуклою квадратичною функцією, яка набуває лише невід'ємних значень. Найменшого значення, рівного нулеві, вона набуває у точці $\boldsymbol{x}^* = (2, 3)$, яка лежить всередині області D.

Отже, оптимальне значення цільової функції може досягатися у внутрішній точці допустимої області *ЗНЛП*.

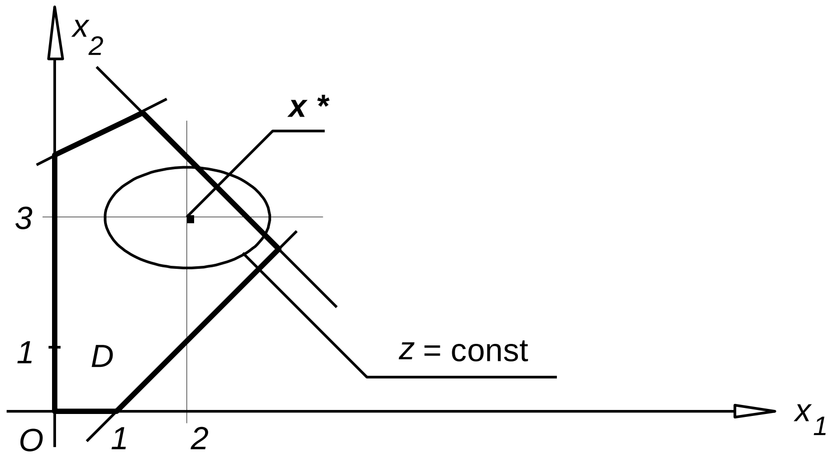


Рис. 6.3

Приклад 6.4. *Знайти максимум функції*

$$z = 25\,(x_1 \quad 2)^2 + (x_2 \quad 2)^2$$

за умов

$$\begin{aligned} &x_1 + \quad x_2 \geq \quad 2, \\ &x_1 \qquad x_2 \geq \quad 2, \\ &x_1 + \quad x_2 \leq \quad 6, \\ &x_1 \quad 3x_2 \leq \quad 2, \\ &x_1 \geq 0,\ x_2 \geq 0, \end{aligned}$$

(див. рис. 6.4).

Допустимою множиною цієї задачі є чотирикутник $V_1V_2V_3V_4$. Лінії рівня суть еліпси. З геометричної інтерпретації випливає, що функція z має на допустимій множині чотири максимуми, які досягаються у всіх вершинах чотирикутника $V_1V_2V_3V_4$. Максимуми в т. V_1,V_2,V_3 є локальними ($z^*=4, 100, 4$). Глобальним є максимум в т. V_4 ($z^*=226$). Отже *ЗНЛП* може мати кілька оптимумів, і не обов'язково тільки один з них буде визначати її оптимальний розв'язок.

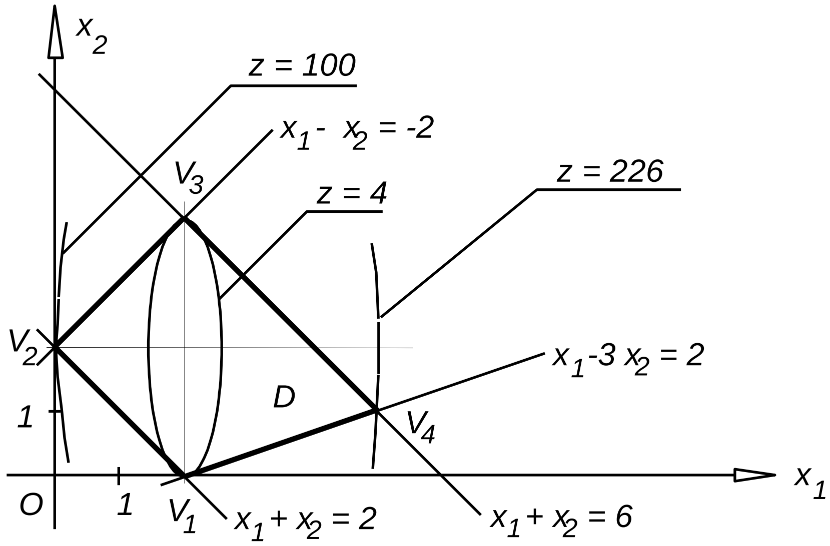


Рис. 6.4

В розглянутих прикладах допустимі множини були опуклими, оскільки визначались системами лінійних нерівностей. У загальному випадку, коли серед обмежень *ЗНЛП* є нелінійні, її допустима множина може бути і не опуклою, і не зв'язною.

Приклад 6.5. Нехай допустима множина *ЗНЛП* визначається системою

$$(x_1 \quad 1)\, x_2 \le 1,$$
$$x_1 + x_2 \ge 4,$$
$$x_1 \ge 0,\ x_2 \ge 0,$$

(див. рис. 6.5).

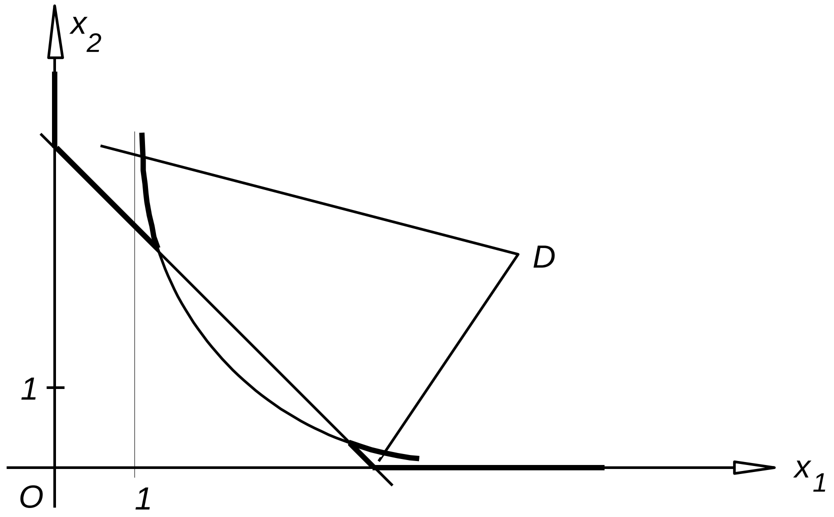


Рис. 6.5

З графічної інтерпретації видно, що D не є опуклою і не є зв'язною множиною.

Класичні методи оптимізації

Необхідні та достатні умови екстремуму

Задача безумовної мінімізації

Розглянемо задачу:

$$f(\boldsymbol{x}) \to \min,\ \boldsymbol{x} = (x_1,...,x_n) \in E^n. \qquad (8.1)$$

Теорема 8.1 (необхідні умови мінімуму). *Нехай функція $f(\boldsymbol{x})$ має мінімум (локальний або глобальний) в точці $\boldsymbol{x}^0$. Тоді:*

1) *якщо $f(\boldsymbol{x}) \in C^1$, то в точці $\boldsymbol{x}^0$ $\nabla f(\boldsymbol{x}^0) = 0$;*
2) *якщо $f(\boldsymbol{x}) \in C^2$, то в точці $\boldsymbol{x}^0$ $\nabla f(\boldsymbol{x}^0) = 0$ і гессіан $H_f(\boldsymbol{x}^0)$ є невід'ємно визначеною матрицею, тобто $\forall \boldsymbol{y} \in E^n$ $\boldsymbol{y}^T H_f(\boldsymbol{x}^0)\boldsymbol{y} \geq 0$.*

Доведення. Доведемо 1). Скористаємося першим формулюванням теореми Тейлора. Маємо $\forall \boldsymbol{y} \in E^n$ при відповідних $\theta \in (0, 1)$

$$f(\boldsymbol{x}^0 + \boldsymbol{y}) - f(\boldsymbol{x}^0) = \nabla^T f(\boldsymbol{x}^0 + \theta \boldsymbol{y}).$$

Оскільки $\boldsymbol{x}^0$ — точка мінімуму $f(\boldsymbol{x})$, то існує окіл $S(\boldsymbol{x}^0)$ точки $\boldsymbol{x}^0$, в якому $\forall \boldsymbol{y} \in E^n$, таких що $\boldsymbol{x}^0 + \boldsymbol{y} \in S(\boldsymbol{x}^0)$, буде

$$f(\boldsymbol{x}^0 + \boldsymbol{y}) - f(\boldsymbol{x}^0) \geq 0.$$

Тому за цих умов при відповідних $\theta \in (0, 1)$ також буде

$$\nabla^T f(\boldsymbol{x}^0 + \theta \boldsymbol{y})\,\boldsymbol{y} \geq 0.$$

Оскільки функція $f(\boldsymbol{x})$ диференційовна, то градієнт $\nabla f(\boldsymbol{x}^0 + \theta \boldsymbol{y})$ при досить малій нормі $\boldsymbol{y}$ буде зберігати знак в околі $S(\boldsymbol{x}^0)$, як неперервна функція. Тоді $\forall \boldsymbol{y}$ з цього околу буде $\nabla^T f(\boldsymbol{x}^0)\,\boldsymbol{y} \geq 0$.

Припустимо від супротивного, що $\nabla f(\boldsymbol{x}^0) \neq 0$.

Тоді в силу довільності $\boldsymbol{y}$ завжди $\exists \boldsymbol{y} \neq 0$ з як завгодно малою нормою, при якому буде $\nabla^T f(\boldsymbol{x}^0)\boldsymbol{y} < 0$, що суперечить невід'ємності добутку $\nabla^T f(\boldsymbol{x}^0)\boldsymbol{y}$ при всіх $\boldsymbol{y}$ з досить малою нормою, а, отже, і означенню мінімуму в точці $\boldsymbol{x}^0$. Тому необхідно $\nabla f(\boldsymbol{x}^0) = 0$.

Доведемо 2). Скористаємося другою теоремою Тейлора. Маємо $\forall \boldsymbol{y} \in E^n$ при відповідних $\theta \in (0, 1)$

$$f(\boldsymbol{x}^0 + \boldsymbol{y}) - f(\boldsymbol{x}^0) = \nabla^T f(\boldsymbol{x}^0)\boldsymbol{y} + (1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0 + \theta \boldsymbol{y})\boldsymbol{y}.$$

Оскільки $\boldsymbol{x}^0$ – точка мінімуму функції $f(\boldsymbol{x})$, то існує окіл $S(\boldsymbol{x}^0)$ точки $\boldsymbol{x}^0$, в якому $\forall \boldsymbol{y} \in E^n$ таких, що $\boldsymbol{x}^0 + \boldsymbol{y} \in S(\boldsymbol{x}^0)$, буде

$$f(\boldsymbol{x}^0 + \boldsymbol{y}) - f(\boldsymbol{x}^0) \geq 0.$$

Крім того за доведеним $\nabla f(\boldsymbol{x}^0) = 0$. Тому за цих умов при відповідних $\theta \in (0,1)$ також буде виконуватись нерівність

$$(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0 + \theta \boldsymbol{y})\boldsymbol{y} \geq 0.$$

Неперервність других похідних ($f(\boldsymbol{x}) \in C^2$) гарантує збереження знаку функцією $H_f(\boldsymbol{x})$ і в точці $\boldsymbol{x}^0$, тобто буде

$$(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0)\boldsymbol{y} \geq 0$$

для довільних $\boldsymbol{y}$, що не виводять за межі досить малого околу точки $\boldsymbol{x}^0$.

Якщо матриця $H_f(\boldsymbol{x}^0)$ не є невід'ємно визначеною, то в силу довільності вектора $\boldsymbol{y}$ завжди знайдеться вектор $\boldsymbol{y}$ з як завгодно малою нормою, для якого буде $(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0)\boldsymbol{y} < 0$, а, отже, і $(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0 + \theta \boldsymbol{y})\boldsymbol{y} < 0$, що суперечить означенню мінімуму функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^0$. Теорему доведено.

Означення 8.1. *Точка $\boldsymbol{x}^0$, яка задовольняє умову $\nabla f(\boldsymbol{x}^0) = 0$ називається стаціонарною точкою диференційовної функції $f(\boldsymbol{x})$.*

Інколи умову стаціонарності формулюють так: $\boldsymbol{x}^0$ *- стаціонарна точка диференційовної функції $f(\boldsymbol{x})$, якщо $\nabla^T f(\boldsymbol{x}^0)\boldsymbol{y} = 0 \quad \forall \boldsymbol{y} \neq 0$*

Зауваження. *Необхідними умовами максимуму функції $f(\boldsymbol{x}) \in C^2$ в точці $\boldsymbol{x}^0$ є умова стаціонарності і недодатна визначеність її гессіана:*

$$\nabla f(\boldsymbol{x}^0) = 0, \ \forall \boldsymbol{y} \in E^n \ \ \boldsymbol{y}^T H_f(\boldsymbol{x}^0)\boldsymbol{y} \leq 0.$$

Доводиться аналогічно.

Теорема 8.2 (достатня умова локального мінімуму). *Нехай:*

1) $f(\boldsymbol{x}) \in C^2$ $(\boldsymbol{x} \in E^n)$,
2) *в точці $\boldsymbol{x}^0 \in E^n$ виконується умова стаціонарності $\nabla f(\boldsymbol{x}^0) = 0$ (або $\nabla^T f(\boldsymbol{x}^0)\boldsymbol{y} = 0 \quad \forall \boldsymbol{y} \neq 0$),*
3) *гессіан $H_f(\boldsymbol{x}^0)$ є додатно визначеною матрицею.*

Тоді точка $\boldsymbol{x}^0$ є точкою строгого локального мінімуму функції $f(\boldsymbol{x})$.

<u>Доведення</u>. Нехай точка $\boldsymbol{x}^0$ задовольняє умови теореми. Розглянемо різницю $f(\boldsymbol{x}^0 + \boldsymbol{y}) - f(\boldsymbol{x}^0) \quad \forall \boldsymbol{y} \in E^n$.

За другим формулюванням теореми Тейлора та з урахуванням умови 2) маємо $\forall \boldsymbol{y} \in E^n$ і деяких $\theta \in (0, 1)$

$$f(\boldsymbol{x}^0+\boldsymbol{y}) - f(\boldsymbol{x}^0)=(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0+\theta\boldsymbol{y})\,\boldsymbol{y}.$$

За умовами теореми гессіан $H_f(\boldsymbol{x}^0)$ є додатно визначеною матрицею, тобто

$$(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0)\boldsymbol{y}>0 \quad \forall \boldsymbol{y}\in E^n.$$

Оскільки $f(\boldsymbol{x})\in C^2$, то неперервність гессіана $H_f(\boldsymbol{x})$ гарантує збереження його знаку і в деякому околі точки $\boldsymbol{x}^0$. Тоді при $\theta\in(0,1)$ і $\forall\boldsymbol{y}\in E^n$ з досить малою нормою будемо мати

$$(1/2)\boldsymbol{y}^T H_f(\boldsymbol{x}^0+\theta\boldsymbol{y})\,\boldsymbol{y}>0.$$

Це означає, що для будь-якого $\boldsymbol{y}$, такого що $\boldsymbol{x}^0+\boldsymbol{y}$ попадає в досить малий окіл точки $\boldsymbol{x}^0$, має місце умова

$$f(\boldsymbol{x}^0+\boldsymbol{y})-f(\boldsymbol{x}^0)>0,$$

тобто точка $\boldsymbol{x}^0$ є точкою строгого локального мінімуму функції $f(\boldsymbol{x})$.

Зауваження. *Достатніми умовами строгого локального максимуму функції* $f(\boldsymbol{x})\in C^2$ *в точці* $\boldsymbol{x}^0\in E^n$ *є умова стаціонарності і від'ємна визначеність її гессіана*

$$\nabla f(\boldsymbol{x}^0)=0,\ \forall\boldsymbol{y}\in E^n\ \ \boldsymbol{y}^T H_f(\boldsymbol{x}^0)\,\boldsymbol{y}<0.$$

Доводиться аналогічно.

Критерій Сильвестра. *Нехай* $\Delta_1(\boldsymbol{x}),\ldots,\Delta_n(\boldsymbol{x})$ — *послідовні головні мінори матриці* $H_f(\boldsymbol{x})$.

Для того щоб матриця $H_f(\boldsymbol{x})$ *була додатно визначеною необхідно і достатньо, щоб виконувались умови*

$$\Delta_1(\boldsymbol{x})>0,\ldots,\Delta_n(\boldsymbol{x})>0 \quad \forall\boldsymbol{x}\in E^n.$$

Для того, щоб матриця $H_f(\boldsymbol{x})$ *була від'ємно визначеною необхідно і достатньо, щоб виконувались умови*

$$\Delta_1(\boldsymbol{x})<0,\ \Delta_2(\boldsymbol{x})>0,\ldots,(-1)^n\Delta_n(\boldsymbol{x})>0 \quad \forall\boldsymbol{x}\in E^n.$$

Задача умовної мінімізації

Розглянемо задачу на умовний екстремум

$$f^0(\boldsymbol{x})\to\min, \tag{8.2}$$

$$f^i(\boldsymbol{x})=0,\quad i=1,\ldots,m, \tag{8.3}$$

$$\boldsymbol{x}\in E^n,$$

в якій $f^i(\boldsymbol{x})\in C^1,\quad i=0,1,\ldots,m.$

Якщо систему (8.3) можна перетворити до еквівалентного вигляду

$$\begin{cases} x_1 = x_1(x_{m+1}, x_{m+2}, \ldots, x_n \\ x_2 = x_2(x_{m+1}, x_{m+2}, \ldots, x_n \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m = x_m(x_{m+1}, x_{m+2}, \ldots, x_n \end{cases} \tag{8.4}$$

виразивши за допомогою (8.3) перші m змінних x_i ($i=1,\dots,m$) через інші змінні x_i ($i=m+1,\dots,n$), то задачу на умовний екстремум (8.2), (8.3) можна звести до задачі на безумовний екстремум для функції

$$f(x_{m+1},x_{m+2},\dots,x_n) = f^0(x_1(x_{m+1},\dots,x_n),\dots,x_m(x_{m+1},\dots,x_n),x_{m+1},x_{m+2},\dots,x_n)$$

змінних $(x_{m+1},x_{m+2},\dots,x_n) \in E^{n-m}$, яку можна дослідити за допомогою необхідних і достатніх умов, доведених раніше. Але такий підхід має обмежене застосування із-за того, що явний вираз однієї групи змінних через інші змінні за допомогою системи (8.3) можна отримати далеко не завжди.

Більш загальний підхід до дослідження задачі відшукання умовного екстремуму дифференційовної функції дає метод Лагранжа. Цей метод полягає у заміні задачі умовного екстремуму (8.2), (8.3) задачею безумовного екстремуму для функції Лагранжа задачі (8.2), (8.3).

Введемо функцію Лагранжа задачі (8.2), (8.3)

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \lambda_0 f^0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f^i(\boldsymbol{x})$$

змінних $(x_1,x_2,\dots,x_n,\lambda_0,\lambda_1,\dots,\lambda_m) = (\boldsymbol{x},\boldsymbol{\lambda}) \in E^{n+m+1}$.

Справедлива теорема.

Теорема 8.3 (необхідні умови умовного екстремуму). *Якщо $\boldsymbol{x}^*$ – точка локального мінімуму або максимуму функції $f^0(\boldsymbol{x})$ за умов $f^i(\boldsymbol{x})=0$, $i=1,\dots,m$, то необхідно існують змінні $(\lambda_0^*,\lambda_1^*,\dots,\lambda_m^*)=\boldsymbol{\lambda}^* \neq 0$, які називаються множниками Лагранжа, такі, що*

$$\frac{\partial L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)}{\partial x_j} = \lambda_0^* \frac{\partial f^0(\boldsymbol{x}^*)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i^* \frac{\partial f^i(\boldsymbol{x}^*)}{\partial x_j} = 0, \quad j=1,\dots,n, \qquad (8.5)$$

або

$$\lambda_0^* \nabla f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f^i(\boldsymbol{x}^*) = 0,$$

тобто вектори $\nabla f^0(\boldsymbol{x}^*), \nabla f^1(\boldsymbol{x}^*), \dots, \nabla f^m(\boldsymbol{x}^*)$ є лінійно залежними.

<u>Доведення</u>. Припустимо противне, тобто, що вказані вектори $\nabla f^0(\boldsymbol{x}^*)$, $\nabla f^1(\boldsymbol{x}^*), \dots, \nabla f^m(\boldsymbol{x}^*)$ лінійно незалежні. Тоді функціональна матриця $\left\| \frac{\partial \boldsymbol{f}_i}{\partial x_j} \right\|$, $i=0,1,\dots,m$, $j=1,\dots,n$, має ранг $m+1$, а, отже існує мінор цієї матриці $m+1$ порядку, відмінний від нуля, який можна вважати якобіаном системи функцій $f^0(\boldsymbol{x}) - f^0(\boldsymbol{x}^*) - t, f^1(\boldsymbol{x}), \dots, f^m(\boldsymbol{x})$ по деякій підмножині $m+1$ змінних множини змінних $x_1,\dots,x_n$ в точці $\boldsymbol{x}^*$, $t=0$.

Тоді за теоремою про неявні функції система

$$\begin{cases} f^0(x) - f^0(x^*) - t = 0, \\ f^1(x) = 0, \\ \dots\dots\dots\dots\dots \\ f^m(x) = 0 \end{cases}$$

має розв'язок при всіх t із околу точки $t=0$, тобто таких, що $|t| < t_0$, де t_0 – досить мале число.

Отже, існує вектор-функція $\boldsymbol{x} = \boldsymbol{x}(t) = (\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_n(t))$, яка визначена і диференційовна при всіх t, $|\boldsymbol{t}| < t_0$, і така, що $\boldsymbol{x}(0) = \boldsymbol{x}^*$, $f^0(\boldsymbol{x}(t)) = f^0(\boldsymbol{x}^*) + t$ і при $i = 1,...,m$ $f^i(\boldsymbol{x}(t)) = 0$. Звідси маємо $\forall t$, $0 < t \le t_0$

$$f^0(\boldsymbol{x}(t)) = f^0(\boldsymbol{x}^*) + t > f^0(\boldsymbol{x}^*) > f^0(\boldsymbol{x}^*) - t = f^0(\boldsymbol{x}(-t)).$$

що суперечить умові теореми про те, що $\boldsymbol{x}^*$ є точкою локального мінімуму або максимуму. Отже, наше припущення невірне і умови (8.5) мають місце.

Теорему доведено.

Таким чином, підозрілими на умовний екстремум можуть бути лише ті точки $\bar{\boldsymbol{x}}$, для яких існують множники $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_0, \bar{\lambda}_1, \ldots, \bar{\lambda}_m) \ne 0$, такі що точка $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}) \in E^{n+m+1}$ задовольняє систему $m+n$ рівнянь

$$\begin{cases} \lambda_0 \dfrac{\partial f^0(\boldsymbol{x})}{\partial \boldsymbol{x}_j} + \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial f^i(\boldsymbol{x})}{\partial \boldsymbol{x}_j} = 0, \quad j = 1, \ldots, n, \\ f^i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m. \end{cases} \tag{8.6}$$

Зауважимо, що якщо $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ розв'язок системи (8.6), то $(\alpha \bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ $\forall \alpha > 0$ також є розв'язком цієї системи.

Тому множники $\boldsymbol{\lambda}$ можна підпорядкувати якій-небудь додатковій умові нормування, наприклад,

$$\lambda_0 \ge 0, \ \left\| \boldsymbol{\lambda} \right\|^2 = \sum_{i=0}^{m} \lambda_i^2 = 1. \tag{8.7}$$

Якщо система (8.6) має розв'язки $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$, такі що $\bar{\lambda}_0 \ne 0$, то задачу мінімізації

$$f^0(\boldsymbol{x}) \to \min, \tag{8.8}$$

за умов

$$f^i(\boldsymbol{x}) = 0, \ i = 1, \ldots, m, \tag{8.9}$$

називають невиродженою в точці $\bar{\boldsymbol{x}}$.

У невиродженій задачі умову нормування (8.7) можна замінити більш простою умовою $\bar{\lambda}_0 = 1$. Зауважимо, що для невиродженості задачі (8.8), (8.9) в точці $\boldsymbol{x}$ достатньо, щоб вектори $\nabla f^1(\bar{\boldsymbol{x}}), \nabla f^2(\bar{\boldsymbol{x}}), \ldots, \nabla f^m(\bar{\boldsymbol{x}})$, були лінійно незалежними, тобто, щоб рівність

$$\alpha_1 \nabla f^1(\bar{\boldsymbol{x}}) + \alpha_2 \nabla f^2(\bar{\boldsymbol{x}}) + \ldots + \alpha_m \nabla f^m(\bar{\boldsymbol{x}}) = 0$$

була можлива тільки при $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$.

Умови (8.6) з умовою нормування (8.7) (або $\lambda_0 = 1$ у невиродженому випадку) визначають систему $n+m+1$ рівнянь з $n+m+1$ невідомими $(\boldsymbol{x}, \boldsymbol{\lambda}) = = (x_1, x_2, \ldots, x_n, \lambda_0, \lambda_1, \ldots, \lambda_m)$. Розв'язавши її, ми знайдемо точки $\bar{\boldsymbol{x}}$, підозрілі на умовний екстремум, і відповідні їм множники Лагранжа $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_0, \bar{\lambda}_1, ..., \bar{\lambda}_m) \ne 0$.

Щоб з'ясувати, чи буде в дійсності в точках $\bar{\boldsymbol{x}}$ мінімум або максимум, треба застосувати **достатні умови мінімуму (максимуму)** до функції Лагранжа по змінній $\boldsymbol{x}$, які можна сформулювати так:

Теорема 8.4. *Нехай:*

1) $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ *задовольняє систему* (7.6);

2) *в точці* $\bar{\mathbf{x}}$ *задача невироджена, тобто* $\bar{\lambda}_0 > 0$;

3) *функція* $L(\mathbf{x},\boldsymbol{\lambda})$ *в околі точки* $\bar{\mathbf{x}}$ *двічі диференційовна по* $\mathbf{x}$ *і має неперервні всі частинні похідні другого порядку в самій точці* $\bar{\mathbf{x}}$;

4) *гессіан* $H_L(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ *по* $\mathbf{x}$ *функції Лагранжа* $L(\mathbf{x},\boldsymbol{\lambda})$ *в точці* $(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ *додатно (від'ємно) визначена матриця.*

Тоді точка $\bar{\mathbf{x}}$ *є точкою локального мінімуму (максимуму) функції* $f^0(\mathbf{x})$ *при умовах* (7.3).

Приклад 8.1. *Нехай у просторі* E^n *дано* m *точок* $\mathbf{x}^i = (x_1^i,\dots,x_n^i)$, $i=1,\dots,m$. *Потрібно знайти точку* $\mathbf{x}\in E^n$, *сума квадратів відстаней якої від даних точок була б мінімальною.*

За умовою задачі треба в E^n мінімізувати функцію

$$f(\mathbf{x}) = \sum_{i=1}^{m} \left\| \mathbf{x} - \mathbf{x}^i \right\|^2 = \sum_{i=1}^{m} (\mathbf{x} - \mathbf{x}^i, \mathbf{x} - \mathbf{x}^i).$$

Знаходимо її градієнт

$$\nabla f(\mathbf{x}) = 2\sum_{i=1}^{m} (\mathbf{x} - \mathbf{x}^i)$$

та записуємо необхідні умови екстремуму $\nabla f(\mathbf{x}) = 0$ або

$$2\sum_{i=1}^{m} (\mathbf{x} - \mathbf{x}^i) = 0,$$

звідки отримуємо

$$m\mathbf{x} - \sum_{i=1}^{m} \mathbf{x}^i = 0.$$

Отже, стаціонарною буде точка

$$\mathbf{x} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}^i \equiv \mathbf{x}^0.$$

Знайдемо гессіан $H_f(\mathbf{x})$ функції $f(\mathbf{x})$ в точці $\mathbf{x}^0$. Маємо

$$\frac{\partial f}{\partial x_j} = 2\sum_{i=1}^{m}(x_j - x_j^i) = 2m x_j - 2\sum_{i=1}^{m} x_j^i,$$

$$\frac{\partial^2 f}{\partial x_k \partial x_j} = 2m\delta_{kj}, \quad k, j = 1,\dots,n,$$

де δ_{kj} — символ Кронекера

$$\delta_{kj} = \begin{cases} 1, & k = j, \\ 0, & k \neq j, \end{cases} \quad k, j = 1,\dots,n.$$

Тоді $\forall \mathbf{x} \in E^n$ маємо $H_f(\mathbf{x}) = 2m\mathbf{E}$, де $\mathbf{E}$ — одинична матриця n-го порядку, і $\forall \mathbf{y} \in E^n$ $(\mathbf{y} \neq \mathbf{0})$ $\mathbf{y}^{\mathrm{T}} H_f(\mathbf{x}^0)\mathbf{y} = 2m\|\mathbf{y}\|^2 > 0$.

За достатніми умовами в точці $\boldsymbol{x}^0$ функція $f(\boldsymbol{x})$ має строгий локальний мінімум.

Оскільки всі головні мінори матриці $H_f(\boldsymbol{x})=2m\mathbf{E}\quad \forall \boldsymbol{x}\in E^n$ додатні $(m>0)$ $\Delta_1=2m>0$, $\Delta_2=4m^2>0,\ldots,$ $\Delta_n=(2m)^n>0$, то функція $f(\boldsymbol{x})$ є опуклою в E^n, і точка

$$\boldsymbol{x}^0=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}^i$$

є точкою глобального мінімуму функції $f(\boldsymbol{x})$.

Приклад 8.2. *Нехай треба знайти на n-вимірній сфері*

$$X=\{\boldsymbol{x}\in E^n : ||\boldsymbol{x}||^2=1\}$$

точку, сума квадратів відстаней від якої до m даних точок $\boldsymbol{x}^1,\ldots,\boldsymbol{x}^m\in E^n$ *була б мінімальною.*

Задача зводиться до мінімізації функції

$$f^0(\boldsymbol{x})=\sum_{i=1}^{m}\left\|\boldsymbol{x}-\boldsymbol{x}^i\right\|^2$$

за умови $||\boldsymbol{x}||^2=1$ або $(\boldsymbol{x},\boldsymbol{x})=1$.

Перепишемо її в скалярному вигляді:

$$f^0(\boldsymbol{x})=\sum_{i=1}^{m}\sum_{k=1}^{n}\left(x_k-x_k^i\right)^2\to\min,$$

$$f^1(\boldsymbol{x})=\sum_{k=1}^{n}(x_k)^2-1=0.$$

Складемо функцію Лагранжа цієї задачі: $L(\boldsymbol{x},\boldsymbol{\lambda})=\lambda_0 f^0(\boldsymbol{x})+\lambda_1 f^1(\boldsymbol{x})=$

$$=\lambda_0\sum_{i=1}^{m}\left\|\boldsymbol{x}-\boldsymbol{x}^i\right\|^2+\lambda_1((\boldsymbol{x},\boldsymbol{x})-1)=\lambda_0\sum_{i=1}^{m}\sum_{k=1}^{n}\left(x_k-x_k^i\right)^2+\lambda_1(\sum_{k=1}^{n}(x_k)^2-1).$$

Запишемо необхідні умови (7.6). Маємо $\nabla_x L(\boldsymbol{x},\boldsymbol{\lambda})=\lambda_0\nabla f^0(\boldsymbol{x})+\lambda_1\nabla f^1(\boldsymbol{x})=$

$$=\lambda_0 2\sum_{i=1}^{m}(\boldsymbol{x}-\boldsymbol{x}^i)+\lambda_1 2\boldsymbol{x}=2m\lambda_0\boldsymbol{x}-2\lambda_0\sum_{i=1}^{m}\boldsymbol{x}^i+2\lambda_1\boldsymbol{x},$$

де $\lambda_0\geq 0$, $\lambda_0^2+\lambda_1^2=1$.

Оскільки точка

$$\boldsymbol{x}^0=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}^i$$

є стаціонарною точкою функції $f^0(\boldsymbol{x})$ (див. приклад 8.1), то

$$\sum_{i=1}^{m}\boldsymbol{x}^i=m\boldsymbol{x}^0,$$

і остаточно отримаємо

$$\begin{cases} \nabla_x L(\boldsymbol{x},\boldsymbol{\lambda}) = 2\lambda_0 m(\boldsymbol{x} - \boldsymbol{x}^0) + 2\lambda_1 \boldsymbol{x} = 0, \\ (\boldsymbol{x},\boldsymbol{x}) = 1, \\ \lambda_0 \ge 0,\ \lambda_0^2 + \lambda_1^2 = 1. \end{cases}$$

З'ясуємо, чи може бути задача виродженою. Із умови $(\boldsymbol{x},\boldsymbol{x}) = 1$ випливає, що $\boldsymbol{x} \neq 0$. Тоді, якщо покласти $\lambda_0 = 0$, перше рівняння дає $\lambda_1 = 0$, що суперечить умові, що вектор $\boldsymbol{\lambda}$ повинен бути ненульовим.

Отже задача невироджена і $\lambda_0 > 0$. Тоді покладемо $\lambda_0 = 1$ і відкинемо умову нормування. Отримаємо

$$\begin{cases} 2m\boldsymbol{x} - 2m\boldsymbol{x}^0 + 2\lambda_1 \boldsymbol{x} = 0, \\ (\boldsymbol{x},\boldsymbol{x}) = 1. \end{cases} \tag{8.11}$$

З першого рівняння маємо $(m+\lambda_1)\boldsymbol{x} - m\boldsymbol{x}^0 = 0$, тобто $\boldsymbol{x} = m\boldsymbol{x}^0/(m+\lambda_1)$. Підставляючи цей вираз для $\boldsymbol{x}$ в друге рівняння, отримаємо після нескладних перетворень: $m^2\|\boldsymbol{x}^0\|^2 = (m+\lambda_1)^2$, або $m\|\boldsymbol{x}^0\| = |m+\lambda_1|$. Розглянемо можливі випадки:

1) $m\|\boldsymbol{x}^0\| = m+\lambda_1$, тоді

$$\lambda_1^{(1)} = m(\|\boldsymbol{x}^0\| - 1) \text{ і } x^{(1)} = \frac{1}{\|\boldsymbol{x}^0\|}\boldsymbol{x}^0,\ \boldsymbol{x}^0 \neq 0;$$

2) $m\|\boldsymbol{x}^0\| = -m-\lambda_1$, тоді

$$\lambda_1^{(2)} = -m(\|\boldsymbol{x}^0\| + 1) \text{ і } x^{(2)} = -\frac{1}{\|\boldsymbol{x}^0\|}\boldsymbol{x}^0,\ \boldsymbol{x}^0 \neq 0.$$

Обчислимо гессіан функції Лагранжа по $\boldsymbol{x}$ в знайдених точках $(\boldsymbol{x}^{(1)},\lambda_1^{(1)})$, $(\boldsymbol{x}^{(2)},\lambda_1^{(2)})$. Маємо $\forall \boldsymbol{x} \in E^n$ $H_L^x(\boldsymbol{x},\boldsymbol{\lambda}) = 2m\boldsymbol{E} + 2\lambda_1\boldsymbol{E}$. В точці $(\boldsymbol{x}^{(1)},\lambda_1^{(1)})$

$$H_L^x(\boldsymbol{x}^{(1)},\lambda_1^{(1)}) = 2m\boldsymbol{E} + 2(\|\boldsymbol{x}^0\| - 1)m\boldsymbol{E} = 2m\|\boldsymbol{x}^0\|\boldsymbol{E}.$$

Головні мінори цієї матриці додатні $\Delta_i = (2m\|\boldsymbol{x}^0\|)^i > 0$, тому в точці $\boldsymbol{x}^{(1)}$ функція $f^0(\boldsymbol{x})$ має строгий локальний умовний мінімум.

В точці $(\boldsymbol{x}^{(2)},\lambda_1^{(2)})$

$$H_L^x(\boldsymbol{x}^{(2)},\lambda_1^{(2)}) = 2m\boldsymbol{E} - 2(\|\boldsymbol{x}^0\| + 1)m\boldsymbol{E} = -2m\|\boldsymbol{x}^0\|\boldsymbol{E}.$$

Головні мінори цієї матриці змінюють знак:

$\Delta_1 = -2m\|\boldsymbol{x}^0\| < 0$, $\Delta_2 = (-2m\|\boldsymbol{x}^0\|)^2 > 0, \ldots, \Delta_n = (2m\|\boldsymbol{x}^0\|)^n$ ($\Delta_n < 0$ при $n = 2k-1$, $\Delta_n > 0$ при $n = 2k$), тому в точці $\boldsymbol{x}^{(2)}$ функція $f^0(\boldsymbol{x})$ має строгий локальний умовний максимум.

Оскільки X — замкнена обмежена множина і $f^0(\boldsymbol{x})$ неперервна на X, то $f^0(\boldsymbol{x})$ досягає на ній свого абсолютного мінімуму і абсолютного максимуму. Точки абсолютного максимуму і абсолютного мінімуму повинні бути розв'язками системи (8.10).

Оскільки при $\boldsymbol{x}^0 \neq 0$ розв'язків тільки два $\boldsymbol{x}^{(1)}$ і $\boldsymbol{x}^{(2)}$, то $\boldsymbol{x}^{(1)} = \boldsymbol{x}^0/\|\boldsymbol{x}^0\|$ є точкою глобального мінімуму, а $\boldsymbol{x}^{(2)} = -\boldsymbol{x}^0/\|\boldsymbol{x}^0\|$ — точкою глобального максимуму функції $f^0(\boldsymbol{x})$ на множині X при $\boldsymbol{x}^0 \neq 0$.

Розглянемо випадок $\boldsymbol{x}^0 = 0$. Тоді система (8.11) набуває вигляду

$$\begin{cases} \lambda_0 = 1, \\ (m + \lambda_1)\boldsymbol{x} = 0, \\ (\boldsymbol{x}, \boldsymbol{x}) = 1, \end{cases}$$

розв’язками якої є $\lambda_0 = 1$, $\lambda_1 = -m$, $\boldsymbol{x}$ □ довільна точка, для якої $||\boldsymbol{x}|| = 1$.

Отже, необхідні умови не дали ніякої корисної інформації, підозрілими на екстремум є всі точки одиничної сфери.

Розглянемо $\forall \boldsymbol{x} \in X$ (при $\boldsymbol{x}^0 = 0$) значення $f^0(\boldsymbol{x})$:

$$f^0(\boldsymbol{x}) = \sum_{i=1}^{m} \left\| \boldsymbol{x} - \boldsymbol{x}^i \right\|^2 = \sum_{i=1}^{m} (\boldsymbol{x} - \boldsymbol{x}^i, \boldsymbol{x} - \boldsymbol{x}^i) = \sum_{i=1}^{m} \left[(\boldsymbol{x}, \boldsymbol{x}) - 2(\boldsymbol{x}, \boldsymbol{x}^i) + (\boldsymbol{x}^i, \boldsymbol{x}^i) \right] =$$

$$= m - 2(\boldsymbol{x}, \sum_{i=1}^{m} \boldsymbol{x}^i) + \sum_{i=1}^{m} ||\boldsymbol{x}^i||^2 = m - 2(\boldsymbol{x}, \boldsymbol{x}^0) + \sum_{i=1}^{m} ||\boldsymbol{x}^i||^2 = m + \sum_{i=1}^{m} ||\boldsymbol{x}^i||^2 = \text{const}.$$

Отже, при $\boldsymbol{x}^0 = 0$ $f^0(\boldsymbol{x}) = \text{const}$ на множині X, тобто у всіх точках цієї множини $f^0(\boldsymbol{x})$ досягає свого глобального мінімуму (в той же час і максимуму).

Приклад 7.3. *Нехай треба знайти точки екстремуму функції* $f^0(x) = x$ *на множині* $X = \{(x, y) \in E^2 : x^3 - y^2 = 0\}$.

Застосуємо метод множників Лагранжа. Будуємо функцію Лагранжа задачі $L(x, y, \lambda_0, \lambda_1) = \lambda_0 x + \lambda_1 (x^3 - y^2)$ та записуємо необхідні умови екстремуму

$$\begin{cases} \dfrac{\partial L}{\partial x} = \lambda_0 + 3\lambda_1 x^2 = 0, \\ \dfrac{\partial L}{\partial y} = -2\lambda_1 y = 0, \\ x^3 - y^2 = 0, \\ \lambda_0 \geq 0, \ \lambda_0^2 + \lambda_1^2 = 1. \end{cases}$$

Якщо припустити, що $\lambda_0 > 0$, то з першого рівняння будемо мати

$\lambda_0 = -3\lambda_1 x^2$, де $\lambda_1 < 0$, $x > 0$.

Тоді з рівняння $x^3 - y^2 = 0$ випливає, що $y \neq 0$.

Отже, $\lambda_1 < 0$, $y \neq 0$ і система не буде сумісною, оскільки не задовольняється рівняння $-2\lambda_1 y = 0$. Виходить, що система сумісна лише при $\lambda_0 = 0$. Тоді $\lambda_1 = 1$ і розв’язком системи відносно x і y буде точка $(0, 0)$.

Таким чином, задача є виродженою в точці $(0, 0)$, яка є підозрілою на екстремум.

Виразивши x через y із обмеження $x^3 - y^2 = 0$, отримаємо $f^0(x) = x = y^{2/3}$ при $-\infty < y < \infty$. Ясно, що $(0, 0)$ є точкою глобального мінімуму функції $f^0(x)$ на множині X.

Геометрична інтерпретація методу Лагранжа

Нехай $\boldsymbol{x} = (x_1, x_2) \in E^2$. Розглянемо задачу

$z = f^0(x_1, x_2) \to \min$,

$f^1(x_1, x_2) = 0$.

Побудуємо для неї функцію Лагранжа

$$L(x_1, x_2, \lambda_0, \lambda_1) = \lambda_0 f^0(x_1, x_2) + \lambda_1 f^1(x_1, x_2)$$

І запишемо необхідні умови умовного екстремуму:

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = \lambda_0 \dfrac{\partial f^0}{\partial x_1} + \lambda_1 \dfrac{\partial f^1}{\partial x_1} = 0, \\ \dfrac{\partial L}{\partial x_2} = \lambda_0 \dfrac{\partial f^0}{\partial x_2} + \lambda_1 \dfrac{\partial f^1}{\partial x_2} = 0, \\ f^1(x_1, x_2) = 0, \\ \lambda_0 \geq 0, \ \lambda_0^2 + \lambda_1^2 = 1. \end{cases} \tag{8.12}$$

Позначимо через $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ довільний розв'язок цієї системи (звичайно, якщо вона взагалі має розв'язки).

Якщо задача вироджена в точці $\bar{\boldsymbol{x}}$, то $\bar{\lambda}_0 = 0$. Тоді $|\bar{\lambda}_1| = 1 > 0$ і система необхідних умов набуває вигляду

$$\begin{cases} \nabla f^1(x_1, x_2) = 0, \\ f^1(x_1, x_2) = 0, \end{cases} \tag{8.13}$$

для довільної цільової функції $f^0(x_1,x_2) \in C^1$.

Отже, незалежно від вигляду функції $f^0(x_1,x_2)$, точки її умовного екстремуму (якщо вони існують) повинні лежати на лінії $f^1(x_1,x_2) = 0$ і градієнт функції обмеження $\nabla f^1(x_1,x_2)$ повинен в цих точках дорівнювати нулю, тобто, якщо розглядати функцію $y = f^1(x_1,x_2)$, то точки її умовного екстремуму повинні лежати на лінії нульового рівня цієї функції і бути її стаціонарними точками.

Нехай тепер задача невироджена в точці $\bar{\boldsymbol{x}}$. Тоді можна покласти $\bar{\lambda}_0 = 1$ і відкинути умову нормування. Система необхідних умов (8.12) набуває вигляду

$$\begin{cases} \dfrac{\partial f^0}{\partial x_1} + \lambda_1 \dfrac{\partial f^1}{\partial x_1} = 0, \\ \dfrac{\partial f^0}{\partial x_2} + \lambda_1 \dfrac{\partial f^1}{\partial x_2} = 0, \\ f^1(x_1, x_2) = 0. \end{cases}$$

З перших двох рівнянь отримуємо

$$\lambda_1 = -\frac{\dfrac{\partial f^0}{\partial x_1}}{\dfrac{\partial f^1}{\partial x_1}} = -\frac{\dfrac{\partial f^0}{\partial x_2}}{\dfrac{\partial f^1}{\partial x_2}}$$

звідки

$$-\frac{\dfrac{\partial f^0}{\partial x_1}}{\dfrac{\partial f^0}{\partial x_2}} = -\frac{\dfrac{\partial f^1}{\partial x_1}}{\dfrac{\partial f^1}{\partial x_2}}.$$

В той же час за правилом диференціювання неявних функцій $x_2^0(x_1)$ і $x_2^1(x_1)$, які визначаються неявно співвідношеннями $f^0(x_1,x_2) = h$ (h=const) і $f^1(x_1,x_2) = 0$

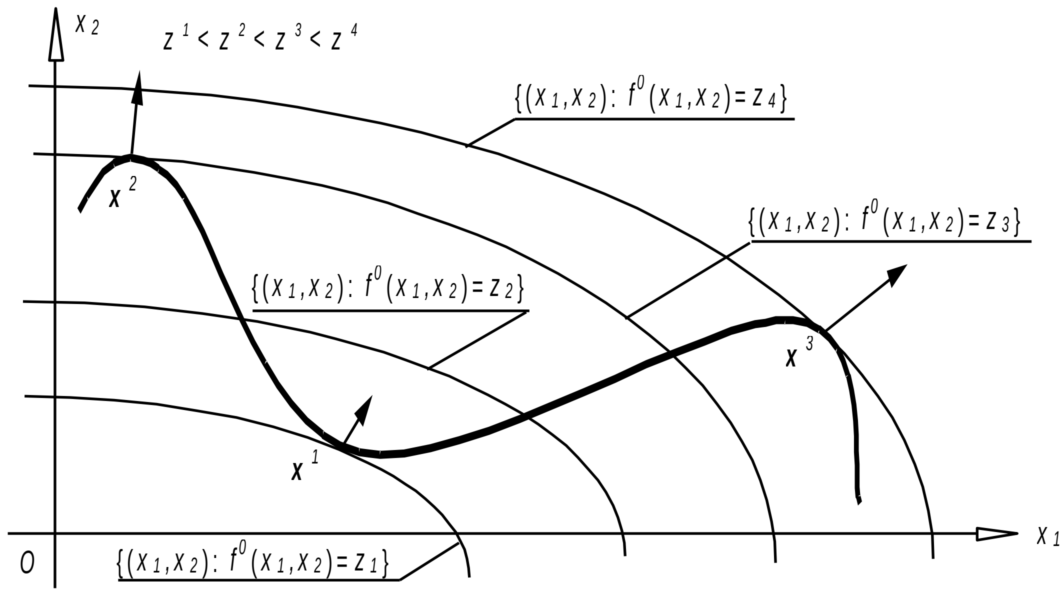


Рис. 8.1

маємо

$$\frac{dx_2^0}{dx_1} - \frac{\frac{\partial f^0}{\partial x_1}}{\frac{\partial f^0}{\partial x_2}}, \quad \frac{dx_2^1}{dx_1} = -\frac{\frac{\partial f^1}{\partial x_1}}{\frac{\partial f^1}{\partial x_2}},$$

Це означає, що кутові коефіцієнти дотичних до ліній рівня $f^0(x_1,x_2)=z^k$ і до функції, заданої неявно рівнянням $f^1(x_1,x_2)=0$ співпадають у точках умовного екстремуму (див. рис. 8.1).

В точці x^1 досягається локальний мінімум, в точці x^2 — локальний максимум, в точці x^3 — глобальний максимум.

Метод множників Лагранжа у випадку обмежень-нерівностей

Метод множників Лагранжа може бути поширений на відшукання умовного екстремуму функції і у випадку, коли частина обмежень є обмеженнями-нерівностями. Нехай маємо задачу

$$f^0(\boldsymbol{x}) \to \min(\max), \tag{8.14}$$

$$f^i(\boldsymbol{x}) = 0,\ i=1,\dots,m, \tag{8.15}$$

$$h^l(\boldsymbol{x}) \le 0,\ l=1,\dots,k, \tag{8.16}$$

$$\boldsymbol{x} \in E^n.$$

Будемо вважати, що функції $f^0(\boldsymbol{x})$, $f^i(\boldsymbol{x})$, $h^l(\boldsymbol{x})$, $i=1,\dots,m,\ l=1,\dots,k,$ визначені і диференційовні у всіх точках простору $\boldsymbol{x} \in E^n$.

Введемо нові допоміжні змінні $\boldsymbol{y}=(y_1,\dots,y_k)$, за допомогою яких систему нерівностей (7.16) приведемо до системи рівнянь

$$h^l(\boldsymbol{x}) + y_l^2 = 0, \quad l = 1,\ldots,k.. \qquad (8.17)$$

Тоді задача відшукання екстремумів функції $f^0(\boldsymbol{x})$ при умовах (8.15), (8.16) зводиться до рівносильної задачі відшукання екстремумів тієї ж функції при обмеженнях (8.15), (8.17). Під еквівалентністю цих задач розуміється те, що якщо $\boldsymbol{x}^*$ — точка локального мінімуму (максимуму) функції $f^0(\boldsymbol{x})$ при обмеженнях (8.15), (8.16), то точка $\bar{\boldsymbol{x}}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*)$, де $\boldsymbol{y}^* = (y_1^*,\ldots,y_k^*)$, $y_l^* = (-h^l(\boldsymbol{x}^*))^{1/2}$, $l = 1,\ldots,k$, буде точкою локального мінімуму (максимуму) функції $f^0(\boldsymbol{x})$ при обмеженнях (8.15), (8.17), і, навпаки, якщо $\bar{\boldsymbol{x}}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*)$ — точка локального мінімуму (максимуму) $f^0(\boldsymbol{x})$ при обмеженнях (8.15), (8.17), то $\boldsymbol{x}^*$ — точка локального мінімуму (максимуму) $f^0(\boldsymbol{x})$ при обмеженнях (8.15), (8.16).

Для відшукання екстремумів функції $f^0(\boldsymbol{x})$ при умовах-рівностях (8.15), (8.17) введемо функцію Лагранжа

$$L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{\mu}) = \lambda_0 f^0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f^i(\boldsymbol{x}) + \sum_{l=1}^{k} \mu_l (h^l(\boldsymbol{x}) + y_l^2)$$

і запишемо необхідні умови екстремуму

$$\begin{cases} \frac{\partial L}{\partial x_j} = \lambda_0 \frac{\partial f^0}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial f^i}{\partial x_j} + \sum_{l=1}^{k} \mu_l \\ \frac{\partial L}{\partial y_l} = 2 y_l \mu_l = 0, \quad l = 1,\ldots,k, \\ \frac{\partial L}{\partial \lambda_i} = f^i(x) = 0, \quad i = 1,\ldots,m, \\ \frac{\partial L}{\partial \mu_l} = h^l(x) + y_l^2 = 0, \quad l = 1,\ldots,k, \\ \lambda_0 \geq 0, \quad \sum_{i=0}^{m} \lambda_i^2 + \sum_{l=1}^{k} \mu_l^2 = 1. \end{cases} \qquad (8.18)$$

Розв'язавши систему (8.18), знайдемо точки $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$, підозрілі на умовний екстремум. Дослідивши їх за допомогою достатніх умов, остаточно отримаємо точки $\bar{\boldsymbol{x}}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*)$ умовного локального мінімуму та умовного локального максимуму функції $f^0(\boldsymbol{x})$ при умовах (8.15), (8.17).

Відкинувши змінні $\boldsymbol{y}^* = (y_1^*,\ldots,y_k^*)$, отримаємо точки екстремуму $f^0(\boldsymbol{x})$ при умовах (8.15), (8.16).

Приклад 8.4. *В n-вимірній одиничній кулі* $X = \{\boldsymbol{x} \in E^n : \|\boldsymbol{x}\|^2 = (\boldsymbol{x},\boldsymbol{x}) \leq 1\}$ *знайти точку, сума квадратів відстаней від якої до m даних точок* $\boldsymbol{x}^1,\ldots,\boldsymbol{x}^m \in E^n$ *була б мінімальною.*

За умовою треба мінімізувати функцію

$$f^0(\boldsymbol{x}) = \sum_{i=1}^{m} \left\| \boldsymbol{x} - \boldsymbol{x}^i \right\|^2$$

при обмеженні $(\boldsymbol{x},\boldsymbol{x}) \leq 1$.

В прикладі 8.1 було показано, що глобальний мінімум функції $f^0(\boldsymbol{x})$ у всьому просторі досягається в точці $\boldsymbol{x}^0 = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}^i$. Тому при $\|\boldsymbol{x}^0\| \leq 1$ точка $\boldsymbol{x}^0$ буде розв'язком сформульованої задачі.

Розглянемо випадок $\|\boldsymbol{x}^0\| > 1$. Введемо змінну $y \in E^1$, за допомогою якої зведемо нерівність $(\boldsymbol{x},\boldsymbol{x}) \leq 1$ до рівності

$$y^2 + (\boldsymbol{x},\boldsymbol{x}) - 1 = 0 \qquad (8.19)$$

і розглянемо задачу мінімізації функції $f^0(\boldsymbol{x})$ в просторі змінних $\bar{\boldsymbol{x}} = (\boldsymbol{x}, y) = (x_1, \dots, x_n, y) \in E^{n+1}$ за умови (8.19). Будуємо функцію Лагранжа цієї задачі

$$L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}) = \lambda_0 \sum_{i=1}^{m} \left\| \boldsymbol{x} - \boldsymbol{x}^i \right\|^2 + \lambda_1 ((\boldsymbol{x}, \boldsymbol{x}) + y^2 - 1)$$

і записуємо необхідні умови умовного екстремуму:

$$\begin{cases} \nabla_x L(\bar{x}, \lambda) = 2m\lambda_0 x - 2\lambda_0 \sum_{i=1}^{m} x^i + 2\lambda_1 x = 2\lambda_0 m(x - x^0 \\ \dfrac{\partial L(\bar{x}, \lambda)}{\partial y} = 2\lambda_1 y = 0, \\ (x, x) + y^2 - 1 = 0, \\ \lambda_0 \geq 0, \ \lambda_0^2 + \lambda_1^2 = 1. \end{cases}$$

Розглянемо випадок $\lambda_0 = 0$. З умови нормування отримаємо $|\lambda_1| = 1 \neq 0$, а тому з першого та другого рівнянь буде $\boldsymbol{x} = 0$, $y = 0$. Але ці значення не задовольняють третє рівняння. Отже, при $\lambda_0 = 0$ система необхідних умов розв'язків не має.

Тому покладемо $\lambda_0 = 1 > 0$ і перепишемо систему необхідних умов, відкинувши умови нормування.

$$\begin{cases} (m + \lambda_1) x = m x^0, \ (\| x^0 \| > 1), \\ \lambda_1 y = 0, \\ \| x \|^2 + y^2 = 1. \end{cases}$$

Розглянемо два випадки: а) $\lambda_1 = 0$ і б) $\lambda_1 \neq 0$.

а) Нехай $\lambda_1 = 0$. Тоді з першого рівняння отримаємо $\boldsymbol{x} = \boldsymbol{x}^0$, а з другого випливає, що y □ довільне. Тому з останнього рівняння будемо мати $\|\boldsymbol{x}\| = \|\boldsymbol{x}^0\| = 1 - y^2 \leq 1$, що суперечить умові $\|\boldsymbol{x}^0\| > 1$.

Отже при $\lambda_1 = 0$ система несумісна.

б) Розглянемо випадок, коли $\lambda_1 \neq 0$. З другого рівняння отримаємо $y = 0$. Тому третє рівняння дає $\|\boldsymbol{x}\|^2 = 1$. Розв'язуючи перше рівняння, отримуємо $\boldsymbol{x} = m\boldsymbol{x}^0/(m+\lambda_1)$. Підставляючи значення $\boldsymbol{x}$ в умову $\|\boldsymbol{x}\|^2 = 1$, будемо мати $m^2 \|\boldsymbol{x}^0\|^2/(m+\lambda_1)^2 = 1$, звідки:

а) $m\|\boldsymbol{x}^0\| = m + \lambda_1$ або $\lambda_1^{(1)} = m(\|\boldsymbol{x}^0\| - 1) > 0$;

б) $m\|\boldsymbol{x}^0\| = -m - \lambda_1$ або $\lambda_1^{(2)} = -m(\|\boldsymbol{x}^0\| + 1) < 0$.

Відповідно отримуємо два розв'язки системи необхідних умов відносно $\boldsymbol{x}$:

а) $\bar{\boldsymbol{x}}^{(1)} = \left(\dfrac{\boldsymbol{x}^0}{\left\| \boldsymbol{x}^0 \right\|}, 0 \right)$ при $\lambda_1^{(1)} = m(\|\boldsymbol{x}^0\| - 1) > 0$;

та

б) $\bar{\boldsymbol{x}}^{(2)} = \left(-\dfrac{\boldsymbol{x}^0}{\left\| \boldsymbol{x}^0 \right\|}, 0 \right)$ при $\lambda_1^{(2)} = -m(\|\boldsymbol{x}^0\| + 1) < 0$.

Отже, екстремум функції $f^0(\boldsymbol{x})$ при умові (8.19) може досягатися лише в точках $\bar{\boldsymbol{x}}^{(1)}$ та $\bar{\boldsymbol{x}}^{(2)}$.

Запишемо гессіан функції Лагранжа по змінних $(\boldsymbol{x}, y) \in E^{n+1}$:

$$H_L^{(\boldsymbol{x}, y)}(x, y, \lambda_1) = \begin{pmatrix} 2(m + \lambda_1)\boldsymbol{E} & 0 \\ 0 & 2\lambda_1 \end{pmatrix}.$$

В точці $(\bar{\boldsymbol{x}}^{(1)}, \lambda_1^{(1)})$ маємо

$$H_L^{(x,y)}\left(\frac{\boldsymbol{x}^0}{\|\boldsymbol{x}^0\|},0,\lambda_1^{(1)}\right)=\begin{pmatrix}2m\|\boldsymbol{x}^0\|\boldsymbol{E} & 0\\ 0 & 2m(\|\boldsymbol{x}^0\|-1)\end{pmatrix}.$$

Всі послідовні головні мінори цієї матриці додатні: $\Delta_1=2m\|x^0\|>0$, $\Delta_2=(2m\|x^0\|)^2>0$, ... , $\Delta_n=(2m\|x^0\|)^n>0$, $\Delta_{n+1}=(2m\|x^0\|)^n(2m(\|x^0\|-1))>0$, тому в точці $(\bar{\boldsymbol{x}}^{(1)},\lambda_1^{(1)})$ функція $f^0(\boldsymbol{x})$ має локальний умовний мінімум.

В точці $(\bar{\boldsymbol{x}}^{(2)},\lambda_1^{(2)})$ маємо

$$H_L^{(x,y)}\left(-\frac{\boldsymbol{x}^0}{\|\boldsymbol{x}^0\|},0,\lambda_1^{(2)}\right)=\begin{pmatrix}-2m\|\boldsymbol{x}^0\|\boldsymbol{E} & 0\\ 0 & -2m(\|\boldsymbol{x}^0\|+1)\end{pmatrix}.$$

Послідовні головні мінори цієї матриці змінюють знак: $\Delta_1=-2m\|x^0\|<0$, $\Delta_2=(-2m\|x^0\|)^2>0$, ... , $\Delta_n=(-2m\|x^0\|)^n$ ($\Delta_n<0$ при $n=2k-1$, $\Delta_n>0$ при $n=2k$), $\Delta_{n+1}=(-2m\|x^0\|)^n(-2m(\|x^0\|+1))$, де $-2m(\|x^0\|+1)<0$. Тому в точці $(\bar{\boldsymbol{x}}^{(2)},\lambda_1^{(2)})$ функція $f^0(\boldsymbol{x})$ має локальний умовний максимум.

Отже, при $\|\boldsymbol{x}^0\|>1$ функція $f^0(\boldsymbol{x})$ має на одиничній кулі в точці $\boldsymbol{x}^0/\|\boldsymbol{x}^0\|$ локальний мінімум, а в точці $-\boldsymbol{x}^0/\|\boldsymbol{x}^0\|$ □ локальний максимум.

Оскільки точок, підозрілих на екстремум, тільки дві, а множина X опукла, замкнена і обмежена, то ці точки і будуть точками глобальних екстремумів функції $f^0(\boldsymbol{x})$ на одиничній кулі.

Методи одновимірної оптимізації

Постановка задачі

Задача одновимірної оптимізації полягає у відшуканні мінімуму (або максимуму) функції

$$y=f(x)\quad (x\in E^1)$$

на відрізку $[a,b]$.

Ця задача має самостійне значення і в той же час її доводиться розв'язувати при оптимізації функцій багатьох змінних.

Відразу треба зауважити, що класичний підхід до розв'язування задач одновимірної оптимізації, що грунтується на відшуканні коренів рівняння

$$f'(x)=0\quad (f(x)\in C^1),$$

далеко не завжди може бути реалізований на практиці. По-перше, в практичних задачах оптимізації часто взагалі невідомо, чи є функція $y=f(x)$ диференційовною, наприклад, якщо вона задана таблично. По-друге, задача розв'язування рівняння $f'(x)=0$ з обчислювальної точки зору має такий же порядок складності, як і вихідна задача. Ось чому є потреба в застосуванні методів оптимізації, відмінних від класичних, тобто таких, які не зв'язані з похідною.

Ми розглянемо деякі з таких методів одновимірної оптимізації стосовно, так званих, *унімодальних функцій*.

Означення 8.1. *Функцію* $y = f(x)$, *визначену на відрізку* $[a, b]$, *назвемо унімодальною, якщо вона має на ньому єдину точку мінімуму* $x^* \in [a, b]$ *і задовольняє умову*

$$\forall x_1, x_2 \text{ таких, що } a \le x_1 < x_2 \le x^*,\ f(x_1) > f(x_2);$$

$$\forall x_1, x_2 \text{ таких, що } x^* \le x_1 < x_2 \le b,\ f(x_1) < f(x_2).$$

Очевидно, що це означення "пристосоване" для задачі мінімізації. Для задачі максимізації відповідне означення змінюється елементарно.

Зауважимо, що унімодальна функція може не бути неперервною і може не бути опуклою.

Основна властивість унімодальної функції

Чисельні методи мінімізації унімодальної функції грунтуються на основній її властивості, яка безпосередньо випливає із її означення (див. рис. 7.1).

Нехай функція $y = f(x)$ *унімодальна на відрізку* $[a, b]$, *має на ньому мінімум в точці* x^* *і точки* l *та* r *з цього відрізка такі, що* $a < l < r < b$.

Тоді:

якщо $f(l) > f(r)$, *то* $x^* \in [l, b]$,

якщо $f(l) < f(r)$, *то* $x^* \in [a, r]$,

якщо $f(l) = f(r)$, *то* $x^* \in [l, r]$.

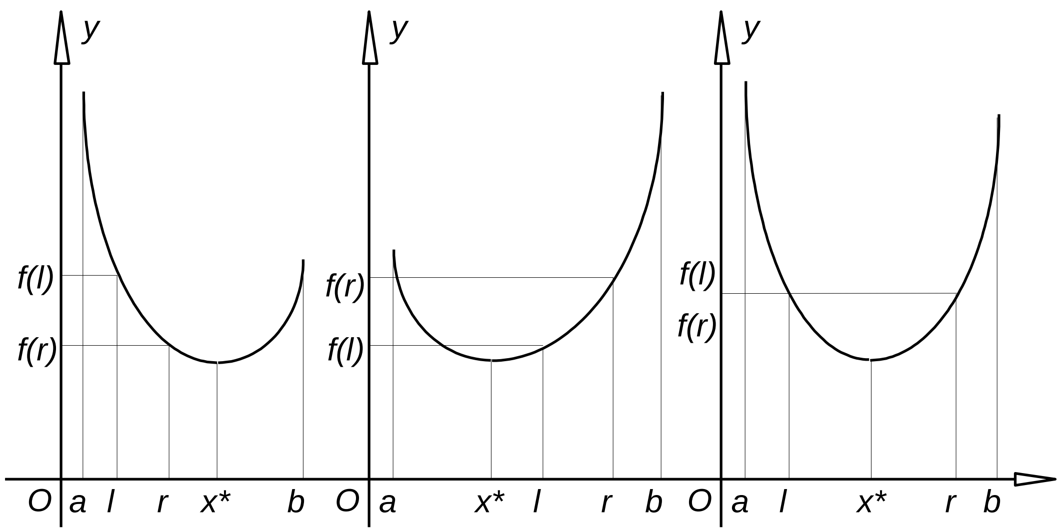


Рис. 7.1

Сформульована властивість дозволяє будувати послідовні алгоритми, на кожному кроці яких скорочується інтервал пошуку мінімуму.

Розглянемо деякі з них.

Метод дихотомії або ділення відрізків навпіл

Нехай x^* - точка мінімуму функції $f(x)$ на відрізку $[a, b]$. За вихідний інтервал пошуку вибирають відрізок $[a, b]$. Покладають

$$a^0 = a,\ b^0 = b.$$

Припустимо, що виконано s кроків алгоритму і обчислений інтервал $[a^s, b^s]$, який містить x^*. Розглянемо $s+1$ крок.

Ділимо відрізок $[a^s, b^s]$ навпіл і для деякого досить малого числа $\delta>0$ будуємо точки

$$l^s = (a^s+b^s-\delta)/2 \quad \text{та} \quad r^s = (a^s+b^s+\delta)/2.$$

За основною властивістю унімодальної функції покладаємо:

$$a^{s+1}=l^s, \quad b^{s+1}=b^s, \quad \text{якщо} \quad f(l^s)>f(r^s),$$
$$a^{s+1}=a^s, \quad b^{s+1}=r^s, \quad \text{якщо} \quad f(l^s)<f(r^s),$$
$$a^{s+1}=l^s, \quad b^{s+1}=r^s, \quad \text{якщо} \quad f(l^s)=f(r^s).$$

Отримаємо відрізок $[a^{s+1}, b^{s+1}]$, який містить точку x^*.

Після здійснення n кроків методу буде знайдено інтервал $[a^n, b^n]$, якому належить точка мінімуму x^* і при цьому, як неважко бачити,

$$b^n - a^n = (b-a)/2^n+(1-2^{-n})\delta.$$

Поклавши наближено $x^* \approx (b^n+a^n)/2$, матимемо похибку обчислення x^*, не більшу, ніж число $\varepsilon=(b^n-a^n)/2$. Число $y^n=f((b^n+a^n)/2)$ приймають за мінімальне значення функції $f(x)$ на інтервалі $[a,b]$.

Зауважимо, що на кожному кроці методу дихотомії потрібно обчислювати значення функції у двох точках.

Більш ефективними з обчислювальної точки зору в порівнянні з розглянутим є методи *золотого перетину* та *Фібоначчі*.

Метод золотого перетину

Нагадаємо спочатку, що означає - знайти точки золотого перетину відрізка. Нехай маємо проміжок $[a,b]$. Точкою золотого перетину відрізка $[a,b]$ називають таку точку цього відрізка, яка ділить його у середньому пропорціональному відношенні, тобто так, що відношення довжини всього відрізка до його більшої частини рівне відношенню більшої частини до меншої. Зауважимо, що таких точок існує дві на $[a,b]$. Позначимо через r ту з них, для якої виконується умова $r-a>b-r$. Невідому довжину відрізка $[a,r]$ позначимо через x, тоді довжина відрізка $[r,b]$ буде рівна $b-a-x$ і для визначення невідомої величини x отримаємо рівняння

$$\frac{b-a}{x}=\frac{x}{b-a-x} \quad \text{або} \quad x^2+(b-a)x-(b-a)^2 = 0.$$

Звідки $x_{1,2}=(b-a)\dfrac{\pm\sqrt{5}-1}{2}$. Оскільки $x>0$ і $b-a>0$, то підходить лише один корінь $x=(b-a)\dfrac{\sqrt{5}-1}{2}$. Отже, $r=a+(b-a)\dfrac{\sqrt{5}-1}{2}$. Інша точка золотого перетину відрізка $[a,b]$, яку ми позначимо через l, лежить на відстані x від точки b, тобто $l=b-(b-a)\dfrac{\sqrt{5}-1}{2}$. Не важко переконатись у тому, що точка l здійснює золотий переріз відрізка $[a,r]$, а точка r у свою чергу є точкою золотого перетину відрізка $[a,b]$.

Розглянемо тепер ітераційну схему алгоритму золотого перетину.

Нехай x^* – точка мінімуму функції $f(x)$ на відрізку $[a,b]$. На початку обчислень покладають $a^0 = a$, $b^0 = b$.

На s-у кроці визначають величини

$$l^s = b^s - \tau(b^s - a^s),$$
$$r^s = a^s + \tau(b^s - a^s),$$

де стала $\tau = \frac{\sqrt{5}-1}{2} \approx 0{,}618$. Покладають

$$a^{s+1} = l^s, \quad b^{s+1} = b^s, \quad \text{якщо } f(l^s) > f(r^s),$$
$$a^{s+1} = a^s, \quad b^{s+1} = r^s, \quad \text{якщо } f(l^s) \le f(r^s),$$

Ітерації продовжують доти, поки не буде виконуватись нерівність

$$b^n - a^n \le \varepsilon,$$

де $\varepsilon > 0$ – задане число, яке визначає похибку розв'язку задачі.

На кожному кроці *МЗП*, починаючи з 1-го, обчислюється лише одне значення функції $f(x)$, тому що одна з точок золотого перерізу на попередньому кроці здійснює золотий переріз проміжка на наступному кроці.

За наближений розв'язок задачі приймають

$$x^* = (a^n + b^n)/2, \quad y^* = f(x^*).$$

При розв'язуванні задачі максимізації функції $f(x)$ необхідно замінити її на функцію $-f(x)$.

Приклад 7.1. *Методом золотого перерізу з точністю $\varepsilon \le 0.05$ знайти мінімум функції $f(x) = e^{-x} - 2\cos x$ на проміжку $[0,1]$.*

Результати обчислень наведені в наступній таблиці.

Таблиця 7.1

a^s	b^s	l^s	r^s	$f(l^s)$	$f(r^s)$	знак	ε
0.000	1.000	0.382	0.618	1.1773	1.0911	<	0.309
0.000	0.618	0.236	0.382	1.1548	1.1733	>	0.191
0.236	0.618	0.382	0.472	1.1733	1.1576	<	0.118
0.236	0.472	0.326	0.382	1.1729	1.1733	>	0.073
0.326	0.472	0.382	0.416	1.1733	1.1697	<	0.045

Отже, $x^* = (0.326 + 0.416)/2 = 0.371$, $f(x^*) = 1.1739$.

Метод Фібоначчі

Зауважимо, що існує кілька ітераційних схем методу Фібоначчі, які в основному відрізняються швидкістю збіжності. Розглянемо одну з найбільш простих, не зупиняючись на питаннях збіжності. Отже, нехай x^* - точка мінімуму функції $f(x)$ на відрізку $[a,b]$.

Як і у вже розглянутих методах на початку обчислень покладають $a^0 = a$, $b^0 = b$.

На s-у кроці визначають величини

$$l^s = b^s - \tau_s(b^s - a^s),$$
$$r^s = a^s + \tau_s(b^s - a^s),$$

де $\tau_s = \dfrac{F_{N-s-1}}{F_{N-s}}$, $s = 0,\dots,N-3$, $\tau_{N-2} = \dfrac{1+\delta}{2}$, N — задане число ітерацій, $\varepsilon > 0$, F_j — числа Фібоначчі, що задаються рекурентним співвідношенням

$$F_j = F_{j-1} + F_{j-2},\quad j = 2, 3, \dots\quad F_0 = F_1 = 1.$$

Покладають

$$a^{s+1} = l^s,\quad b^{s+1} = b^s,\quad \text{якщо } f(l^s) > f(r^s),$$
$$a^{s+1} = a^s,\quad b^{s+1} = r^s,\quad \text{якщо } f(l^s) < f(r^s),$$
$$a^{s+1} = l^s,\quad b^{s+1} = r^s,\quad \text{якщо } f(l^s) = f(r^s).$$

За наближений розв'язок задачі приймають

$$x^* = (a^N + b^N)/2,\quad y^* = f(x^*).$$

Для *МФ* у випадку заздалегідь фіксованого числа ітерацій довжина кінцевого інтервалу пошуку мінімальна.

При розв'язуванні задачі максимізації функції $f(x)$ необхідно замінити її на функцію $-f(x)$.

Метод випадкового пошуку

Метод випадкового пошуку (МВП) застосовується для знаходження мінімуму (*максимуму*) *довільної функції* $y = f(\mathbf{x})$, *що задана в будь-якій допустимій області D*.

Розглянемо реалізацію даного методу для функції однієї змінної. Нехай довільна функція $f(x)$ задана на проміжку $[a, b]$. За допомогою давача випадкових чисел, рівномірно розподілених на проміжку $[0,1]$, будується послідовність випадкових чисел $x\{k\}$, $k = 1,\dots,N$, рівномірно розподілених на проміжку $[a, b]$. Обчислюються та порівнюються між собою значення функції $f(x)$ в точках $x\{k\}$. Мінімальне з них приймається за оцінку мінімуму функції $f(x)$ на проміжку $[a, b]$.

Якщо N прямує до нескінченності, отримана оцінка по ймовірності збігається до глобального мінімуму функції, що розглядається.

При розв'язуванні задачі максимізації функції $f(x)$ необхідно замінити її на функцію $-f(x)$.

Лекція 30. *Елементи опуклого аналізу. Опуклі функції, їх властивості. Субградієнт функкії, його властивості. – 2год. [1].*

Завдання для самостійної роботи. *Означення опуклої множини. Означення проекції точки на опуклу множину. Теорема про проекцію точки на опуклу множину. Теорема про відділяючу гіперплощину. Означення опорної гіперплощини для множини. Теорема про опорну гіперплощину. Теорема про гіперплощину, що розділяє дві опуклі множини. Означення опуклої (донизу) та увігнутої (опуклої доверху) функцій. Приклади опуклих функцій. Основні властивості опуклих функцій. Теорема про опуклість множини, що визначається опуклою нерівністю. Нерівність Ієнсена. Властивість неперервності. Існування похідної за напрямком. Критерій опуклості диференційовної функції. Субградієнт функції. Теорема про існування субградієнта. Теорема про властивості множини субградієнтів. Теорема про обчислення похідної за напрямком. Геометричний зміст субградієнта. Критерій глобального мінімуму опуклої функції. Критерій глобального мінімуму опуклої диференційовної функції. -4год. [1-3,7,9].*

Елементи опуклого аналізу

При дослідженні задач математичного програмування велике значення має така глобальна властивість функцій f і f_i, як опуклість (увігнутість) та тісно зв'язана з нею властивість опуклості множин.

Означення 6.4. *Множина $W \subset E^n$ називається опуклою, якщо для всіх $\mathbf{x}, \mathbf{y} \in W$ та $\lambda \in [0, 1]$ виконується умова $\lambda \mathbf{x} + (1 - \lambda)\, \mathbf{y} \in W$.*

Означення 6.5. *Проекцією точки $\mathbf{x}^0$ на опуклу множину W називають таку точку $\mathbf{x}^* \in W$, що*

$$\|x^0 - x^*\| = \inf_{x \in W} \|x^0 - x\| = d.$$

При цьому d називають відстанню точки $\mathbf{x}^0$ від множини W.

Теорема 6.1 (про проекцію точки на множину). *Для довільної опуклої замкненої множини W і довільної точки $\mathbf{x}^0$ існує єдина точка $\mathbf{x}^* \in W$, що є проекцією $\mathbf{x}^*$ на W.*

<u>Доведення</u>. Якщо $\mathbf{x}^0 \in W$, то, очевидно, що $\mathbf{x}^* = \mathbf{x}^0$ і $d = 0$.
Нехай $\mathbf{x}^0 \notin W$. Тоді $\exists\, c > 0$ таке, що $\forall\, \mathbf{x} \in W$

$$\|\mathbf{x}^0 - \mathbf{x}\| > c > 0.$$

Оскільки евклідова норма вектора є неперервною функцією $\mathbf{x}$ і внаслідок попередньої нерівності є обмеженою знизу на множині W, то існує її точна нижня грань

$$\inf_{\mathbf{x} \in W} \|\mathbf{x}^0 - \mathbf{x}\| = d.$$

Оскільки множина W замкнена, тобто включає в себе всі свої граничні точки і точки межі, то inf досягається принаймні в одній точці $\boldsymbol{x}^* \in W$

$$\inf_{x \in W} \left\| x^0 - x \right\| = \left\| x^0 - x^* \right\| = d.$$

Дійсно, оскільки $\inf \| \boldsymbol{x}^0 - \boldsymbol{x} \|$, $\boldsymbol{x} \in W$, існує, то завжди знайдеться послідовність $\{\boldsymbol{x}^k\}$, $\boldsymbol{x}^k \in W$, яка по нормі збігається до d

$$\lim_{k \to \infty} \left\| \boldsymbol{x}^0 - \boldsymbol{x}^k \right\| = d.$$

Далі, так як послідовність $\{\boldsymbol{x}^k\}$ обмежена, то з неї можна виділити послідовність $\{\boldsymbol{x}^{k_i}\}$, яка буде збігатись до деякої точки $\boldsymbol{x}^* \in W$

$$\lim_{i \to \infty} \boldsymbol{x}^{k_i} = \boldsymbol{x}^* \in W.$$

До того ж

$$\lim_{i \to \infty} \left\| \boldsymbol{x}^0 - \boldsymbol{x}^{k_i} \right\| = d,$$

що означає

$$\| \boldsymbol{x}^0 - \boldsymbol{x}^* \| = d.$$

Доведемо єдиність точки $\boldsymbol{x}^* \in W$. Від супротивного: припустимо, що існують дві точки $\boldsymbol{x}^1 \in W$ і $\boldsymbol{x}^2 \in W$, причому $\boldsymbol{x}^1 \neq \boldsymbol{x}^2$, для яких

$$\| \boldsymbol{x}^0 - \boldsymbol{x}^1 \| = \| \boldsymbol{x}^0 - \boldsymbol{x}^2 \| = d.$$

Множина W – опукла за умовою теореми, тому опукла комбінація точок $\boldsymbol{x}^1$ і $\boldsymbol{x}^2$ належить їй:

$$\boldsymbol{z} = (1/2)\,(\boldsymbol{x}^1 + \boldsymbol{x}^2) \in W.$$

Зауважимо, що точки $\boldsymbol{x}^0$, $\boldsymbol{x}^1$, $\boldsymbol{x}^2$, $\boldsymbol{z}$ лежать в одній площині.

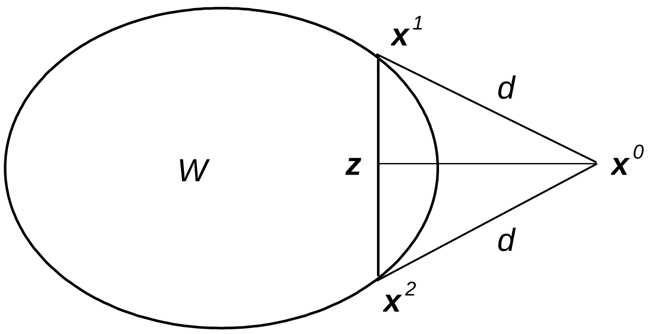


Рис. 6.8

Розглянемо $\| \boldsymbol{x}^0 - \boldsymbol{z} \|$. Оскільки

$$\| \boldsymbol{x}^0 - \boldsymbol{x}^1 \| = \| \boldsymbol{x}^0 - \boldsymbol{x}^2 \| = d,$$

то трикутник $\Delta \boldsymbol{x}^0 \boldsymbol{x}^1 \boldsymbol{x}^2$ рівнобедрений і вектори $\boldsymbol{x}^0 - \boldsymbol{z}$ та $\boldsymbol{x}^1 - \boldsymbol{z}$, а також $\boldsymbol{x}^0 - \boldsymbol{z}$ та $\boldsymbol{x}^2 - \boldsymbol{z}$ є ортогональними. Тоді за теоремою Піфагора маємо

$$\| \boldsymbol{x}^1 - \boldsymbol{z} \|^2 + \| \boldsymbol{x}^0 - \boldsymbol{z} \|^2 = \| \boldsymbol{x}^0 - \boldsymbol{x}^1 \|^2,$$

$$\|\boldsymbol{x}^2 - \boldsymbol{z}\|^2 + \|\boldsymbol{x}^0 - \boldsymbol{z}\|^2 = \|\boldsymbol{x}^0 - \boldsymbol{x}^2\|^2,$$

Так як $\boldsymbol{x}^1 \neq \boldsymbol{x}^2$, то $\boldsymbol{x}^1 \neq \boldsymbol{z}$ і $\boldsymbol{x}^2 \neq \boldsymbol{z}$. Тому

$$\|\boldsymbol{x}^0 - \boldsymbol{z}\|^2 < \|\boldsymbol{x}^0 - \boldsymbol{x}^1\|^2,$$

$$\|\boldsymbol{x}^0 - \boldsymbol{z}\|^2 < \|\boldsymbol{x}^0 - \boldsymbol{x}^2\|^2.$$

Додаючи останні нерівності, отримаємо

$$\|\boldsymbol{x}^0 - \boldsymbol{z}\|^2 < (1/2)\,(\|\boldsymbol{x}^0 - \boldsymbol{x}^1\|^2 + \|\boldsymbol{x}^0 - \boldsymbol{x}^2\|^2) = d^2,$$

що суперечить означенню inf. Теорема доведена.

Теорема 6.2 (про відділяючу гіперплощину). *Нехай W — замкнена опукла множина і $\boldsymbol{x}^0 \notin W$. Тоді існує гіперплощина*

$$(\boldsymbol{a}, \boldsymbol{x}) + b = 0$$

така, що

$$(\boldsymbol{a}, \boldsymbol{x}^0) + b > 0,$$

$$(\boldsymbol{a}, \boldsymbol{y}) + b < 0, \ \forall\, \boldsymbol{y} \in W.$$

Іншими словами, гіперплощина $(\boldsymbol{a}, \boldsymbol{x}) + b = 0$ строго відділяє точку $\boldsymbol{x}^0$ від множини W.

Доведення. Оскільки W замкнена опукла множина, то на основі теореми про проекцію існує точка $\boldsymbol{x}^* \in W$ така, що

$$\|x^0 - x^*\| = \inf_{y \in W} \|x^0 - y\| = d > 0.$$

Розглянемо гіперплощину

$$L(\boldsymbol{x}) = (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x} - \boldsymbol{x}^*) = 0,$$

яка проходить через точку $\boldsymbol{x}^*$ і має вектор нормалі $\boldsymbol{x}^0 - \boldsymbol{x}^*$ (див. рис. 6.9).

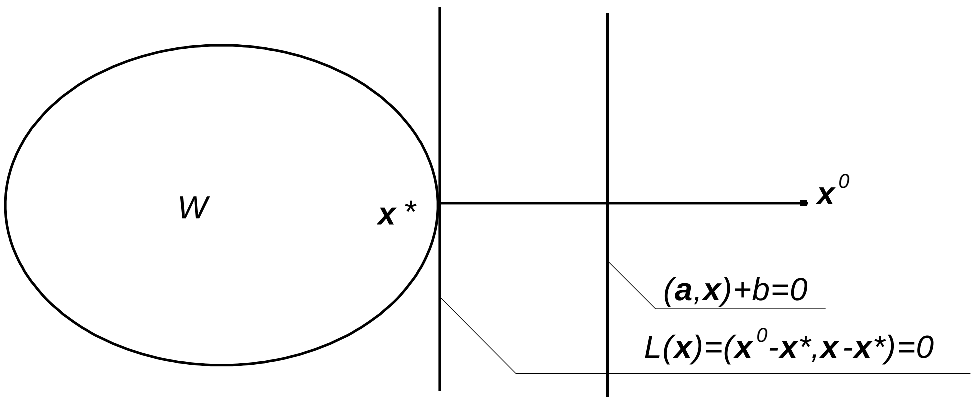


Рис. 6.9

Маємо

$$L(\boldsymbol{x}^0) = (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x}^0 - \boldsymbol{x}^*) = \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|^2 = d^2 > 0.$$

Покажемо, що $\forall\, \boldsymbol{y} \in W$ буде $L(\boldsymbol{y}) \leq 0$.

Від супротивного припустимо, що існує точка $\boldsymbol{y} \in W$, для якої $L(\boldsymbol{y}) > 0$. Сполучимо точку $\boldsymbol{y}$ з точкою $\boldsymbol{x}^*$ відрізком. Довільну точку відрізка позначимо

$$\boldsymbol{x}(\alpha) = \alpha \boldsymbol{y} + (1 - \alpha)\, \boldsymbol{x}^*,$$

де $\alpha \in [0, 1]$.

Розглянемо точки відрізка $\boldsymbol{x}(\alpha)$ відмінні від $\boldsymbol{x}^*$. Оскільки при $\alpha = 0$ $\boldsymbol{x}(\alpha) = \boldsymbol{x}^*$ і $L(\boldsymbol{x}^*) = 0$, то з подальшого розгляду виключимо випадок $\alpha = 0$.

Далі, W — опукла множина, тому $\boldsymbol{x}(\alpha) \in W$ при $\alpha \in (0,1]$.

Розглянемо $\|\boldsymbol{x}^0 - \boldsymbol{x}(\alpha)\|^2$. Маємо

$$\|\boldsymbol{x}^0 - \boldsymbol{x}(\alpha)\|^2 = \|\boldsymbol{x}^0 - \alpha \boldsymbol{y} - (1-\alpha)\boldsymbol{x}^*\|^2 = \|(\boldsymbol{x}^0 - \boldsymbol{x}^*) - \alpha(\boldsymbol{y} - \boldsymbol{x}^*)\|^2 =$$
$$= \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|^2 + \alpha^2\|\boldsymbol{y} - \boldsymbol{x}^*\|^2 - 2\alpha(\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{y} - \boldsymbol{x}^*).$$

Звідси $\forall \alpha \in (0,1]$

$$\|\boldsymbol{x}^0 - \boldsymbol{x}(\alpha)\|^2 - \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|^2 = \alpha^2\|\boldsymbol{y} - \boldsymbol{x}^*\|^2 - 2\alpha(\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{y} - \boldsymbol{x}^*).$$

Оскільки $\boldsymbol{x}^*$ — проекція $\boldsymbol{x}^0$ на W, то $\|\boldsymbol{x}^0 - \boldsymbol{x}(\alpha)\|^2 - \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|^2 > 0$ $\forall \alpha \in (0,1]$. Тому також $\alpha^2\|\boldsymbol{y} - \boldsymbol{x}^*\|^2 - 2\alpha(\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{y} - \boldsymbol{x}^*) > 0$ $\forall \alpha \in (0,1]$.

Отже $\forall \alpha \in (0,1]$ будемо мати $\alpha\|\boldsymbol{y} - \boldsymbol{x}^*\|^2 - 2L(\boldsymbol{y}) > 0$.

Але $\alpha\|\boldsymbol{y} - \boldsymbol{x}^*\|^2$ за рахунок вибору α може стати меншим за довільне наперед задане додатне число, зокрема, меншим ніж $2L(\boldsymbol{y}) > 0$.

Тоді $\|\boldsymbol{y} - \boldsymbol{x}^*\|^2 - 2L(\boldsymbol{y})$ буде від'ємним, а, значить, буде від'ємним і вираз $\|\boldsymbol{x}^0 - \boldsymbol{x}(\alpha)\|^2 - \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|^2$, що суперечить означенню проекції точки на множину.

Отже $\forall \boldsymbol{y} \in W$ маємо $L(\boldsymbol{y}) \le 0$.

Розглянемо тепер гіперплощину $L(\boldsymbol{x}) - c = 0$, де $0 < c < d^2$. Оскільки

$$L(\boldsymbol{x}^0) - c = d^2 - c > 0,$$

а $\forall \boldsymbol{y} \in W$ буде $L(\boldsymbol{y}) - c \le -c < 0$, то гіперплощина $L(\boldsymbol{x}) - c = 0$ відділяє точку $\boldsymbol{x}^0$ від множини W.

Остаточно маємо

$$L(\boldsymbol{x}) - c = (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x} - \boldsymbol{x}^*) - c = (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x}) - (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x}^*) - c = (\boldsymbol{a}, \boldsymbol{x}) + b,$$

де $\boldsymbol{a} = \boldsymbol{x}^0 - \boldsymbol{x}^*$, $b = -(\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x}^*) - c$.

Отже, гіперплощина $(\boldsymbol{a}, \boldsymbol{x}) + b = 0$ є шуканою.

Теорема доведена.

Означення 6.6. *Гіперплощина називається опорною, якщо вона проходить через точки межі площини W так, що всі точки множини W лежать по одну сторону від цієї гіперплощини.*

У доведеній теоремі гіперплощина $L(\boldsymbol{x}) = (\boldsymbol{x}^0 - \boldsymbol{x}^*, \boldsymbol{x} - \boldsymbol{x}^*) = 0$ є опорною, оскільки вона проходить через $\boldsymbol{x}^* \in W$ і $\forall \boldsymbol{y} \in W$ виконується умова $L(\boldsymbol{y}) \le 0$.

Теорема 6.3 (про опорну гіперплощину). *Через кожну точку межі опуклої замкненої множини W можна провести принаймні одну опорну гіперплощину.*

Доведення. Нехай $\bar{\boldsymbol{x}} \in W$ — точка межі множини W. Тоді в довільному ε-околі точки $\bar{\boldsymbol{x}}$ знайдуться точки $\boldsymbol{x}(k) \notin W$ ($k = 1, 2, \ldots$), для яких

$$\lim_{k \to \infty} \boldsymbol{x}(k) = \bar{\boldsymbol{x}}.$$

Кожній точці $\boldsymbol{x}(k)$ ($k = 1, 2, \ldots$) поставимо у відповідність гіперплощину

$$(\boldsymbol{x}(k) - \boldsymbol{x}^*(k), \boldsymbol{x} - \boldsymbol{x}^*(k)) = 0,$$

де $\boldsymbol{x}^*(k)$ — проекція точки $\boldsymbol{x}(k)$ на множину W.

Для кожної з таких гіперплощин, як було доведено в теоремі про відділяючу гіперплощину, виконується нерівність

$$(\boldsymbol{x}(k) - \boldsymbol{x}^*(k), \boldsymbol{y} - \boldsymbol{x}^*(k)) \le 0, \ \forall \boldsymbol{y} \in W.$$

Нормуючи вектори $\boldsymbol{x}(k) - \boldsymbol{x}^*(k)$, отримаємо для всіх k і $\boldsymbol{y} \in W$

$$\left(\frac{\boldsymbol{x}(k) - \boldsymbol{x}^*(k)}{|\boldsymbol{x}(k) - \boldsymbol{x}^*(k)|}, \boldsymbol{y} - \boldsymbol{x}^*(k) \right) \le 0 \ .$$

Покладемо

$$\boldsymbol{a}_k = \frac{\boldsymbol{x}(k) - \boldsymbol{x}^*(k)}{|\boldsymbol{x}(k) - \boldsymbol{x}^*(k)|}, \ k = 1, 2, \ldots,$$

і розглянемо послідовність $\{\boldsymbol{a}_k\}$. Оскільки ця послідовність обмежена (вона належить замкненій сфері радіуса 1), то з неї можна виділити збіжну підпослідовність $\{\boldsymbol{a}_{k_i}\}$.

Нехай

$$\lim_{i \to \infty} \boldsymbol{a}_{k_i} = \lim_{i \to \infty} \frac{\boldsymbol{x}(k_i) - \boldsymbol{x}^*(k_i)}{|\boldsymbol{x}(k_i) - \boldsymbol{x}^*(k_i)|} = \boldsymbol{a} \quad (|\boldsymbol{a}| = 1) \ .$$

Для всіх членів послідовності $\{\boldsymbol{a}_{k_i}\}$ маємо

$$(\boldsymbol{a}_{k_i}, \boldsymbol{y} - \boldsymbol{x}^*(k_i)) \le 0, \ \forall \boldsymbol{y} \in W.$$

Переходячи до границі при $i \to \infty$, отримаємо

$$(\boldsymbol{a}, \boldsymbol{y} - \bar{\boldsymbol{x}}) \le 0, \ \forall \boldsymbol{y} \in W,$$

оскільки з умови

$$\lim_{k \to \infty} \boldsymbol{x}(k) = \bar{\boldsymbol{x}}$$

випливає умова

$$\lim_{i \to \infty} \boldsymbol{x}^*(k_i) = \bar{\boldsymbol{x}}.$$

Отже, шуканою гіперплощиною є гіперплощина $L(\boldsymbol{x}) = (\boldsymbol{a}, \boldsymbol{x} - \bar{\boldsymbol{x}}) = 0$.

Теорема доведена.

<u>Наслідок</u>. *Якщо $W \subset E^n$ – замкнена, опукла, обмежена множина, то через довільну точку $\boldsymbol{x}^0 \in E^n$ (скінченну), $\boldsymbol{x}^0 \notin W$, можна провести опорну гіперплощину, тобто існує вектор $\boldsymbol{c} \ne \boldsymbol{0}$ такий, що $\forall \boldsymbol{y} \in W$ буде $(\boldsymbol{c}, \boldsymbol{y} - \boldsymbol{x}^0) \le 0$ або $(\boldsymbol{c}, \boldsymbol{y}) \le (\boldsymbol{c}, \boldsymbol{x}^0)$.*

<u>Доведення</u>. Розглянемо конус з вершиною в точці $\boldsymbol{x}^0$, породжений замиканням $\overline{W}$, тобто множину точок $\boldsymbol{K} = \{\alpha \boldsymbol{y} + (1-\alpha)\boldsymbol{x}^0 : \boldsymbol{y} \in \overline{W}, \ \alpha \ge 0\}$. Оскільки W опукла і замкнена, то конус $\boldsymbol{K}$ також опуклий і замкнений. Тоді на основі теореми про опорну гіперплощину через довільну скінченну точку його межі $\boldsymbol{x}^0 \in Гр\boldsymbol{K}$ можна провести опорну гіперплощину. Кожна така опорна гіперплощина буде проходити через вершину конуса $\boldsymbol{K}$ точку $\boldsymbol{x}^0$ і через деяку точку межі множини W, що і треба було довести.

Зауважимо, що обмеженість W є істотною, оскільки в іншому випадку точок межі конуса $\boldsymbol{K}$, через які можна було б провести опорну гіперплощину, що проходила б також і через точку $\boldsymbol{x}^0$, може взагалі не існувати.

Теорема 6.4 (про гіперплощину, яка розділяє дві опуклі множини, що не перетинаються). *Якщо множина X^0 внутрішніх точок опуклої множини X не порожня і не перетинається з опуклою множиною Y ($X^0 \cap Y = \varnothing$), то для множин X та Y існує розділяюча їх гіперплощина, тобто існує вектор* $\boldsymbol{c} \neq \boldsymbol{0}$ *такий, що*

$$(\boldsymbol{c}, \boldsymbol{x}) \le (\boldsymbol{c}, \boldsymbol{y}), \quad \forall\, \boldsymbol{x} \in X,\ \forall\, \boldsymbol{y} \in Y.$$

<u>Доведення</u>. Розглянемо множину $Z = X^0 - Y$:

$$Z = \{\boldsymbol{z} : \boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y},\ \boldsymbol{x} \in X^0,\ \boldsymbol{y} \in Y\}.$$

Вона є опуклою (перевіряється безпосередньо) і, оскільки $X^0 \cap Y = \varnothing$, то $\boldsymbol{0}$ не є її внутрішньою точкою.

Якщо $\boldsymbol{0} \notin \overline{Z}$, то в силу наслідку теореми про опорну гіперплощину, або, якщо $\boldsymbol{z} = \boldsymbol{0}$ є точкою межі Z, то в силу теореми про опорну гіперплощину, існує ненульовий вектор $\boldsymbol{c} \neq \boldsymbol{0}$ такий, що

$$(\boldsymbol{c}, \boldsymbol{z} - \boldsymbol{0}) \le 0, \quad \forall\, \boldsymbol{z} \in Z.$$

Тобто $\forall\, \boldsymbol{x} \in X^0$, і $\forall\, \boldsymbol{y} \in Y$ будемо мати

$$(\boldsymbol{c}, \boldsymbol{x} - \boldsymbol{y}) \le 0 \text{ або } (\boldsymbol{c}, \boldsymbol{x}) \le (\boldsymbol{c}, \boldsymbol{y}).$$

Ця нерівність залишається справедливою і для замикань множин X і Y, оскільки граничний перехід не порушує нестрогих нерівностей.

Теорема доведена.

Опуклі функції та їх основні властивості

Розглянемо тепер *означення, приклади та властивості опуклих функцій.*

Означення 6.7. *Функцію* $f(\boldsymbol{x})$, $\boldsymbol{x} \in X \subset E^n$, *де* X – *опукла множина, називають опуклою (опуклою донизу), якщо* $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ і $\forall\ \alpha \in [0,1]$ *виконується нерівність*

$$f(\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \le \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2)$$

(дивись рис. 6.10).

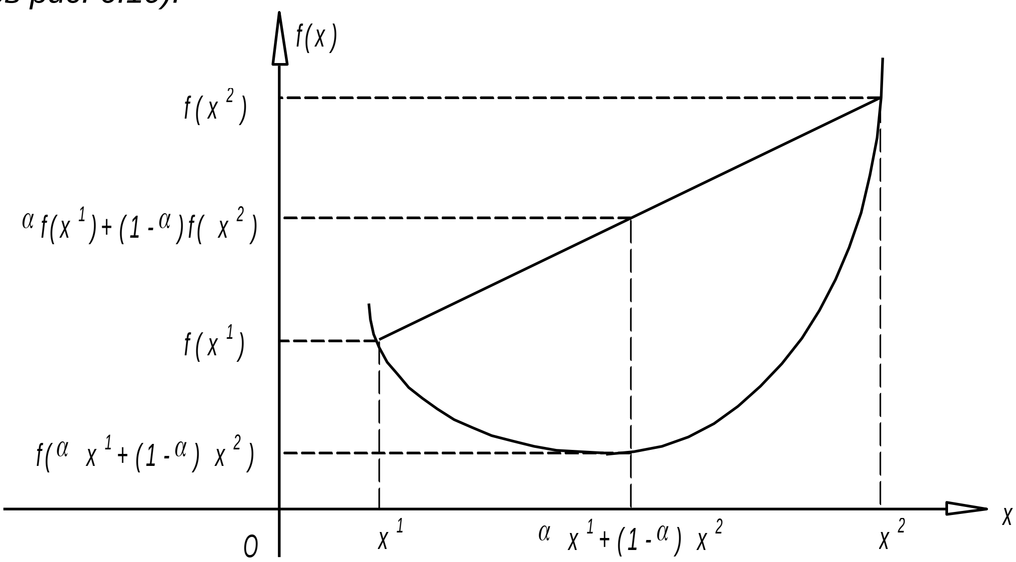


Рис. 6.10

Означення 6.8. *Функцію* $f(\boldsymbol{x})$, *визначену на опуклій множині* $X \subset E^n$ *називають увігнутою (опуклою доверху), якщо* $-f(\boldsymbol{x})$ *є опуклою функцією на* X.

Зауважимо, що опуклість множини X є істотною в цих означеннях, оскільки $\forall$ $\boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ і $\forall\, \alpha \in [0,1]$ точка $\alpha \boldsymbol{x}^1 + (1-\alpha)\, \boldsymbol{x}^2$ повинна належати множині X.

Якщо в означеннях опуклої і увігнутої функцій знаки нестрогої нерівності замінити на знаки строгої нерівності, то будемо мати означення *строго опуклої і строго увігнутої функцій.*

Приклади опуклих функцій

1. *Лінійна функція* $f(\boldsymbol{x}) = \boldsymbol{c}^T \boldsymbol{x} = (\boldsymbol{c}, \boldsymbol{x})$ *є опуклою (і увігнутою) у всьому просторі* E^n.

Перевіряється безпосередньо з використанням означень відповідних функцій.

2. *Квадратична форма* $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x})$ *опукла у всьому просторі* E^n, *якщо вона невід'ємно визначена (тобто, якщо* $\forall\ \boldsymbol{x} \in E^n$ *має місце нерівність* $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \ge 0$).

Дійсно, утворимо опуклу комбінацію двох довільних точок $\boldsymbol{x}^1$, $\boldsymbol{x}^2 \in E^n$ і розглянемо відповідне їй значення квадратичної форми.

Маємо $\forall\, \alpha \in [0,1]$ і $\boldsymbol{x}^1, \boldsymbol{x}^2 \in E^n$

$$f(\alpha \boldsymbol{x}^1 + (1-\alpha)\, \boldsymbol{x}^2) = (\alpha \boldsymbol{x}^1 + (1-\alpha)\, \boldsymbol{x}^2)^T \boldsymbol{A}\, (\alpha \boldsymbol{x}^1 + (1-\alpha)\, \boldsymbol{x}^2) =$$

$$= (\boldsymbol{x}^2 + \alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2))^T \boldsymbol{A}\, (\boldsymbol{x}^2 + \alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)) =$$

$$= (\boldsymbol{x}^2)^T \boldsymbol{A}\, \boldsymbol{x}^2 + 2\alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}\, (\boldsymbol{x}^2) + \alpha^2\, (\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}\, (\boldsymbol{x}^1 - \boldsymbol{x}^2).$$

Так як $0 \le \alpha \le 1$, то $\alpha^2 \le \alpha$, і, оскільки $\forall\ \boldsymbol{x} \in E^n$ виконується $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \ge 0$, то

$$\alpha^2\, (\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}\, (\boldsymbol{x}^1 - \boldsymbol{x}^2) \le \alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}\, (\boldsymbol{x}^1 - \boldsymbol{x}^2).$$

Використовуючи останню нерівність, отримаємо:

$$f(\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \le (\boldsymbol{x}^2)^T \boldsymbol{A}\boldsymbol{x}^2 + 2\alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}(\boldsymbol{x}^2) + \alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}(\boldsymbol{x}^1 - \boldsymbol{x}^2) =$$

$$= (\boldsymbol{x}^2)^T \boldsymbol{A}\, \boldsymbol{x}^2 + \alpha(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \boldsymbol{A}\, (\boldsymbol{x}^1 - \boldsymbol{x}^2) = (\boldsymbol{x}^2)^T \boldsymbol{A}\, \boldsymbol{x}^2 + \alpha(\boldsymbol{x}^1)^T \boldsymbol{A}\, \boldsymbol{x}^1 -$$

$$- \alpha(\boldsymbol{x}^2)^T \boldsymbol{A}\, \boldsymbol{x}^2 = \alpha(\boldsymbol{x}^1)^T \boldsymbol{A}\boldsymbol{x}^1 + (1-\alpha)(\boldsymbol{x}^2)^T \boldsymbol{A}\boldsymbol{x}^2 = \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2),$$

що і треба було довести.

3. *Сума опуклих функцій*

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} f_i(\boldsymbol{x}),$$

визначених на опуклій множині $X \subset E^n$, *є опуклою функцією на* X.

Дійсно, $\forall\, \alpha \in [0,1]$ і $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ маємо

$$f(\alpha\ \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) = \sum_{i=1}^{k} f_i(\alpha\ \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \le$$

$$\le \sum_{i=1}^{k} (\alpha\ f_i(\boldsymbol{x}^1) + (1-\alpha)\ f_i(\boldsymbol{x}^2)) = \alpha \sum_{i=1}^{k} f_i(\boldsymbol{x}^1) + (1-\alpha) \sum_{i=1}^{k} f_i(\boldsymbol{x}^2) =$$

$$= \alpha f(\boldsymbol{x}^1) + (1-\alpha)\, f(\boldsymbol{x}^2).$$

4. *Квадратична функція* $f(\boldsymbol{x})=\boldsymbol{c}^T\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=(\boldsymbol{c},\boldsymbol{x})+(\boldsymbol{x},\boldsymbol{A}\boldsymbol{x})$ *є опуклою функцією у всьому просторі* E^n, *якщо квадратична форма* $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ *невід'ємно визначена.*

Це твердження є наслідком трьох попередніх тверджень.

5. *Якщо* $f(\boldsymbol{x})$ *опукла на опуклій множині* X, *то функція*

$$g(\boldsymbol{x}) = \max\{f(\boldsymbol{x}),0\}$$

також опукла на X.

Дійсно, $\forall\ \alpha\in[0,1]$ і $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ маємо

$$g(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2) = \max\{f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2),0\} \le$$
$$\le \max\{\alpha f(\boldsymbol{x}^1)+(1-\alpha)f(\boldsymbol{x}^2),0\} \le$$
$$\le \alpha\max\{f(\boldsymbol{x}^1),0\} + (1-\alpha)\max\{f(\boldsymbol{x}^2),0\} = \alpha g(\boldsymbol{x}^1) + (1-\alpha)g(\boldsymbol{x}^2).$$

6. *Якщо* $g(\boldsymbol{x})$ *опукла і невід'ємна на опуклій множині* X, *то функція* $f(\boldsymbol{x})=g^2(\boldsymbol{x})$ *також опукла на* X.

Дійсно, $\forall\ \alpha\in[0,1]$ і $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ маємо

$$f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2) = g^2(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)\le(\alpha g(\boldsymbol{x}^1)+(1-\alpha)g(\boldsymbol{x}^2))^2=$$
$$= \alpha^2g^2(\boldsymbol{x}^1)+2\alpha(1-\alpha)g(\boldsymbol{x}^1)g(\boldsymbol{x}^2)+(1-\alpha)^2g^2(\boldsymbol{x}^2) =$$
$$= \alpha^2g^2(\boldsymbol{x}^1)+ (1-\alpha)^2g^2(\boldsymbol{x}^2)+\alpha(1-\alpha)[g^2(\boldsymbol{x}^1)+g^2(\boldsymbol{x}^2)-$$
$$-(g^2(\boldsymbol{x}^1)-g^2(\boldsymbol{x}^2))^2] = \alpha g^2(\boldsymbol{x}^1)+(1-\alpha)g^2(\boldsymbol{x}^2)-$$
$$-\alpha(1-\alpha)[g(\boldsymbol{x}^1)-g(\boldsymbol{x}^2)]^2 \le \alpha g^2(\boldsymbol{x}^1)+(1-\alpha)g^2(\boldsymbol{x}^2) =$$
$$= \alpha g(\boldsymbol{x}^1)+(1-\alpha)g(\boldsymbol{x}^2).$$

Основні властивості опуклих функцій

1. *Для довільної опуклої функції* $g(\boldsymbol{x})$, *визначеної на опуклій множині* $X\subset E^n$, *множина* $G=\{\boldsymbol{x}\in X: g(\boldsymbol{x})\le 0\}$ *є опуклою множиною.*

Доведення. Розглянемо опуклу комбінацію двох довільних точок із G і покажемо, що вона також належить G.

Маємо $\forall\ \alpha\in[0,1]$ і $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in G$

$$g(\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \le \alpha g(\boldsymbol{x}^1) + (1-\alpha)g(\boldsymbol{x}^2) \le \alpha\cdot 0+(1-\alpha)\cdot 0 = 0,$$

тобто $\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2 \in G$, що і треба було довести.

2. Нерівність Ієнсена. *Якщо* $f(\boldsymbol{x})$ *опукла на опуклій множині* $X\subset E^n$ *та* $\boldsymbol{x}^i\in X$ $(i=1,\ldots,m)$, $\alpha_i\ge 0$ $(i=1,\ldots,m)$, $\alpha_1+\ldots+\alpha_m=1$, *то*

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

Зауважимо, що при $m=2$ нерівність Ієнсена співпадає з нерівністю, яка визначає опуклу функцію. При $m>2$ нерівність доводиться за індукцією.

Без доведення (доведення див. [15], стор. 42).

3. Властивість неперервності. *Якщо функція* $f(\boldsymbol{x})$ *опукла на опуклій множині* $X\subset E^n$, *то вона неперервна у всіх внутрішніх точках множини* X.

Без доведення (доведення див. [15], стор. 43).

4. Існування похідної по напрямку. *Якщо функція* $f(\boldsymbol{x})$ *опукла на опуклій множині* $X\subset E^n$, *то в кожній внутрішній точці* $\boldsymbol{x}\in X$ *вона має похідну по довільному напрямку* $\boldsymbol{r}$ $(\|\boldsymbol{r}\|=1)$

$$D_r f(\boldsymbol{x}) = \lim_{h\to +0}\frac{f(\boldsymbol{x}+h\boldsymbol{r})-f(\boldsymbol{x})}{h}.$$

Без доведення (доведення див. [15], стор. 39).

5. Основна властивість диференційовних опуклих функцій. *Якщо диференційовна на опуклій множині* $X\subset E^n$ *функція* $f(\boldsymbol{x})$ $(f(\boldsymbol{x})\in C_X^1)$ *опукла на* X, *то в кожній внутрішній точці* $\boldsymbol{x}^1$ *цієї множини* ($\boldsymbol{x}^1\in \operatorname{int} X$) *виконується нерівність*

$$\nabla^T f(\boldsymbol{x}^1)(\boldsymbol{x}^2-\boldsymbol{x}^1)\le f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)\ \ \forall\, \boldsymbol{x}^2\in X. \qquad (6.8)$$

<u>Доведення</u>. Оскільки $f(\boldsymbol{x})$ опукла на X, то $\forall\ \boldsymbol{x}^1\in \operatorname{int} X$, $\forall\ \boldsymbol{x}^2\in X$ і $\forall\alpha\in(0,1)$ має місце умова

$$f(\alpha\boldsymbol{x}^2+(1-\alpha)\boldsymbol{x}^1)\le \alpha f(\boldsymbol{x}^2)+(1-\alpha)f(\boldsymbol{x}^1).$$

Перепишемо її у вигляді

$$\frac{f(\boldsymbol{x}^1+\alpha(\boldsymbol{x}^2-\boldsymbol{x}^1))-f(\boldsymbol{x}^1)}{\alpha}\le f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1).$$

Оскільки $f(\boldsymbol{x})\in C_X^1$, то з першої формули Тейлора маємо

$$f(\boldsymbol{x}^1+\alpha(\boldsymbol{x}^2-\boldsymbol{x}^1))-f(\boldsymbol{x}^1)=\alpha\nabla^T f(\boldsymbol{x}^1+\theta\alpha(\boldsymbol{x}^2-\boldsymbol{x}^1))(\boldsymbol{x}^2-\boldsymbol{x}^1),$$

де $\theta\in(0,1)$.

Із останніх двох співвідношень отримаємо

$$\nabla^T f(\boldsymbol{x}^1+\theta\alpha(\boldsymbol{x}^2-\boldsymbol{x}^1))(\boldsymbol{x}^2-\boldsymbol{x}^1)\le f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1).$$

Переходячи до границі при $\alpha\to 0$ в останній нерівності, приходимо до нерівності

$$\nabla^T f(\boldsymbol{x}^1)(\boldsymbol{x}^2-\boldsymbol{x}^1)\le f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1),$$

яку і треба було довести.

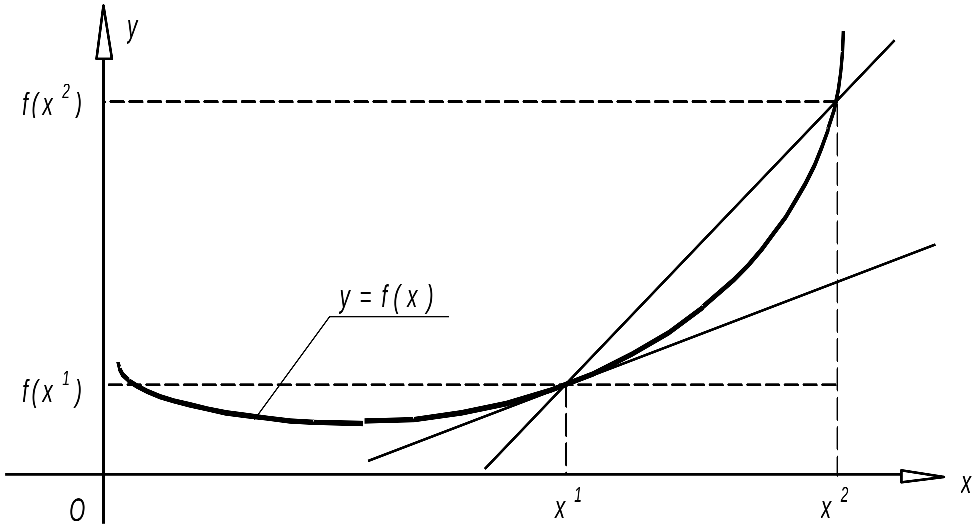


Рис. 6.11

Геометрична інтерпретація нерівності (6.8) для диференційовної опуклої функції $y=f(x)$ однієї змінної ($x\in E^1$) дається на рис. 6.11. При цьому співвідношення (6.8) набуває вигляду

$$f'(x^1)\le\frac{f(x^2)-f(x^1)}{x^2-x^1}\quad (x^2>x^1)$$

і ліва його частина визначає в точці x^1 кутовий коефіцієнт дотичної, а права – кутовий коефіцієнт січної для кривої $y=f(x)$.

6. Вірна і обернена теорема до щойно доведеної. *Якщо для диференційовної на опуклій множині X функції* $f(\boldsymbol{x})$ $(f(\boldsymbol{x})\in C_X^1)$, *в кожній внутрішній точці* $\boldsymbol{x}^1$ *множини X* ($\boldsymbol{x}^1\in \operatorname{int} X$) *має місце нерівність*

$$\nabla^T f(\boldsymbol{x}^1)\,(\boldsymbol{x}^2-\boldsymbol{x}^1)\le f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)$$

$\forall\, \boldsymbol{x}^2\in X$, *то* $f(\boldsymbol{x})$ *опукла на* X.

<u>Доведення</u>. $\forall\ \boldsymbol{x}^1,\boldsymbol{x}^2\in X$ і $\forall\alpha\in[0,1]$ покладемо $\boldsymbol{z}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2$. Тоді $\boldsymbol{z}\in X$, оскільки X – опукла множина. За умовою теореми маємо

$$\nabla^T f(\boldsymbol{z})\,(\boldsymbol{x}^1-\boldsymbol{z})\le f(\boldsymbol{x}^1)-f(\boldsymbol{z}),$$

$$\nabla^T f(\boldsymbol{z})\,(\boldsymbol{x}^2-\boldsymbol{z})\le f(\boldsymbol{x}^2)-f(\boldsymbol{z}).$$

Помножимо першу нерівність на α, другу – на $1-\alpha$ і додамо їх. Отримаємо

$$\nabla^T f(\boldsymbol{z})\,(\alpha\boldsymbol{x}^1+(1-\alpha)\,\boldsymbol{x}^2-\boldsymbol{z})\le\alpha f(\boldsymbol{x}^1)+(1-\alpha)\,f(\boldsymbol{x}^2)-f(\boldsymbol{z}).$$

Оскільки ліва частина останньої нерівності дорівнює $\nabla^T f(\boldsymbol{z})\boldsymbol{0}=0$, то остаточно отримаємо

$$f(\alpha\boldsymbol{x}^2+(1-\alpha)\,\boldsymbol{x}^1)\le\alpha f(\boldsymbol{x}^2)+(1-\alpha)\,f(\boldsymbol{x}^1)$$

$\forall\, \boldsymbol{x}^1,\boldsymbol{x}^2\in X$ і $\forall\alpha\in[0,1]$ (з урахуванням рівності $\boldsymbol{z}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2$).

Властивість доведена.

Співвідношення (6.8) може служити відправним пунктом для визначення важливого поняття *субградієнта* (*узагальненого градієнта*) *функції.*

Субградієнт функції та його основні властивості

Означення 6.9. *Нехай функція* $f(\boldsymbol{x})$ *визначена на множині* $X \subset E^n$, *для якої* $\operatorname{int} X \neq \varnothing$. *Якщо в деякій внутрішній точці* $\boldsymbol{x}^1$ *множини* X ($\boldsymbol{x}^1 \in \operatorname{int} X$) $\exists$ *вектор* $\hat{\nabla} f(\boldsymbol{x}^1)$ *такий, що виконується умова*

$$\hat{\nabla}^T f(\boldsymbol{x}^1)(\boldsymbol{x}^2 - \boldsymbol{x}^1) \leq f(\boldsymbol{x}^2) - f(\boldsymbol{x}^1)$$

$\forall \boldsymbol{x}^2 \in X$, *то вектор* $\hat{\nabla} f(\boldsymbol{x}^1)$ *називають субградієнтом функції* $f(\boldsymbol{x})$ *в точці* $\boldsymbol{x}^1$.

Введемо поняття *надграфіка* функції багатьох змінних.

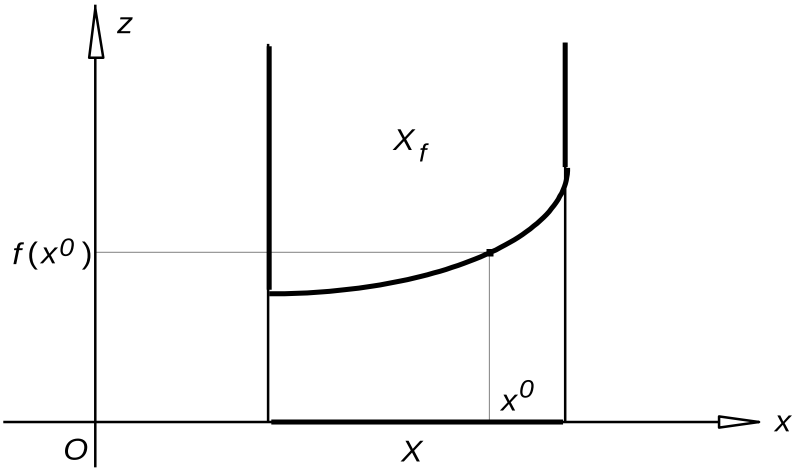


Рис. 6.12

Означення 6.10. *Нехай функція* $z = f(\boldsymbol{x}) = f(x_1,\ldots,x_n)$ *визначена на множині* $X \subset E^n$. *Назвемо надграфіком цієї функції множину* X_f *точок простору* E^{n+1} *таку, що*

$$X_f = \{(z, x_1,\ldots,x_n):\ z \geq f(\boldsymbol{x}),\ \boldsymbol{x} \in X,\ z \in E^1\}.$$

Зауважимо, що за властивістю **1** опуклих функцій надграфік опуклої функції є опуклою множиною, оскільки X_f можна подати у вигляді

$$X_f = \{(z, x_1,\ldots,x_n):\ g(z, x) = -z + f(\boldsymbol{x}) \leq 0,\ (z, \boldsymbol{x}) \in E^1 \times X\}.$$

Теорема 6.5 (про існування субградієнта). *Якщо функція* $f(\boldsymbol{x})$ *опукла на опуклій множині* $X \subset E^n$, *то у кожній внутрішній точці* $\boldsymbol{x}^0 \in X$ *існує субградієнт* $\hat{\nabla} f(\boldsymbol{x}^0)$.

<u>Доведення</u>. Нехай $\boldsymbol{x}^0 \in \operatorname{int} X$ і $(f(\boldsymbol{x}^0), \boldsymbol{x}^0) = \boldsymbol{u}^0 \in E^{n+1}$. Точка $\boldsymbol{u}^0$ є точкою межі надграфіка X_f, оскільки лежить на гіперповерхні $z = f(\boldsymbol{x})$.

Так як для опуклої функції її надграфік є опуклою множиною, то на основі теореми про опорну гіперплощину існує опорна гіперплощина, яка проходить через точку $(f(\boldsymbol{x}^0), \boldsymbol{x}^0)$. Нехай рівняння цієї гіперплощини має вигляд

$$(\boldsymbol{a}, \boldsymbol{u}\quad \boldsymbol{u}^0) = 0,$$

причому для всіх $\boldsymbol{u} \in X_f \subset E^{n+1}$ буде

$(\boldsymbol{a}, \boldsymbol{u} - \boldsymbol{u}^0) \le 0$.

Оскільки $\boldsymbol{a}$ — $(n+1)$-вимірний вектор, подамо його у вигляді $\boldsymbol{a} = (-\lambda, \boldsymbol{r})$, де $\boldsymbol{r}$ — n-вимірний вектор. Тоді рівняння опорної гіперплощини можна переписати у вигляді

$$-\lambda(z - f(\boldsymbol{x}^0)) + (\boldsymbol{r}, \boldsymbol{x} - \boldsymbol{x}^0) = 0,$$

причому $\forall\, (z, \boldsymbol{x}) \in X_f$ повинна мати місце умова

$$-\lambda(z - f(\boldsymbol{x}^0)) + (\boldsymbol{r}, \boldsymbol{x} - \boldsymbol{x}^0) \le 0,$$

де $\lambda > 0$, оскільки нерівність зберігається і при $z \to +\infty$.

Покладемо $\hat{\nabla} = \boldsymbol{r}/\lambda$. Отримаємо $\forall\, (z, \boldsymbol{x}) \in X_f$

$$(\hat{\nabla}, \boldsymbol{x} - \boldsymbol{x}^0) \le z - f(\boldsymbol{x}^0).$$

Зокрема, можна покласти $z = f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in X$ в цій нерівності. Тоді, позначивши $\hat{\nabla} = \hat{\nabla} f(\boldsymbol{x}^0)$, остаточно будемо мати $\forall \boldsymbol{x} \in X$

$$(\hat{\nabla} f(\boldsymbol{x}^0), \boldsymbol{x} - \boldsymbol{x}^0) \le f(\boldsymbol{x}) - f(\boldsymbol{x}^0),$$

тобто: за означенням субградієнта, вектор $\hat{\nabla} f(\boldsymbol{x}^0)$ є субградієнтом функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^0 \in \operatorname{int} X$.

Зауважимо, що, оскільки вектор нормалі $\boldsymbol{a}$ опорної гіперплощини

$$(\boldsymbol{a}, \boldsymbol{u} - \boldsymbol{u}^0) = 0$$

може бути пронормований завжди, а число $\lambda > 0$, то множина векторів вигляду $\hat{\nabla} = \boldsymbol{r}/\lambda$ буде обмеженою завжди. Це означає, що множина субградієнтів $\partial f(\boldsymbol{x}^0)$ опуклої функції $f(\boldsymbol{x})$ $\forall\, \boldsymbol{x}^0 \in \operatorname{int} X$ буде обмеженою завжди.

Теорема доведена.

Теорема 6.6 (про властивості множини субградієнтів). *Якщо функція* $f(\boldsymbol{x})$ опукла *на опуклій множині* X, *то множина її субградієнтів* $\partial f(\boldsymbol{x}^0)$ $\forall \boldsymbol{x}^0 \in \operatorname{int} X$ *є непорожньою, обмеженою, опуклою і замкненою.*

<u>Доведення</u>. Існування принаймні одного субградієнта $\forall\, \boldsymbol{x}^0 \in \operatorname{int} X$, а також обмеженість їх множини доведені в попередній теоремі.

Доведемо опуклість. Нехай $\hat{\nabla}_1, \hat{\nabla}_2 \in \partial f(\boldsymbol{x}^0)$. Це означає, що $\forall \boldsymbol{x} \in X$ виконуються умови

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^0) \ge (\hat{\nabla}_1, \boldsymbol{x} - \boldsymbol{x}^0),$$

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^0) \ge (\hat{\nabla}_2, \boldsymbol{x} - \boldsymbol{x}^0).$$

Візьмемо довільне $\alpha \in [0,1]$. Помножимо першу нерівність на α, другу — на $1 - \alpha$ і додамо їх. Отримаємо $\forall\, \boldsymbol{x} \in X$

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^0) \ge (\alpha \hat{\nabla}_1 + (1 - \alpha)\hat{\nabla}_2, \boldsymbol{x} - \boldsymbol{x}^0),$$

що означає належність опуклої комбінації векторів $\hat{\nabla}_1$, $\hat{\nabla}_2$ до множини $\partial f(\boldsymbol{x}^0)$. Отже, множина $\partial f(\boldsymbol{x}^0)$ опукла.

Доведемо замкненість. Нехай послідовність $\{\hat{\nabla}_s\} \in \partial f(\boldsymbol{x}^0)$ і $\lim \hat{\nabla}_s = \hat{\nabla}$ при $s \to \infty$. Тоді для довільного s та $\forall\, \boldsymbol{x} \in X$ має місце умова

$$f(\boldsymbol{x})\quad f(\boldsymbol{x}^0) \geq (\hat{\nabla}_s, \boldsymbol{x}\quad \boldsymbol{x}^0).$$

Переходячи в ній до границі при $s \to \infty$, отримаємо $\forall\, \boldsymbol{x} \in X$

$$f(\boldsymbol{x})\quad f(\boldsymbol{x}^0) \geq (\hat{\nabla}, \boldsymbol{x}\quad \boldsymbol{x}^0),$$

тобто $\hat{\nabla} \in \partial f(\boldsymbol{x}^0)$. Отже, $\partial f(\boldsymbol{x}^0)$ □ замкнена множина.

Теорема доведена.

Теорема 6.7 (про обчислення похідної по напрямку). *Якщо функція* $f(\boldsymbol{x})$ *опукла на опуклій множині* $X \subset E^n$, *то похідна по довільному напрямку* $\boldsymbol{r}$ $(|\boldsymbol{r}| = 1)$ *цієї функції в довільній внутрішній точці* $\boldsymbol{x}^0 \in X$ *обчислюється за формулою*

$$D_r f(\boldsymbol{x}^0) = \max_{\hat{\nabla} \in \partial f(\boldsymbol{x}^0)} (\hat{\nabla}, \boldsymbol{r}).$$

Без доведення (доведення див. [16], стор. 288).

Зауваження відносно геометричної інтерпретації субградієнта. Як було з'ясовано раніше, градієнт $\nabla f(\boldsymbol{x}^0)$ опуклої диференційовної функції $f(\boldsymbol{x})$ є вектором нормалі дотичної гіперплощини до гіперповерхні рівня в точці $\boldsymbol{x}^0$, спрямованим у напрямку зростання функції $f(\boldsymbol{x})$. Дотична гіперплощина є і опорною у цьому випадку для множини $\{\boldsymbol{x} \in E^n \colon f(\boldsymbol{x}) \leq f(\boldsymbol{x}^0)\}$, до того ж єдиною.

Якщо опукла функція $f(\boldsymbol{x})$ не є диференційовною в точці $\boldsymbol{x}^0$, то дотичної гіперплощини до гіперповерхні рівня не існує. Але буде існувати безліч опорних гіперплощин для множини $\{\boldsymbol{x} \in E^n \colon f(\boldsymbol{x}) \leq f(\boldsymbol{x}^0)\}$, що проходитимуть через точку $\boldsymbol{x}^0$. *Вектор нормалі довільної із таких опорних гіперплощин, спрямований в бік зростання функції* $f(\boldsymbol{x})$ і буде задавати напрямок *її субградієнта* $\hat{\nabla} f(\boldsymbol{x}^0)$ *в точці* $\boldsymbol{x}^0$ (див. рис. 6.13).

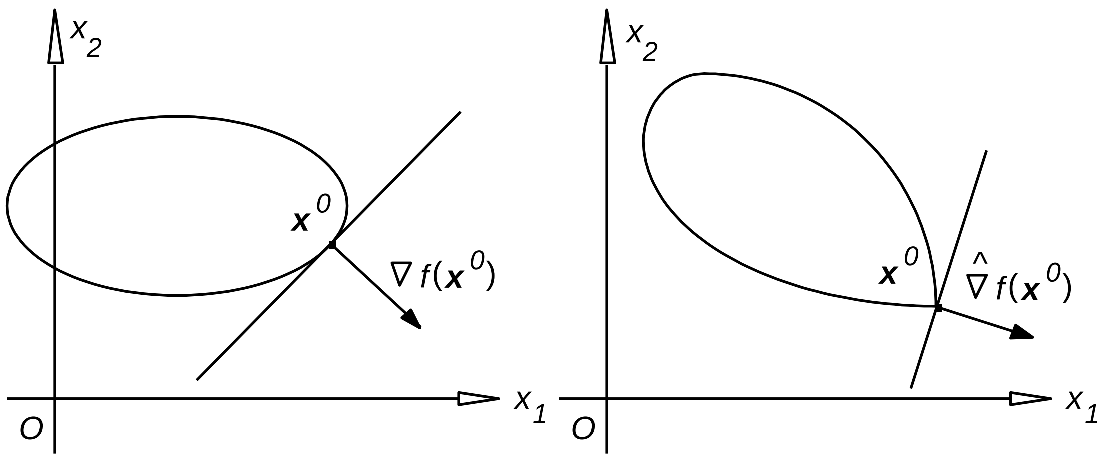


Рис. 6.13

Очевидно, що субградієнт рівний градієнту функції $f(\boldsymbol{x})$ в точках, де вона диференційовна.

Екстремальні властивості опуклих функцій

Теорема 6.8 (про рівність абсолютного і відносного мінімумів). *Якщо функція $f(\boldsymbol{x})$ опукла на опуклій замкненій множині $X \subset E^n$, то довільний відносний мінімум функції $f(\boldsymbol{x})$, що досягається в деякій точці множини X, є її абсолютним мінімумом на множині X.*

<u>Доведення.</u> Нехай $\boldsymbol{x}^0 \in X$ точка відносного мінімуму функції $f(\boldsymbol{x})$. Припустимо від супротивного, що $\boldsymbol{x}^0$ не буде точкою абсолютного мінімуму функції $f(\boldsymbol{x})$ на множині X. Це означає, що існує точка $\boldsymbol{x}^* \in X$ така, що

$$f(\boldsymbol{x}^*) < f(\boldsymbol{x}^0).$$

Оскільки множина X опукла, то опукла комбінація точок $\boldsymbol{x}^*$ і $\boldsymbol{x}^0$ належить X, тобто

$$\forall\ \alpha \in [0,1],\ \alpha \boldsymbol{x}^* + (1-\alpha)\,\boldsymbol{x}^0 \in X.$$

Так як функція $f(\boldsymbol{x})$ опукла і $f(\boldsymbol{x}^*) < f(\boldsymbol{x}^0)$ за припущенням, то

$$f(\alpha\boldsymbol{x}^* + (1-\alpha)\boldsymbol{x}^0) \le \alpha f(\boldsymbol{x}^*) + (1-\alpha) f(\boldsymbol{x}^0) <$$
$$< \alpha f(\boldsymbol{x}^0) + (1-\alpha) f(\boldsymbol{x}^0) = f(\boldsymbol{x}^0),$$

тобто $\forall\ \alpha \in [0,1]$ виконується умова

$$f(\alpha\boldsymbol{x}^* + (1-\alpha)\boldsymbol{x}^0) < f(\boldsymbol{x}^0).$$

Але при α близьких до нуля точка $\alpha\boldsymbol{x}^* + (1-\alpha)\boldsymbol{x}^0$ попадає в досить малий окіл точки $\boldsymbol{x}^0$, а отже, остання умова суперечить означенню відносного мінімуму.

Отже, припущення було невірним і $\boldsymbol{x}^0$ є точкою абсолютного мінімуму функції $f(\boldsymbol{x})$ на множині X.

Теорема 6.9 (необхідні і достатні умови мінімуму опуклої функції). *Якщо функція $f(\boldsymbol{x})$ опукла на опуклій множині $X \subset E^n$ і $\boldsymbol{x}^* \in \operatorname{int} X$, то $\boldsymbol{x}^*$ є точкою абсолютного мінімуму функції $f(\boldsymbol{x})$ на множині X тоді і тільки тоді, коли нуль-вектор належить множині субградієнтів функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^*$ $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*)$.*

<u>Доведення</u>. Необхідність. Нехай $\boldsymbol{x}^*$ є точкою абсолютного мінімуму функції $f(\boldsymbol{x})$ на множині X. Це означає, що $\forall\ \boldsymbol{x} \in X$

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \ge 0$$

або

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \ge (\boldsymbol{0}, (\boldsymbol{x} - \boldsymbol{x}^*)).$$

Тоді за означенням субградієнта функції в точці, нуль є субградієнтом функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^*$, тобто $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*)$.

Достатність. Нехай $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*)$. Тоді $\forall \boldsymbol{x} \in X$ за означенням субградієнта має місце умова

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \ge (\boldsymbol{0}, (\boldsymbol{x} - \boldsymbol{x}^*)),$$

тобто $\forall \boldsymbol{x} \in X\ \ f(\boldsymbol{x}) \ge f(\boldsymbol{x}^*)$, що і треба було довести.

*<u>Наслідок</u>***.** *Якщо функція $f(\boldsymbol{x}^*)$ опукла і диференційовна на опуклій множині $X \subset E^n$, то умова*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0},\ \boldsymbol{x}^* \in \operatorname{int} X,$$

є необхідною і достатньою умовою абсолютного мінімуму функції $f(\boldsymbol{x})$ *в точці* $\boldsymbol{x}^*$.

Доведення. Дійсно, у випадку $f(\boldsymbol{x}) \in C_X^1$ множина субградієнтів функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^*$ є одноелементною, її утворює єдиний градієнт $\nabla f(\boldsymbol{x}^*)$, який в той же час є і субградієнтом. За умовою наслідку він є нулем, а, отже, виконуються умови попередньої теореми.

Лекція 31. *Теорема Куна-Таккера. Розв'язування задач квадратичного програмування. – 2год. [1].*

Завдання для самостійної роботи. *Задача опуклого програмування. Достатні умови оптимальності в задачах НП. Умови регулярності. Теорема Куна-Таккера. Умови оптимальності в задачах ОП (випадок диференційовності всіх функцій задачі), їх застосування до розв'язування задач квадратичного програмування. -4год. [1-3,7,9].*

Опукле програмування

Загальна теорія. Теорема Куна-Таккера

Розглянемо задачу

$$\mathbf{C}: \min\{f^0(\boldsymbol{x}): f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x} \in X\},$$

де X опукла множина. Якщо функції $f^i(\boldsymbol{x})$, $i=0,1,\dots,m$, опуклі на X, то задачу **С** *називають задачею опуклого програмування (ЗОП). До опуклого програмування відносять також задачі максимізації вигляду*

$$\max\{g^0(\boldsymbol{x}): g^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x} \in X\},$$

якщо X опукла множина, функція $g^0(\boldsymbol{x})$ опукла доверху на X, а система обмежень (тип яких, тобто, "≤", "=", "≥" не є суттєвим) може бути зведена до вигляду $f^i(\boldsymbol{x}) \le 0$, $i=1,\dots,m$, $\boldsymbol{x} \in X$, де функції $f^i(\boldsymbol{x})$, $i=1,\dots,m$, опуклі донизу на X.

Як і у класичних задачах оптимізації *функцією Лагранжа* задачі **С** назвемо функцію

$$L(\boldsymbol{x},\boldsymbol{u}) = f^0(\boldsymbol{x}) + \sum_{i=1}^{m} u_i f^i(\boldsymbol{x}),$$

де $\boldsymbol{x} \in X$, $\boldsymbol{u} = (u_1, u_2, \dots, u_m) \ge 0$.

Означення 9.1. *Пара векторів $(\boldsymbol{x}^*, \boldsymbol{u}^*)$ називається сідловою точкою функції Лагранжа на множині $\boldsymbol{x} \in X$, $\boldsymbol{u} \ge 0$, якщо $\boldsymbol{x}^* \in X$, $\boldsymbol{u}^* \ge 0$ і для довільних $\boldsymbol{x} \in X$, $\boldsymbol{u} \ge 0$*

$$L(\boldsymbol{x}^*,\boldsymbol{u}) \le L(\boldsymbol{x}^*,\boldsymbol{u}^*) \le L(\boldsymbol{x},\boldsymbol{u}^*). \quad (9.1)$$

Теорема 9.1 (достатні умови оптимальності). *Якщо функція Лагранжа задачі* **С** *має сідлову точку $(\boldsymbol{x}^*, \boldsymbol{u}^*)$*, то $\boldsymbol{x}^*$ є оптимальним розв'язком задачі **С**, *і при цьому виконується правило доповнюючої нежорсткості*

$$\sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) = 0. \quad (9.2)$$

Доведення. За означенням сідлової точки маємо: $\boldsymbol{x}^* \in X$, $\boldsymbol{u}^* \ge 0$ і для довільних $\boldsymbol{x} \in X$, $\boldsymbol{u} \ge 0$ виконується нерівність

$$f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} u_i f^i(\boldsymbol{x}^*) \le f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) \le f^0(\boldsymbol{x}) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}). \tag{9.3}$$

Треба довести, що $\boldsymbol{x}^* \in D_C = \{\boldsymbol{x} \in E^n: f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x} \in X\}$ і що для довільних $\boldsymbol{x} \in D_C$ буде $f^0(\boldsymbol{x}^*) \le f^0(\boldsymbol{x})$.

Розглянемо ліву частину нерівності (9.3). Маємо для довільних $\boldsymbol{u} \ge 0$

$$\sum_{i=1}^{m} u_i f^i(\boldsymbol{x}^*) \le \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*). \tag{9.4}$$

Із цієї нерівності випливає, що

$$f^i(\boldsymbol{x}^*) \le 0,\ i=1,\dots,m. \tag{9.5}$$

Дійсно, якби для деякого k було б $f^k(\boldsymbol{x}^*) > 0$, то, поклавши $u_i = 0$ при $i \ne k$ і $u_k = M$, де $M > 0$ як завгодно велике число, ми порушили б нерівність (9.4), а разом з нею і умови теореми. Отже, $\boldsymbol{x}^* \in D_C$.

Покладемо у нерівності (9.4) всі $u_i = 0,\ i=1,\dots,m$. Отримаємо

$$0 \le \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*). \tag{9.6}$$

З іншого боку, помноживши кожну з нерівностей (9.5) на відповідне $u_i^* \ge 0$ та додавши їх, отримаємо

$$\sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) \le 0. \tag{9.7}$$

Із (9.6) та (9.7) випливає правило доповнюючої нежорсткості (9.2).

Розглянемо тепер праву частину нерівності (9.3). Врахувавши умову (9.2), для довільних $\boldsymbol{x} \in X$ отримаємо

$$f^0(\boldsymbol{x}^*) \le f^0(\boldsymbol{x}) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}). \tag{9.8}$$

В той же час для довільних $\boldsymbol{x} \in D_C$

$$f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m. \tag{9.9}$$

Тоді, помноживши кожну з нерівностей (9.9) на відповідне $u_i^* \ge 0$ та додавши їх, отримаємо

$$\sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}) \le 0. \tag{9.10}$$

Далі, розглядаючи нерівність (9.8) на множині $D_C \subset X$ та враховуючи (9.10), остаточно отримаємо: $f^0(\boldsymbol{x}^*) \le f^0(\boldsymbol{x})\quad \forall \boldsymbol{x} \in D_C$, тобто $\boldsymbol{x}^*$ є оптимальним розв'язком задачі **C**

Теорема доведена.

Умови регулярності

Означення 9.2. *Якщо для кожного $i=1,\dots,m$ існує точка $\boldsymbol{x}^i \in D_C$, в якій*

$$f^i(\boldsymbol{x}^i) < 0, \tag{9.11}$$

то говорять, що множина D_C задовольняє умову регулярності.

Цю умову ми будемо використовувати в еквівалентній формі, яка називається *умовою регулярності Слейтера.*

Означення 9.3. *Якщо існує точка* $\boldsymbol{x} \in D_C$, *в якій*

$$f^i(\boldsymbol{x}) < 0,\ i = 1,\ldots,m, \tag{9.12}$$

то говорять, що множина D_C *задовольняє умову Слейтера.*

Геометричний зміст умови Слейтера дуже простий множина внутрішніх точок множини D_C, яка задовольняє умову Слейтера, непорожня: $\operatorname{int} D_C \neq \varnothing$.

Теорема 9.2 (про еквівалентність умов регулярності).
Умова регулярності (9.11) та умова Слейтера (9.12) еквівалентні.

Доведення. Із умови Слейтера безпосередньо випливає умова регулярності, для цього досить $\forall i$ покласти $\boldsymbol{x}^i = \boldsymbol{x}$.

Щоб довести, що із умови регулярності випливає умова Слейтера, покладемо

$$\boldsymbol{x} = \sum_{k=1}^{m} \alpha_k \boldsymbol{x}^k,\quad \alpha_k \geq 0,\ k = 1,\ldots,m,\ \sum_{k=1}^{m} \alpha_k = 1.$$

Оскільки всі $\boldsymbol{x}^k \in D_C$, а D_C опукла множина, то $\boldsymbol{x} \in D_C$.

Скористаємось нерівністю Ієнсена для кожного $i=1,\ldots,m$. Маємо при $i=1,\ldots,m$

$$f^i(\boldsymbol{x}) = f^i\left(\sum_{k=1}^{m} \alpha_k \boldsymbol{x}^k\right) \leq \sum_{k=1}^{m} \alpha_k f^i(\boldsymbol{x}^k) < 0$$

внаслідок того, що всі $\alpha_k \geq 0$ і завжди можна вибрати $\alpha_k > 0$ при $k=i$ $\forall i$, а при $k=i$ виконується $f^i(\boldsymbol{x}^i) < 0$ за умовою регулярності.

Теорема доведена.

Теорема 9.3 (Куна-Таккера). *Нехай*

$$\mathbf{C}: \min\{f^0(\boldsymbol{x}) : f^i(\boldsymbol{x}) \leq 0,\ i = 1,\ldots,m,\ \boldsymbol{x} \in X\}$$

задача опуклого програмування, допустима множина

$$D_C = \{\boldsymbol{x} \in E^n : f^i(\boldsymbol{x}) \leq 0,\ i = 1,\ldots,m,\ \boldsymbol{x} \in X\}$$

якої задовольняє умову регулярності Слейтера.

Допустимий розв'язок $\boldsymbol{x}^* \in D_C$ *задачі* **C** *є її оптимальним розв'язком тоді і тільки тоді, коли існує невід'ємний вектор* $\boldsymbol{u}^* \geq 0$ *такий, що точка* $(\boldsymbol{x}^*, \boldsymbol{u}^*)$ *є сідловою точкою функції Лагранжа задачі* **C** *на множині* $\boldsymbol{x} \in X$, $\boldsymbol{u} \geq 0$.

Доведення.

Достатність випливає з теореми про достатні умови оптимальності.

Необхідність. Нехай $\boldsymbol{x}^*$ є оптимальним розв'язком задачі **C**, тобто для довільних $\boldsymbol{x} \in D_C$ виконується $f^0(\boldsymbol{x}^*) \leq f^0(\boldsymbol{x})$.

Введемо у просторі E^{m+1} множини Y і Z за допомогою співвідношень

$$Z = \{\boldsymbol{z} = (z_0, z_1, \ldots, z_m) \in E^{m+1} : z_0 < f^0(\boldsymbol{x}^*),\ z_i < 0,\ i = 1,\ldots,m\},$$

$$Y = \bigcup_{\boldsymbol{x} \in X} Y(\boldsymbol{x}),$$

де $Y(\boldsymbol{x})$ для довільного $\boldsymbol{x} \in X$ визначається так:

$$Y(\boldsymbol{x}) = \{\boldsymbol{y} = (y_0, y_1, \ldots, y_m) \in E^{m+1}:\ f^{\,i}(\boldsymbol{x}) \le y_i,\ i=0,1,\ldots,m\}.$$

Геометрична інтерпретація множин Z та $Y(\boldsymbol{x})$ приводиться на рис. 9.1.

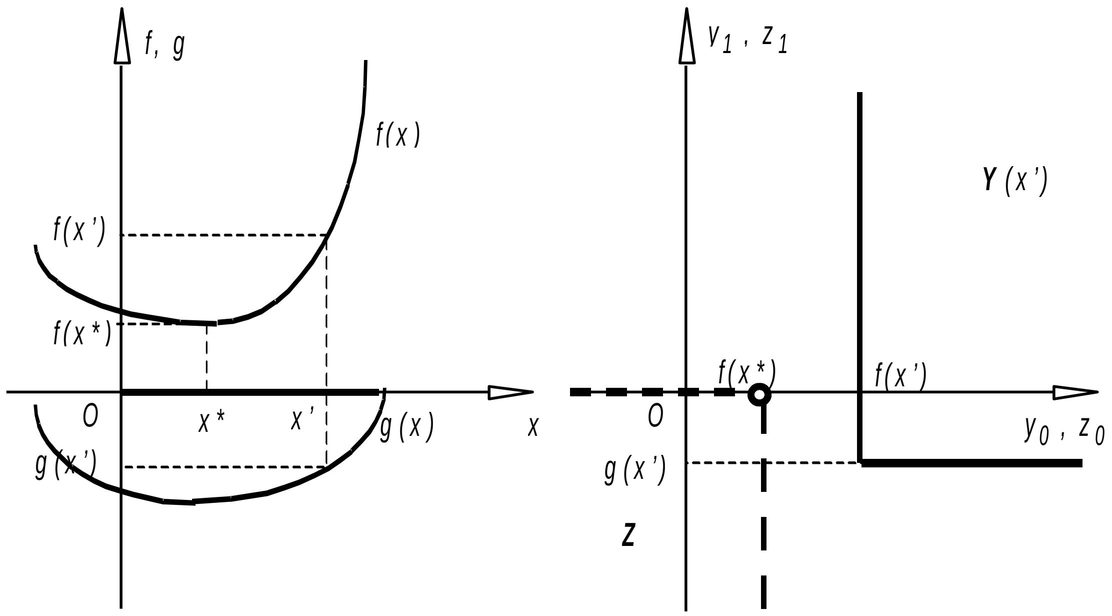


Рис. 9.1

$f(x) \to \min$, $g(x) \le 0$, $x \ge 0$.
$Z = \{\boldsymbol{z} = (z_0, z_1): z_0 < f(x^*),\ z_1 < 0\}$.
$Y = \{\boldsymbol{y} = (y_0, y_1): y_0 \ge f(x),\ y_1 \ge g(x),\ x \ge 0\}$,

Покажемо, що множини Z і Y опуклі.

Опуклість множини Z очевидна: Z є перетином скінченного числа $m+1$ півпросторів у просторі E^{m+1}.

Аналогічно стверджуємо, що для довільного $\boldsymbol{x} \in X$ множина $Y(\boldsymbol{x})$ також опукла як перетин скінченного числа $m+1$ півпросторів у просторі E^{m+1}.

Далі, для довільних $\boldsymbol{y}^1, \boldsymbol{y}^2 \in Y$ утворимо їх опуклу лінійну комбінацію $\overline{\boldsymbol{y}} = \alpha \boldsymbol{y}^1 + (1-\alpha)\boldsymbol{y}^2$, $\alpha \in [0,1]$, і покажемо, що $\overline{\boldsymbol{y}}$ є елементом множини Y.

Оскільки $\boldsymbol{y}^1 \in Y$, то існує $\boldsymbol{x}^1 \in X$, для якого $\boldsymbol{y}^1 \in Y(\boldsymbol{x}^1)$. Аналогічно, існує елемент $\boldsymbol{x}^2 \in X$, для якого $\boldsymbol{y}^2 \in Y(\boldsymbol{x}^2)$. Оскільки X опукла множина, то опукла лінійна комбінація елементів $\boldsymbol{x}^1$ та $\boldsymbol{x}^2$ □ точка $\overline{\boldsymbol{x}} = \alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2$, $\alpha \in [0,1]$, є елементом множини X.

Враховуючи опуклість функцій $f^{\,i}(\boldsymbol{x})$, $i=0,1,\ldots,m,$ отримаємо

$$f^{\,i}(\overline{\boldsymbol{x}}) = f^{\,i}(\alpha \boldsymbol{x}^1 + (1-\alpha)\,\boldsymbol{x}^2) \le \alpha f^{\,i}(\boldsymbol{x}^1) + (1-\alpha)\, f^{\,i}(\boldsymbol{x}^2) \le$$
$$\le \alpha y_i^1 + (1-\alpha)\, y_i^2 = \overline{y_i}\,.$$

Отже, $f^{\,i}(\overline{\boldsymbol{x}}) \le \overline{y_i}$, $i=0,1,\ldots,m$, тобто $\overline{\boldsymbol{y}} \in Y(\overline{\boldsymbol{x}}) \subset Y$, що означає опуклість множини Y.

Доведемо, що множини Z і Y не перетинаються. Розглянемо два випадки: 1) $\boldsymbol{x} \in D_C$ і 2) $\boldsymbol{x} \notin D_C$, але $\boldsymbol{x} \in X$.

1) Для довільних $\boldsymbol{x} \in D_C$ маємо:

- за умовами теореми: $f^0(\boldsymbol{x}^*) \le f^0(\boldsymbol{x})$,
- за означенням множини Z: $z_0 < f^0(\boldsymbol{x}^*)$,
- за означенням множини Y: $f^0(\boldsymbol{x}) \le y_0$.

Остаточно, для довільних $\boldsymbol{x} \in D_C$ отримаємо: $z_0 < f^0(\boldsymbol{x}^*) \le f^0(\boldsymbol{x}) \le y_0$, тобто $z_0 < y_0$.

Це і означає, що множини Z та Y не перетинаються.

2) Оскільки $\boldsymbol{x} \notin D_C$, то знайдеться хоча б один індекс i, для якого виконується $0 < f^i(\boldsymbol{x})$. В той же час за означенням множини Y: $f^i(\boldsymbol{x}) \le y_i$, а за означенням множини Z: $z_i < 0$.

Таким чином, завжди існує індекс i, для якого $z_i < 0 < f^i(\boldsymbol{x}) \le y_i$, тобто $z_i < y_i$. А це означає, що і в цьому випадку множини Z та Y не перетинаються.

Отже, множини Z та Y опуклі і не мають спільних точок. Тоді на основі теореми про розділяючу гіперплощину та її наслідку існує ненульовий вектор $\boldsymbol{c} = (c_0, c_1, \ldots, c_m) \ne 0$ такий, що

$$(\boldsymbol{c}, \boldsymbol{z}) \le (\boldsymbol{c}, \boldsymbol{y}), \tag{9.13}$$

для довільних $z \in \overline{Z}$, $y \in \overline{Y}$, де $\overline{Z}$ і $\overline{Y}$ □ замикання, відповідно, множин Z і Y.

Оскільки множині $\overline{Z}$ належать точки з як завгодно великими по модулю від'ємними координатами, то із того, що нерівність (9.13) має місце для довільних $y \in \overline{Y}$, випливає умова: $\boldsymbol{c} = (c_0, c_1, \ldots, c_m) \ge \boldsymbol{0}$.

В іншому випадку, тобто коли б знайшлося хоча б одне $c_i < 0$ серед координат вектора $\boldsymbol{c}$, то ліва частина нерівності (9.13) стала б необмеженою на множині $\overline{Z}$ зверху, а, значить, стала б неможливою для довільних $y \in \overline{Y}$ нерівність (9.13).

Зазначимо, що нерівність (9.13) виконується і в граничних точках множин Z та Y. Тому, вибравши для довільних $\boldsymbol{x} \in X$: $y_0 = f^0(\boldsymbol{x})$, $z_0 = f^0(\boldsymbol{x}^*)$, $y_i = f^i(\boldsymbol{x})$, $z_i = 0$, $i = 1, \ldots, m$, отримаємо із (9.13) для довільних $\boldsymbol{x} \in X$

$$c_0 f^0(\boldsymbol{x}^*) \le c_0 f^0(\boldsymbol{x}) + \sum_{i=1}^{m} c_i f^i(\boldsymbol{x}). \tag{9.15}$$

Впевнимося, що $c_0 > 0$. Припустимо противне, тобто, що $c_0 = 0$. Тоді з (9.15) для довільних $\boldsymbol{x} \in X$ отримаємо

$$0 \le \sum_{i=1}^{m} c_i f^i(\boldsymbol{x}). \tag{9.16}$$

Тим більше нерівність (9.16) матиме місце для довільних $\boldsymbol{x} \in D_C \subset X$.

З іншого боку для довільних $\boldsymbol{x} \in D_C$ завжди

$$f^i(\boldsymbol{x}) \le 0,\ i = 1, \ldots, m. \tag{9.17}$$

Помноживши кожну з нерівностей (9.17) на відповідне $c_i \ge 0$ і додавши їх, отримаємо для довільних $\boldsymbol{x} \in D_C$

$$\sum_{i=1}^{m} c_i f^i(\boldsymbol{x}) \le 0. \tag{9.18}$$

Із (9.16) та (9.18) витікає, що для довільних $\boldsymbol{x} \in D_C$ має місце рівність

$$\sum_{i=1}^{m} c_i f^i(\boldsymbol{x}) = 0. \qquad (9.19)$$

Кожний доданок цієї суми недодатний, тому сума буде дорівнювати нулю лише тоді, коли всі її доданки для будь-яких $\boldsymbol{x} \in D_C$ дорівнюють нулю

$$c_i f^i(\boldsymbol{x}) = 0,\ i=1,\ldots,m. \qquad (9.20)$$

Оскільки $\boldsymbol{c} = (c_0, c_1, \ldots, c_m) \neq 0$ і всі $c_i \geq 0$, $i=1,\ldots,m$, то умова (9.20) буде виконуватись тільки тоді, коли для тих i, для яких $c_i > 0$, буде $f^i(\boldsymbol{x}) = 0\ \forall \boldsymbol{x} \in D_C$. А це суперечить умові регулярності Слейтера для множини D_C, за якою існує допустима точка $\bar{\boldsymbol{x}} \in D_C$ така, що $f^i(\bar{\boldsymbol{x}}) < 0$, $i=1,\ldots,m$. Отже, наше припущення невірне і $c_0 > 0$.

Покладемо $u_i^* = c_i / c_0$, $i=1,\ldots,m$, $\boldsymbol{u}^* = (u_1^*, \ldots, u_m^*)$. Очевидно, що $\boldsymbol{u}^* \geq \boldsymbol{0}$. Тоді нерівність (9.15) набуде такого вигляду для довільних $\boldsymbol{x} \in X$

$$f^0(\boldsymbol{x}^*) \leq f^0(\boldsymbol{x}) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}). \qquad (9.21)$$

Покладемо в (9.21) $\boldsymbol{x} = \boldsymbol{x}^*$. Отримаємо

$$0 \leq \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*). \qquad (9.22)$$

За умовою теореми $\boldsymbol{x}^* \in D_C$, тобто $f^i(\boldsymbol{x}^*) \leq 0$, $i=1,\ldots,m$, наслідком чого при довільних $u_i \geq 0$, $i=1,\ldots,m$, є нерівність

$$\sum_{i=1}^{m} u_i f^i(\boldsymbol{x}^*) \leq 0. \qquad (9.23)$$

При $u_i = u_i^*$, $i=1,\ldots,m$, із (9.23) отримаємо

$$\sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) \leq 0. \qquad (9.24)$$

Система нерівностей (9.22), (9.24) дає умову доповнюючої нежорсткості

$$\sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) = 0. \qquad (9.25)$$

Враховуючи (9.25) із (9.21) отримаємо для довільних $\boldsymbol{x} \in X$

$$f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*) \leq f^0(\boldsymbol{x}) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x})$$

або

$$L(\boldsymbol{x}^*, \boldsymbol{u}^*) \leq L(\boldsymbol{x}, \boldsymbol{u}^*). \qquad (9.26)$$

Враховуючи (9.23), (9.25) отримаємо $\forall u_i \geq 0$, $i=1,\ldots,m$,

$$f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} u_i f^i(\boldsymbol{x}^*) \leq f^0(\boldsymbol{x}^*) + \sum_{i=1}^{m} u_i^* f^i(\boldsymbol{x}^*)$$

або

$$L(\boldsymbol{x}^*, \boldsymbol{u}) \le L(\boldsymbol{x}^*, \boldsymbol{u}^*). \qquad (9.27)$$

Порівняння (9.26) і (9.27) дає подвійну нерівність

$$L(\boldsymbol{x}^*, \boldsymbol{u}) \le L(\boldsymbol{x}^*, \boldsymbol{u}^*) \le L(\boldsymbol{x}, \boldsymbol{u}^*)$$

$\forall \boldsymbol{x} \in X$ та $\forall \boldsymbol{u} \ge 0$, тобто $(\boldsymbol{x}^*, \boldsymbol{u}^*)$ є сідловою точкою функції Лагранжа задачі **C**.

Теорема доведена.

Зауважимо, що необхідною і достатньою умовою існування сідлової точки скалярної функції двох векторних аргументів, як було доведено раніше (див. розділ 5), є рівність мінімаксів

$$\min_{\boldsymbol{x}\in X} \max_{\boldsymbol{u}\ge 0} L(\boldsymbol{x},\boldsymbol{u}) = \max_{\boldsymbol{u}\ge 0} \min_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{u}), \qquad (9.28)$$

причому компоненти сідлової точки $\boldsymbol{x}^*$ та $\boldsymbol{u}^*$ є, відповідно, точками зовнішніх екстремумів в мінімаксах

$$\begin{aligned} \boldsymbol{x}^* &= \arg\min_{\boldsymbol{x}\in X} (\max_{\boldsymbol{u}\ge 0} L(\boldsymbol{x},\boldsymbol{u})), \\ \boldsymbol{u}^* &= \arg\max_{\boldsymbol{u}\ge 0} (\min_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{u})). \end{aligned} \qquad (9.29)$$

Значення функції $L(\boldsymbol{x}^*, \boldsymbol{u}^*)$ в сідловій точці дорівнює спільному значенню мінімаксів.

Зауваження до теореми Куна-Таккера.

Теорема Куна-Таккера дає необхідні і достатні умови оптимальності в задачах опуклого програмування (*ОП*). Якщо умову Слейтера відкинути, то теорема Куна-Таккера дає тільки достатні умови оптимальності.

Приклад 9.1. *Нехай маємо задачу*

$$\mathbf{C}: \min\{-x:\ x^2 \le 0,\ x \ge 0\}.$$

Це задача *ОП*. Умова Слейтера для допустимої області задачі **C** не виконується, оскільки множина її внутрішніх точок порожня $\{x: x^2 < 0\} = \varnothing$. Сама допустима множина складається із однієї точки $x = 0$. Оптимальний розв'язок задачі **C** існує і рівний нулю: $x^* = 0$.

Розглянемо функцію Лагранжа $L(x,u) = -x + ux^2$ цієї задачі в області $x \ge 0$, $u \ge 0$. В сідловій точці (x^*, u^*), якщо вона існує, функція Лагранжа досягає мінімуму по x при $u = u^*$ і максимуму по u при $x = x^*$.

Знайдемо стаціонарні точки функції $L(x,u)$

$$\begin{cases} \dfrac{\partial L}{\partial x} = 2xu - 1 = 0, \\ \dfrac{\partial L}{\partial u} = x^2 = 0. \end{cases}$$

Маємо $x = 0$, але не існує таких $u \ge 0$, для яких виконувалось би перше рівняння. Отже, функція Лагранжа $L(x,u)$ не має стаціонарних точок, а тому вона не має і сідлових точок.

Отже, остаточно, розв'язування задачі математичного програмування еквівалентне відшуканню сідлової точки функції Лагранжа цієї задачі за умови, що така точка існує.

Тому важливими для практичного застосування є необхідні і достатні умови існування у функції сідлової точки. Нагадаємо, що у загальному випадку необхідною і достатньою умовою існування сідлової точки у скалярної функції двох векторних аргументів є рівність мінімаксів.

Сформулюємо такі умови для задачі *ОП*

$$\mathbf{C}: \min\{f^0(\boldsymbol{x}) : f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x} \ge 0\}$$

у випадку диференційовності функцій $f^i(\boldsymbol{x})$, $i=0,1,\dots,m$.

Теорема 9.5 (необхідні і достатні умови існування сідлової точки функції Лагранжа задачі ОП).

Якщо функції $f^i(\boldsymbol{x}) \in C_X^1$, $i=0,1,\dots,m$, *де* $X=\{\boldsymbol{x}\in E^n : \ \boldsymbol{x}\ge 0\}$, *то необхідною і достатньою умовою того, що точка* $(\boldsymbol{x}^*,\boldsymbol{u}^*)$ *є сідловою точкою функції Лагранжа задачі* **C** *в області* $\boldsymbol{x}\ge 0$, $\boldsymbol{u}\ge 0$ *є виконання умов*

$$\begin{cases} \nabla_X L(\boldsymbol{x}^*,\boldsymbol{u}^*) \ge 0, & (\boldsymbol{x}^*, \nabla_X L(\boldsymbol{x}^*,\boldsymbol{u}^*)) = 0, \quad \boldsymbol{x}^* \ge 0, \\ \nabla_U L(\boldsymbol{x}^*,\boldsymbol{u}^*) \le 0, & (\boldsymbol{u}^*, \nabla_U L(\boldsymbol{x}^*,\boldsymbol{u}^*)) = 0, \quad \boldsymbol{u}^* \ge 0, \end{cases} \qquad (9.36)$$

де $\nabla_X L(\boldsymbol{x},\boldsymbol{u})$ *і* $\nabla_U L(\boldsymbol{x},\boldsymbol{u})$ □ *градієнти функції* $L(\boldsymbol{x},\boldsymbol{u})$ *відповідно по* $\boldsymbol{x}$ *та по* $\boldsymbol{u}$.

Без доведення (доведення див. [15], стор. 64).

Зазначимо, що якщо допустима множина задачі **C**

$$D_C = \{\boldsymbol{x}\in E^n : f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x}\ge 0\}$$

задовольняє умову регулярності Слейтера, то умови (9.36) будуть необхідними і достатніми умовами існування оптимальної точки $\boldsymbol{x}^*$ задачі **C**. Враховуючи це, можна вказати такий *спосіб розв'язування задачі ОП* для випадку, що розглядається:

а) побудувати функцію Лагранжа задачі **C**;

б) записати для неї достатні умови оптимальності (9.36), які використовують, як правило, в еквівалентному вигляді

$$\begin{cases} \frac{\partial L(x,u)}{\partial x_j} \ge 0, \quad j=1,\dots,n, \qquad x_j \frac{\partial L(x,u)}{\partial x_j} = 0, \quad j=1 \\ \frac{\partial L(x,u)}{\partial u_i} = f^i(x) \le 0, \quad i=1,\dots,m, \qquad u_i \frac{\partial L(x,u)}{\partial u_i} = \\ x_j \ge 0, \quad j=1,\dots,n, \qquad u_i \ge 0, \quad i=1,\dots,m; \end{cases}$$

(9.37)

в) розв'язати систему (9.37) відносно $\boldsymbol{x}$ (звичайно, якщо вона має розв’язок взагалі) і, тим самим, знайти розв’язок $\boldsymbol{x}^*$ задачі **C** .

Задача опуклого квадратичного програмування

Зауважимо, що процедуру розв'язування системи (9.37) для довільних опуклих диференційовних функцій $f^i(\boldsymbol{x})$, $i=0,1,\dots,m$, реалізувати практично неможливо (мається на увазі аналітично). В той же час ця система ефективно розв'язується, якщо вона є лінійною. Але вона буде лінійною лише для задачі лінійного програмування, яку взагалі недоцільно розв’язувати таким шляхом. Однак система (9.37) може бути частково лінійною. Це буде у випадку, коли $f^0(\boldsymbol{x})$ □ квадратична функція, а $f^i(\boldsymbol{x})$, $i=1,\dots,m$, □ лінійні функції, тобто, коли оптимізаційна задача буде задачею квадратичного програмування.

Отже, розглянемо *задачу опуклого квадратичного програмування* у вигляді

$$f^0(\boldsymbol{x}) = (\boldsymbol{c},\boldsymbol{x}) + (\boldsymbol{x},\ \boldsymbol{D}\boldsymbol{x}) \to \min, \tag{9.38}$$

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \le \boldsymbol{0}, \tag{9.39}$$

$$\boldsymbol{x} \ge \boldsymbol{0}, \tag{9.40}$$

де $\boldsymbol{c}\in E^n$, $\boldsymbol{x}\in E^n$, $\boldsymbol{D}=\|d_{ij}\|$, $i,j=1,\ldots,n$, – симетрична невід'ємно визначена матриця, $\boldsymbol{A}=\|a_{ij}\|$, $i=1,\ldots,m$, $j=1,\ldots,n$, – прямокутна матриця виміру $m\times n$, $\boldsymbol{b}\in E^m$.

Побудуємо функцію Лагранжа задачі (9.38)–(9.40)

$$L(\boldsymbol{x},\boldsymbol{u}) = (\boldsymbol{c},\boldsymbol{x}) + (\boldsymbol{x},\boldsymbol{D}\boldsymbol{x}) + (\boldsymbol{u},\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

Знайдемо градієнти функції Лагранжа по $\boldsymbol{x}$ та по $\boldsymbol{u}$

$$\nabla_x L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{c} + 2\boldsymbol{D}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{u},\ \nabla_u L(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

та запишемо необхідні і достатні умови існування сідлової точки

$$\boldsymbol{c} + 2\boldsymbol{D}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{u} \ge \boldsymbol{0}, \tag{9.41}$$

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \le \boldsymbol{0}, \tag{9.42}$$

$$(\boldsymbol{c} + 2\boldsymbol{D}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{u},\boldsymbol{x}) = 0, \tag{9.43}$$

$$(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b},\boldsymbol{u}) = 0, \tag{9.44}$$

$$\boldsymbol{x} \ge \boldsymbol{0},\ \boldsymbol{u} \ge \boldsymbol{0}.$$

Введемо невід'ємні вектори балансних змінних $\boldsymbol{v}=(v_1,\ldots,v_n)\ge\boldsymbol{0}$ та $\boldsymbol{w}=(w_1,\ldots,w_n)\ge\boldsymbol{0}$, за допомогою яких нерівності (9.41), (9.42) перетворимо в рівності. Отримаємо систему

$$\boldsymbol{c} + 2\boldsymbol{D}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0}, \tag{9.45}$$

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} + \boldsymbol{w} = \boldsymbol{0}, \tag{9.46}$$

$$\boldsymbol{x} \ge \boldsymbol{0},\ \boldsymbol{u} \ge \boldsymbol{0},\ \boldsymbol{v} \ge \boldsymbol{0},\ \boldsymbol{w} \ge \boldsymbol{0}. \tag{9.47}$$

Із (9.45), (9.46) маємо $\boldsymbol{v}=\boldsymbol{c}+2\boldsymbol{D}\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{u}$, $\boldsymbol{w}=-\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$. Тому рівності (9.43), (9.44) набувають вигляду

$$(\boldsymbol{v},\boldsymbol{x}) = 0,\ (\boldsymbol{w},\boldsymbol{u}) = 0$$

або в силу умов (9.47)

$$v_j\ x_j = 0,\ j=1,\ldots,n, \tag{9.48}$$

$$w_i\ u_i = 0,\ i=1,\ldots,m. \tag{9.49}$$

Зауважимо, що деякі з обмежень (9.39) можуть бути рівностями. Тоді на відповідні їм множники Лагранжа u_i не накладаються умови невід'ємності. Крім того, оскільки в цьому випадку відповідні таким обмеженням умови (9.49) виконуватимуться тотожно, то нема потреби їх враховувати і тому їх можна відкинути.

Отже, задача відшукання сідлової точки функції Лагранжа задачі (9.38)–(9.40) звелась до відшукання розв'язку системи (9.45), (9.46) за умов (9.48), (9.49). Система (9.45), (9.46) складається з $m+n$ лінійних рівнянь відносно $2(m+n)$ невідомих x_j, v_j ($j=1,\ldots,n$), u_i, w_i ($i=1,\ldots,m$). Крім того, як випливає із (9.48), (9.49), довільний розв'язок цієї системи повинен задовольняти умови:

$\forall j$, якщо $x_j > 0$, то $v_j = 0$, або, якщо $v_j > 0$, то $x_j = 0$; (9.50)

$\forall i$, якщо $u_i > 0$, то $w_i = 0$, або, якщо $w_i > 0$, то $u_i = 0$. (9.51)

Тоді шуканим розв'язком системи (9.45), (9.46) може бути довільний допустимий базисний її розв'язок, для якого змінні x_j та v_j з однаковими індексами j, а також змінні u_i та w_i з однаковими індексами i, не є водночас базисними. Знайти такий розв'язок можна шляхом побудови для системи (9.45), (9.46) допоміжної задачі методом штучного базису з подальшим розв'язуванням її симплекс-методом.

Оптимальний розв'язок допоміжної задачі може не задовольняти умови (9.50), (9.51). Тоді потрібно добитись виконання умов (9.50), (9.51) проведенням необхідної кількості додаткових ітерацій симплекс-методу по заміні векторів базису оптимального розв'язку. При цьому для введення в базис можна вибирати лише ті небазисні вектори, які мають нульові оцінки. Це гарантуватиме незмінність значення цільової функції.

Приклад 9.1. *Розв'язати задачу*

$$z = -5x_1 - 2x_2 + x_1^2 - x_1 x_2 + x_2^2 \to \min,$$
$$2x_1 + 3x_2 \le 15,$$
$$x_1 + 2x_2 = 8,$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Функція $z = z(x_1, x_2)$ опукла, оскільки

$$z_{11} = 2 > 0,\quad \begin{vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 > 0,\quad \text{де } z_{ij} = \frac{\partial^2 z}{\partial x_i \partial x_j},$$

і за критерієм Сильвестра матриця

$$\boldsymbol{D} = \begin{Vmatrix} 1 & -1/2 \\ -1/2 & 1 \end{Vmatrix}$$

додатно визначена, а, отже, додатно визначеною є і квадратична форма $x_1^2 - x_1 x_2 + x_2^2$. Тому задача, що розглядається, є задачею опуклого квадратичного програмування.

Будуємо її функцію Лагранжа: $L(x_1, x_2, u_1, u_2) =$

$$= -5x_1 - 2x_2 + x_1^2 - x_1 x_2 + x_2^2 + u_1(2x_1 + 3x_2 - 15) + u_2(x_1 + 2x_2 - 8)$$

та записуємо необхідні і достатні умови існування сідлової точки

$$\begin{aligned}
&\frac{\partial L}{\partial x_1} = -5 + 2x_1 - x_2 + 2u_1 + u_2 \ge 0, && \frac{\partial L}{\partial x_1} x_1 = 0,\\
&\frac{\partial L}{\partial x_2} = -2 - x_1 + 2x_2 + 3u_1 + 2u_2 \ge 0, && \frac{\partial L}{\partial x_2} x_2 = 0,\\
&\frac{\partial L}{\partial u_1} = 2x_1 + 3x_2 - 15 \le 0, && \frac{\partial L}{\partial u_1} u_1 = 0,\\
&\frac{\partial L}{\partial u_2} = x_1 + 2x_2 - 8 = 0, && x_1 \ge 0,\ x_2 \ge 0,\ u_1 \ge 0.
\end{aligned} \qquad (9.52)$$

Зауважимо, що умова $\frac{\partial L}{\partial u_2} u_2 = 0$ виконується тотожно, оскільки $\frac{\partial L}{\partial u_2} = 0$, тому її можна відкинути.

Введемо балансні змінні $v_1 \ge 0$, $v_2 \ge 0$, $w_1 \ge 0$ та перепишемо систему умов (9.52) у вигляді системи рівнянь

[illegible]

(9.53)

та умов

$$x_1 v_1 = 0, x_2 v_2 = 0, \; u_1 w_1 = 0. \qquad (9.54)$$

Змінна u_2 вільна за знаком, тому подамо її у вигляді $u_2 = u_3 - u_4$, де $u_3 \ge 0$, $u_4 \ge 0$, та перепишемо систему (9.53)

[illegible]

Допоміжну задачу для відшукання допустимого базисного розв'язку цієї системи будуємо методом штучного базису. Маємо

[illegible]

(9.55)

Зазначимо, що шуканий розв'язок повинен задовольняти умови (9.54).

Задачу (9.55) розв'язуємо симплекс-методом (див. таблицю 9.1). Символом "*" позначені дані, які можна не обчислювати. Остання ітерація здійснена з метою задоволення умови $u_1 w_1 = 0$.

Оптимальний розв'язок допоміжної задачі $x_1 = 24/7$, $x_2 = 16/7$, $u_3 = 3/7$, $w_1 = 9/7$, $u_1 = u_4 = v_1 = v_2 = y_1 = y_2 = y_3 = 0$ визначає оптимальний розв'язок вихідної задачі квадратичного програмування: $x_1 = 24/7$, $x_2 = 16/7$. Оптимальне значення цільової функції рівне $z = -88/7$.

Таблиця 9.1.

$x_б$	x_1	x_2	u_1	u_3	u_4	v_1	v_2	w1	y_1	y_2	y_3	β	θ
y_1	2	1	2	1	1	1	0	0	1	0	0	5	5/2
y_2	1	2	**3**	2	2	0	1	0	0	1	0	2	2/3
w_1	2	3	0	0	0	0	0	1	0	0	0	15	
y_3	1	2	0	0	0	0	0	0	0	0	1	8	
Δ	2	3	5	3	3	1	1	0	0	0	0	15	
y_1	**8/3**	7/3	0	1/3	1/3	1	2/3	0	1	*	0	11/3	11/8
u_1	1/3	2/3	1	2/3	2/3	0	1/3	0	0	*	0	2/3	
w_1	2	3	0	0	0	0	0	1	0	*	0	15	15/2
y_3	1	2	0	0	0	0	0	0	0	*	1	8	8
Δ	11/3	1/3	0	1/3	1/3	1	2/3	0	0	*	0	35/3	
x_1	1	7/8	0	1/8	1/8	3/8	2/8	0	*	*	0	11/8	

u_1	0	3/8	1	5/8	5/8	1/8	2/8	0	*	*	0	9/8	3
w_1	0	38/8	0	2/8	2/8	6/8	4/8	1	*	*	0	98/8	98/38
y_3	0	**23/8**	0	1/8	1/8	3/8	2/8	0	*	*	1	53/8	53/23
Δ	0	23/8	0	1/8	1/8	3/8	2/8	0	*	*	0	53/8	
x_1	1	0	0	2/23	2/23	6/23	4/23	0	*	*	*	78/23	
u_1	0	0	1	**14/23**	14/23	4/23	5/23	0	*	*	*	6/23	3/7
w_1	0	0	0	1/23	1/23	3/23	2/23	1	*	*	*	30/23	30
x_2	0	1	0	1/23	1/23	3/23	2/23	0	*	*	*	53/23	53
Δ	0	0	0	0	0	0	0	0	*	*	*	0	
x_1	1	0	*	0	*	*	*	0	*	*	*	24/7	
u_3	0	0	*	1	*	*	*	0	*	*	*	3/7	
w_1	0	0	*	0	*	*	*	1	*	*	*	9/7	
x_2	0	1	*	0	*	*	*	0	*	*	*	16/7	
Δ	0	0	*	1/3	*	*	*	0	*	*	*	0	

Лекція 32. *Градієнтні методи безумовної оптимізації.– 2год. [1].*

Завдання для самостійної роботи. *Підхожі напрямки. Градієнтні методи. Метод з подрібненням кроку. Теорема про збіжність методу з подрібненням кроку. Метод найшвидшого спуску. Теорема про збіжність методу найшвидшого спуску. Тести на зупинку процедури найшвидшого спуску. Недоліки градієнтних методів. -4год. [1-3,7,9].*

Градієнтні методи безумовної оптимізації

У цьому параграфі буде розглянута задача відшукання безумовного екстремуму диференційовної функції $z=f(\boldsymbol{x})$, $\boldsymbol{x}\in E^n$, $f(\boldsymbol{x})\in C^1$.

Ця задача (хоча б принципово) може бути розв'язана класичними методами. Ці методи називаються непрямими, оскільки вони використовують необхідні умови екстремуму $\nabla f(\boldsymbol{x})=\boldsymbol{0}$. Однак слід зауважити, що для реальних задач розв'язування цієї системи є не менш складною проблемою, ніж розв'язування вихідної задачі. Непрямі методи застосовують в основному тоді, коли розв'язок екстремальної задачі необхідно знайти в аналітичному вигляді. Для розв'язування складних практичних задач, як правило, використовують прямі методи, які зв'язані з безпосереднім порівнянням функції в двох чи більше точках.

Нехай маємо деяку точку $\boldsymbol{x}^s\in E^n$. З'ясуємо, як при розв'язуванні задачі мінімізації

$$f(\boldsymbol{x})\to\min,\ \boldsymbol{x}\in E^n,\ f(\boldsymbol{x})\in C^1,$$

перейти до нової точки $\boldsymbol{x}^{s+1}$ так, щоб виконувалась нерівність

$$f(\boldsymbol{x}^{s+1})< f(\boldsymbol{x}^s).$$

Подамо $\boldsymbol{x}^{s+1}$ у вигляді $\boldsymbol{x}^{s+1}=\boldsymbol{x}^s+\rho\boldsymbol{d}$, де вектор $\boldsymbol{d}=(d_1,\ldots,d_n)^T$ визначає напрямок зміщення, а число $\rho>0$ □ крок зміщення із точки $\boldsymbol{x}^s$ в точку $\boldsymbol{x}^{s+1}$.

Означення 10.1. *Напрямок* $\boldsymbol{d}$ *назвемо підхожим (для задачі мінімізації), якщо існує* $\rho>0$, для якого $f(\boldsymbol{x}^s+\rho\boldsymbol{d})<f(\boldsymbol{x}^s)$.

Оскільки функція $f(\boldsymbol{x})$ диференційовна, то за теоремою Тейлора маємо

$$f(\boldsymbol{x}^s+\rho\boldsymbol{d})=f(\boldsymbol{x}^s)+\rho\nabla^T f(\boldsymbol{\xi})\boldsymbol{d},$$

де $\boldsymbol{\xi}=\boldsymbol{x}^s+\theta\rho\boldsymbol{d}$, $\theta\in(0,1)$. Звідси витікає, що напрямок буде підхожим, якщо

$$\nabla^T f(\boldsymbol{\xi})\boldsymbol{d}<0. \qquad (10.1)$$

При досить малих $\rho>0$ точка $\boldsymbol{\xi}$ попадає в такий окіл точки $\boldsymbol{x}^s$, в якому функція $\nabla^T f(\boldsymbol{x})\boldsymbol{d}$ зберігає знак як неперервна ($f(\boldsymbol{x})\in C^1$). Тому в точці $\boldsymbol{x}^s$ теж буде мати місце нерівність подібна (10.1)

$$\nabla^T f(\boldsymbol{x}^s)\boldsymbol{d}<0. \qquad (10.2)$$

Отже, ми з'ясували, що при переході від точки $\boldsymbol{x}^s$ до точки $\boldsymbol{x}^{s+1}=\boldsymbol{x}^s+\rho\boldsymbol{d}$, $\rho>0$, напрямок $\boldsymbol{d}$ підхожий, якщо має місце нерівність (10.2), тобто, якщо похідна по напрямку $\boldsymbol{d}$ від функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^s$ є від'ємною:

$$D_d f(\boldsymbol{x}^s) = \nabla^T f(\boldsymbol{x}^s)\boldsymbol{d} = (\nabla f(\boldsymbol{x}^s), \boldsymbol{d}) < 0 .$$

Зауважимо, що якщо $\boldsymbol{d}$ — підхожий напрямок, то і $\rho\boldsymbol{d}$ для довільних $\rho > 0$ також є підхожим напрямком. Тому множина всіх підхожих напрямків утворює конус підхожих напрямків з вершиною в точці $\boldsymbol{x}^s$.

З'ясуємо тепер *геометричну інтерпретацію* умови (10.2).

Як відомо, градієнт $\nabla f(\boldsymbol{x}^s)$ є вектором нормалі до гіперповерхні рівня функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^s$ (у двовимірному випадку — лінії рівня), спрямованим у бік зростання функції $f(\boldsymbol{x})$.

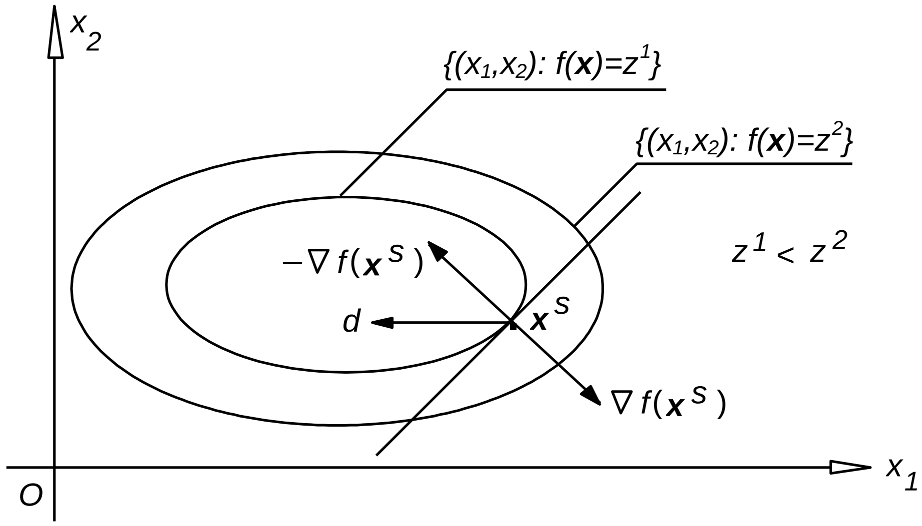


Рис. 10.1

Поряд з градієнтом розглянемо антиградієнт $-\nabla f(\boldsymbol{x}^s)$. Візьмемо за напрямок $\boldsymbol{d}$ довільний вектор з початком у точці $\boldsymbol{x}^s$, який утворює гострий кут з антиградієнтом $-\nabla f(\boldsymbol{x}^s)$. Маємо у цьому випадку

$$-\nabla^T f(\boldsymbol{x}^s)\boldsymbol{d} = (-\nabla f(\boldsymbol{x}^s), \boldsymbol{d}) = |\nabla f(\boldsymbol{x}^s)||\boldsymbol{d}|\cos \widehat{(\nabla f(\boldsymbol{x}^s), \boldsymbol{d})} > 0$$

або

$$\nabla^T f(\boldsymbol{x}^s)\boldsymbol{d} < 0 ,$$

що означає, що $\boldsymbol{d}$ — підхожий напрямок (див. рис. 10.1).

Нехай $|\boldsymbol{d}| = 1$. Тоді похідна за напрямком $\boldsymbol{d}$

$$D_d f(\boldsymbol{x}^s) = \nabla^T f(\boldsymbol{x}^s)\boldsymbol{d} = (\nabla f(\boldsymbol{x}^s), \boldsymbol{d}) = |\nabla f(\boldsymbol{x}^s)||\boldsymbol{d}|\cos \widehat{(\nabla f(\boldsymbol{x}^s), \boldsymbol{d})}$$

матиме найменше значення, рівне $-|\nabla f(\boldsymbol{x}^s)|$, якщо напрямок $\boldsymbol{d}$ збігатиметься з напрямком антиградієнта $-\nabla f(\boldsymbol{x}^s)$ ($\widehat{(\nabla f(\boldsymbol{x}^s), \boldsymbol{d})} = \pi$).

Отже швидкість спадання функції $f(\boldsymbol{x})$ у напрямі антиградієнта $-\nabla f(\boldsymbol{x}^s)$ є найбільшою.

Якщо за підхожий напрямок $\boldsymbol{d}$ *для мінімізації функції* $f(\boldsymbol{x})$ *використовується антиградієнт (для максимізації — градієнт), то відповідний метод називають градієнтним. Початкову точку* $\boldsymbol{x}^0$ *в градієнтному методі вибирають довільно, а* *всі інші послідовні наближення до точки мінімуму обчислюються за формулою*

$$\boldsymbol{x}^{s+1}=\boldsymbol{x}^{s}-\rho_s\nabla f(\boldsymbol{x}^{s}),\ s=1,2,\ldots \tag{10.3}$$

Такий перехід від точки $\boldsymbol{x}^s$ до точки $\boldsymbol{x}^{s+1}$ зменшує значення функції $f(\boldsymbol{x})$, якщо крок ρ_s досить малий. Розглянемо способи регулювання кроку ρ_s на довільній ітерації градієнтного методу.

Градієнтний метод з подрібненням кроку

Фіксується досить малий крок $\rho_0>0$ і, починаючи з точки $\boldsymbol{x}^0$, деяке число раз реалізується процедура (10.3) (див. рис. 10.2). На кожній ітерації обчислюється значення функції $f(\boldsymbol{x}^s)$. Процедуру (10.3) продовжують доти, поки $f(\boldsymbol{x}^s)$ зменшується. При цьому точка $\boldsymbol{x}^s$, як правило, прямує в окіл локального мінімуму,

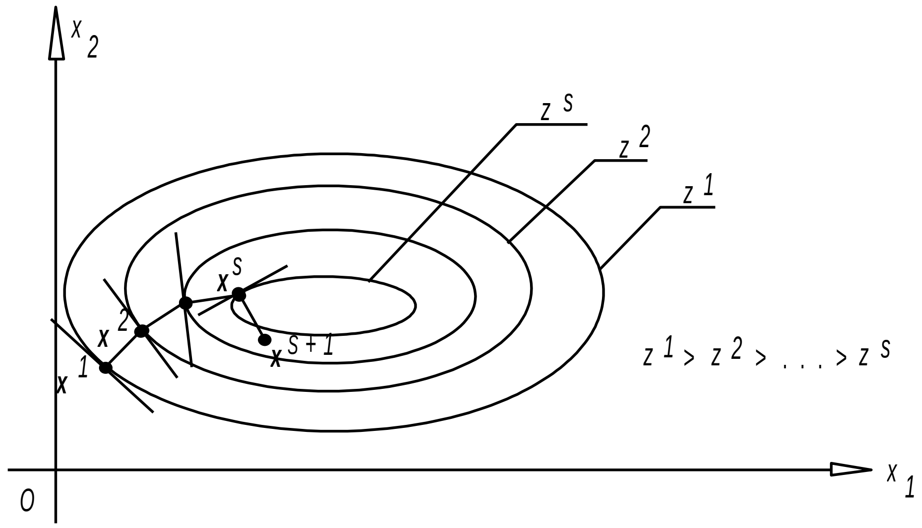


Рис. 10.2

розмір якого одного порядку з ρ_0. Точка $\boldsymbol{x}^s$ досягає цього околу при такому $\bar{s}$, при якому буде

$$f(\boldsymbol{x}^{\bar{s}+1})\geq f(\boldsymbol{x}^{\bar{s}}).$$

Якщо досягнута точність недостатня, то зменшують крок, тобто вибирають $0<\rho_1<\rho_0$, і продовжують ітерації з новим кроком за правилом (10.3) доти, поки точка $\boldsymbol{x}^s$ не попаде в окіл локального мінімуму, розмір якого не більший за задану похибку.

Теорема 10.1 (про збіжність градієнтного методу з подрібненням кроку). *Нехай*:

1) *функція* $f(\boldsymbol{x})$ *опукла і* $f(\boldsymbol{x})\in C^2$, $\boldsymbol{x}\in E^n$;
2) *множина* $R(\boldsymbol{x}^0)=\{\boldsymbol{x}\in E^n: f(\boldsymbol{x})\leq f(\boldsymbol{x}^0)\}$ *обмежена*;
3) *на множині* $R(\boldsymbol{x}^0)$ $\forall\eta$, $|\eta|=1$, *гессіан* $H_f(\boldsymbol{x})$ *задовольняє умову* $(H_f(\boldsymbol{x}))\eta,\eta)\leq M$ $(M>0)$.

Тоді:

якщо крок ρ *методу задовольняє умову* $0<\rho<2/M$, *то*

$$\lim_{s\to\infty} f(\boldsymbol{x}^s)=\min f(\boldsymbol{x}).$$

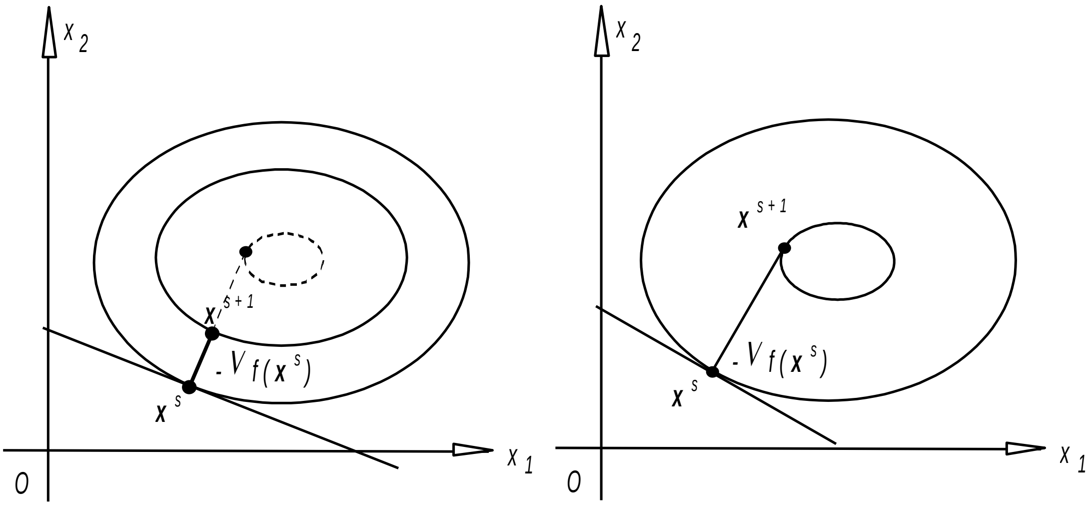
Без доведення (доведення див. [16], стор. 314).

Метод найшвидшого спуску

В цьому градієнтному методі величина кроку ρ_S в процедурі (10.3) вибирається за правилом

$$\rho_S = \arg\min_{\rho>0} f(\boldsymbol{x}^S - \rho\nabla f(\boldsymbol{x}^S)). \tag{10.4}$$

При фіксованому кроці ρ ми повинні зупинятися в точці $\boldsymbol{x}^{S+1}$ на кожній ітерації, незважаючи на те, що напрямок $-\nabla f(\boldsymbol{x}^S)$ ще веде до зменшення значення цільової функції.



крок ρ_S фіксований — крок ρ_S задовольняє (10.4)

Рис. 10.3

В методі найшвидшого спуску рух після точки $\boldsymbol{x}^{S+1}$ у напрямку антиградієнта $-\nabla f(\boldsymbol{x}^S)$ уже не приводить до менших значень цільової функції (див. Рис. 10.3). Тому метод найшвидшого спуску відносять до так званих *повнокрокових методів*.

Теорема 10.2 (про збіжність методу найшвидшого спуску).
Нехай:

1) *функція* $f(\boldsymbol{x})$ $(\boldsymbol{x}\in E^n)$ *неперервно диференційовна* $(f(\boldsymbol{x})\in C^1)$;
2) *мінімум* $f(\boldsymbol{x})$ *існує* $(\min f(\boldsymbol{x}) > -\infty)$;
3) *множина* $R(\boldsymbol{x}^0)=\{\boldsymbol{x}\in E^n\colon f(\boldsymbol{x}) \le f(\boldsymbol{x}^0)\}$ *обмежена*.

Тоді:

а) *якщо метод найшвидшого спуску закінчується за скінченне число ітерацій N, то* $\|\nabla f(\boldsymbol{x}^N)\| = 0$;
б) *якщо метод не закінчується за скінченне число ітерацій, то послідовність* $\{f(\boldsymbol{x}^S)\}$ *збігається, і для кожної граничної точки* $\boldsymbol{x}'$ *послідовності* $\{\boldsymbol{x}^S\}$ *виконується умова* $\|\nabla f(\boldsymbol{x}')\| = 0$.

Доведення. Можливі дві альтернативи: або метод найшвидшого спуску збігається за скінченне число ітерацій, або не збігається.

Розглянемо першу альтернативу. Нехай процедура (10.3) завершилась за скінченне число ітерацій в точці $\boldsymbol{x}^N$ виконанням умови (10.4). Це означає, що для довільних $\rho > 0$

$$f(\boldsymbol{x}^N - \rho\nabla f(\boldsymbol{x}^N)) \geq f(\boldsymbol{x}^N). \qquad (10.5)$$

За теоремою Тейлора для довільного $\rho > 0$ існує $\theta \in (0,1)$ таке, що

$$f(\boldsymbol{x}^N - \rho\nabla f(\boldsymbol{x}^N)) - f(\boldsymbol{x}^N) = -\rho\nabla^T f(\boldsymbol{x}^N - \rho\theta\nabla f(\boldsymbol{x}^N))\nabla f(\boldsymbol{x}^N).$$

Тоді за умови (10.5) для довільних $\rho > 0$ при $\theta \in (0,1)$ повинно бути

$$\nabla^T f(\boldsymbol{x}^N - \rho\nabla f(\boldsymbol{x}^N))\nabla f(\boldsymbol{x}^N) \leq 0. \qquad (10.6)$$

Оскільки $f(\boldsymbol{x}) \in C^1$, то $\nabla f(\boldsymbol{x})$ неперервна функція. Тому $\nabla f(\boldsymbol{x})$ зберігає знак у двох досить близьких точках, тобто при досить малих $\rho > 0$ градієнти

$$\nabla f(\boldsymbol{x}^N - \rho\nabla f(\boldsymbol{x}^N)) \text{ та } \nabla f(\boldsymbol{x}^N)$$

однакові за знаком.

Отже, при досить малих $\rho > 0$ поряд з (10.6) матиме місце умова

$$\nabla^T f(\boldsymbol{x}^N)\nabla f(\boldsymbol{x}^N) \leq 0, \qquad (10.7)$$

тобто

$$\|\nabla f(\boldsymbol{x}^N)\|^2 \leq 0,$$

що може бути лише у випадку

$$\|\nabla f(\boldsymbol{x}^N)\| = 0.$$

Розглянемо тепер іншу альтернативу.

За умови 2) теореми послідовність $\{f(\boldsymbol{x}^S)\}$ обмежена знизу. За побудовою послідовність $\{f(\boldsymbol{x}^S)\}$ спадаюча, оскільки $f(\boldsymbol{x}^{S+1}) < f(\boldsymbol{x}^S)$. Тому вона є збіжною. Нехай

$$\lim_{S\to\infty} f(\boldsymbol{x}^S) = f(\boldsymbol{x}').$$

Розглянемо послідовність $\{\boldsymbol{x}^S\}$. Оскільки $\{f(\boldsymbol{x}^S)\}$ збігається, то з $\{\boldsymbol{x}^S\}$ можна виділити збіжну, наприклад, до точки $\boldsymbol{x}'$, підпослідовність $\{\boldsymbol{x}^{S_k}\}$

$$\lim_{k\to\infty} \boldsymbol{x}^{S_k} = \boldsymbol{x}'.$$

Очевидно, що при цьому буде

$$\lim_{S\to\infty} f(\boldsymbol{x}^S) = \lim_{k\to\infty} f(\boldsymbol{x}^{S_k}) = f(\boldsymbol{x}').$$

Від супротивного припустимо, що $\|\nabla f(\boldsymbol{x}')\| > 0$. Розглянемо тепер підпослідовність $\{\boldsymbol{x}^{S_k+1}\}$. Ця підпослідовність обмежена, оскільки вона за побудовою належить множині $R(\boldsymbol{x}^0)$, яка за умовою 3) теореми обмежена. Тоді, в свою чергу, з $\{\boldsymbol{x}^{S_k+1}\}$ можна вибрати збіжну, скажімо, до $\boldsymbol{x}''$ підпослідовність. Оскільки

$$\boldsymbol{x}^{S_k+1} = \boldsymbol{x}^{S_k} - \rho_{S_k}\nabla f(\boldsymbol{x}^{S_k}),$$

то, спрямувавши $k \to \infty$, отримаємо $\boldsymbol{x}'' = \boldsymbol{x}' - \rho'\nabla f(\boldsymbol{x}')$, де

$$\rho' = \arg\min_{\rho>0} f(\boldsymbol{x}' - \rho\nabla f(\boldsymbol{x}')).$$

Так як за припущенням $\|\nabla f(\boldsymbol{x}')\| > 0$, то $\nabla f(\boldsymbol{x}'') \neq 0$, а, отже, $f(\boldsymbol{x}'') < f(\boldsymbol{x}')$, що суперечить умові про збіжність $\{f(\boldsymbol{x}^S)\}$ до $f(\boldsymbol{x}')$.

Теорема доведена.

Тести на зупинку процедури найшвидшого спуску

Оскільки збіжність методу найшвидшого спуску в загальному випадку не буде скінченною, то необхідно визначити ознаку припинення процесу ітерацій. Наведемо кілька найбільш уживаних критеріїв:

1) $\max\limits_{i=1,\ldots,n} \left| \dfrac{\partial f}{\partial x_i} \right| < \varepsilon \quad (\varepsilon > 0 \text{ задане});$

2) $\|\nabla f(x)\|^2 = \sum\limits_{i=1}^{n} \left(\dfrac{\partial f}{\partial x_i} \right)^2 < \varepsilon \quad (\varepsilon > 0 \text{ задане});$

3) $\left| f(\boldsymbol{x}^{s+1}) - f(\boldsymbol{x}^s) \right| < \varepsilon \quad (\varepsilon > 0 \text{ задане}).$

Недоліки градієнтних методів

Мабуть головним недоліком градієнтного методу є те, що він у найкращому випадку забезпечує збіжність послідовності $\{\boldsymbol{x}^S\}$ лише до точки локального мінімуму функції $f(\boldsymbol{x})$.

В загальному випадку послідовність $\{\boldsymbol{x}^S\}$ збігається, як було доведено, до стаціонарної точки $\boldsymbol{x}'$функції $f(\boldsymbol{x})$, в якій $\nabla f(\boldsymbol{x}') = 0$.

Повну гарантію збіжності $\{\boldsymbol{x}^S\}$ до точки глобального мінімуму функції $f(\boldsymbol{x})$ може дати, наприклад, вимога опуклості функції $f(\boldsymbol{x})$.

Одним із суттєвих недоліків методу найшвидшого спуску є те, що для деяких типів функцій збіжність його може виявитись повільною. Дійсно, якщо число ρ_S мінімізує функцію $g(\rho) = f(\boldsymbol{x}^S) - \rho\nabla f(\boldsymbol{x}^S)$, то повинно бути

$$\frac{dg(\rho_S)}{d\rho} = (\nabla f(x^{s+1}), \nabla f(x^s)) = 0,$$

що означає ортогональність напрямків зміщення на послідовних ітераціях (див. рис. 10.4). Тоді число ітерацій, необхідних для мінімізації погано обумовлених функцій "яристого" типу може бути дуже великим (хоча існують способи покращення збіжності, див., наприклад, [17], стор. 39).

Недоліком градієнтних методів є також те, що неможливе їх безпосереднє застосування до оптимізації недиференційовних функцій або умовної оптимізації диференційовних функцій.

<u>*Зауваження*</u>. Задача відшукання кроку на кожній ітерації методу найшвидшого спуску є задачею одновимірної оптимізації і може бути розв'язана одним із розглянутих раніше методів одновимірної оптимізації.

Приклад 10.1. *Визначити за допомогою градієнтного методу мінімум функції*

$$f(\boldsymbol{x}) = -4x_1 - 2x_2 + x_1^2 + x_2^2,$$

почавши ітераційний процес з точки $\boldsymbol{x}^0 = (4, 5)$.

Знаходимо частинні похідні

$$\frac{\partial f}{\partial x_1} = -4 + 2x_1, \frac{\partial f}{\partial x_2} = -2 + 2x_2.$$

1-а ітерація. Обчислюємо градієнт функції $f(\boldsymbol{x})$ в початковій точці

$$\nabla f(\boldsymbol{x}^0) = (4,8).$$

Він відмінний від нуля, тому будуємо промінь $\boldsymbol{x}^I(\rho)$, що виходить з точки $\boldsymbol{x}^0$ у напрямку антиградієнта

$$\boldsymbol{x}^I(\rho) = \boldsymbol{x}^0 - \rho\nabla f(\boldsymbol{x}^0) = (4,5) - \rho(4,8) = (4-4\rho, 5-8\rho), \ \rho > 0.$$

Обчислюємо градієнт функції

$$\nabla f(\boldsymbol{x}^I(\rho)) = (-4+2(4-4\rho), -2+2(5-8\rho)) = (4-8\rho, 8-16\rho)$$

та розглядаємо функцію $f(\boldsymbol{x})$ в точках променя $\boldsymbol{x}^I(\rho)$ при $\rho > 0$. Функція $f(\boldsymbol{x})$ в цьому випадку буде функцією однієї змінної ρ. Оскільки

$$\frac{df(\boldsymbol{x}^S - \rho\nabla f(\boldsymbol{x}^S))}{d\rho} = -(\nabla f(\boldsymbol{x}^S - \rho\nabla f(\boldsymbol{x}^S)), \nabla f(\boldsymbol{x}^S)),$$

то в силу необхідних умов екстремуму

$$\frac{df(\boldsymbol{x}^I(\rho))}{d\rho} = -(\nabla f(\boldsymbol{x}^I(\rho)), \nabla f(\boldsymbol{x}^0)) = 0,$$

тобто $-((4-8\rho, 8-16\rho), (4,8)) = 0$, звідки $160\rho - 80 = 0$ або $\rho_0 = 0.5$. Оскільки

$$\frac{d^2 f(\boldsymbol{x}^I(\rho))}{d\rho^2} = \frac{d}{dx}(160\rho - 80) = 160 > 0,$$

то $\rho_0 = 0.5$ є точкою мінімуму $f(\boldsymbol{x}^I(\rho))$. Обчислюємо $\boldsymbol{x}^1$

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 - \rho_0\nabla f(\boldsymbol{x}^0) = (4,5) - 0.5(4,8) = (2,1).$$

2-а ітерація. Обчислюємо градієнт функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^1$.

$$\nabla f(\boldsymbol{x}^1) = (-4+2\times 2, -2+2\times 1) = (0,0).$$

Оскільки він рівний нулю, то точка $\boldsymbol{x}^1$ є стаціонарною точкою функції $f(\boldsymbol{x})$. До того ж $f(\boldsymbol{x})$ опукла вниз, тому $\boldsymbol{x}^1$ буде точкою глобального мінімуму $f(\boldsymbol{x})$.

Приклад 10.2. *Визначити за допомогою градієнтного методу мінімум функції* $f(\boldsymbol{x}) = (x_1-3)^2 + 4(x_2-2)^2$, *почавши ітераційний процес з точки* $\boldsymbol{x}^0 = (0,0)$.

Результати розв'язування задачі методом найшвидшого спуску приведені в таблиці 10.1 та ілюструється рисунком 10.4. Зауважимо, що розрахунки ведуться за формулами

$$\nabla f(\boldsymbol{x}) = (2x_1-6, 8x_2-16),$$

$$\boldsymbol{x}^{S+1} = \boldsymbol{x}^S - \rho_S\nabla f(\boldsymbol{x}^S), \ s = 1,2,\ldots,$$

$$\rho_S = \arg\min_{\rho>0} f(\boldsymbol{x}^S - \rho\nabla f(\boldsymbol{x}^S)).$$

Таблиця 10.1

s	$\boldsymbol{x}^s$	$f(\boldsymbol{x}^s)$	$\nabla f(\boldsymbol{x}^s)$	ρ_s	$\|\nabla f(\boldsymbol{x}^s)\|$
0	(0,0)	25	(□6,□16)	0.138	17.088
1	(0.826,2.204)	4.893	(□4.348, 1.632)	0.365	4.644
2	(2.412,1.609)	0.957	(□1.176,□3.128)	0.138	3.342
3	(2.574,2.041)	0.188	(□0.852, 0.328)		0.913
∞	(3,2)	0	(0,0)		0

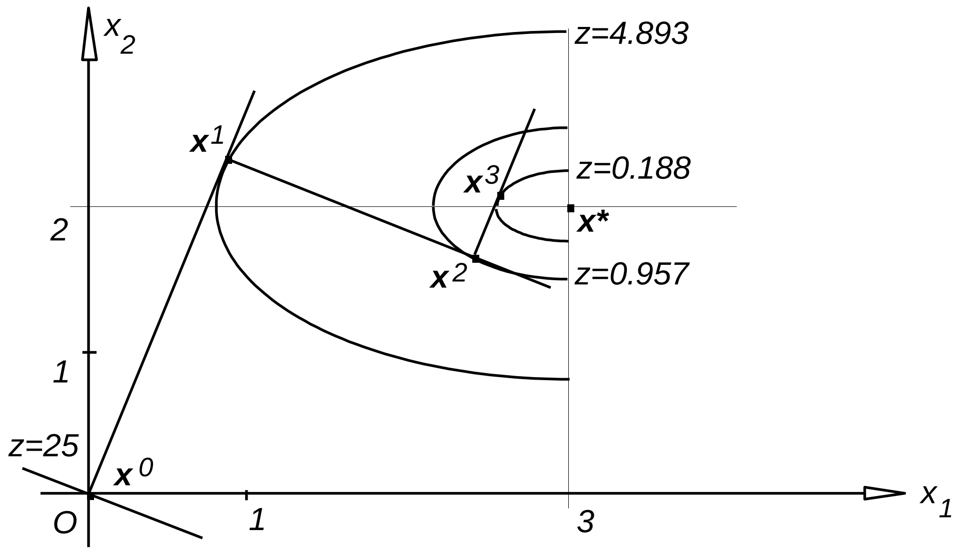


Рис. 10.4

Лекція 33. *Метод можливих напрямків Зойтендейка.– 2год. [1].*

Завдання для самостійної роботи. *Активні обмеження. Можливі напрямки. Допоміжна задача для відшукання можливих і підхожих напрямків. Обчислення кроку методу. Вибір початкового наближення. 4год. [1-3,7,9].*

Методи можливих напрямків

Метод Зойтендейка

Розглянемо задачу *НЛП*

$$\min\{f^0(\boldsymbol{x}):\ f^i(\boldsymbol{x})\le 0,\ i=1,\dots,m,\ \boldsymbol{x}\in E^n\} \qquad (11.1)$$

за умови, що функції $f^i(\boldsymbol{x})$ диференційовні, $f^i(\boldsymbol{x})\in C^1$, $i=0,1,\dots,m$.

Зауважимо, що до вигляду (11.1) легко зводиться і *ЗНЛП* з умовами невід'ємності змінних $x_j\ge 0$, $j=1,\dots,n$. Для цього достатньо включити в загальну систему обмежень $f^i(\boldsymbol{x})\le 0$, $i=1,\dots,m$, умови $-x_j\le 0$, $j=1,\dots,n$.

Нехай точка $\boldsymbol{x}^s\in X$, де $X=\{\boldsymbol{x}\in E^n:\ f^i(\boldsymbol{x})\le 0,\ i=1,\dots,m\}$. Обмеження $f^i(\boldsymbol{x})\le 0$ назвемо *активним* в точці $\boldsymbol{x}^s$, якщо $f^i(\boldsymbol{x}^s)=0$. Позначимо через $I_s=\{i: f^i(\boldsymbol{x}^s)=0\}$ множину індексів активних обмежень в точці $\boldsymbol{x}^s$. Очевидно, що тільки ці обмеження визначають напрямки просування із допустимої точки $\boldsymbol{x}^s$ в іншу допустиму точку $\boldsymbol{x}^{s+1}$.

Будемо, поки-що, вважати $\boldsymbol{x}^{s+1}$ довільною точкою простору E^n і подамо її у вигляді $\boldsymbol{x}^{s+1}=\boldsymbol{x}^s+\rho\boldsymbol{r}^s$, де $\boldsymbol{r}^s$ — довільний вектор напрямку зміщення із точки $\boldsymbol{x}^s$, а $\rho>0$ — крок зміщення. Напрямок $\boldsymbol{r}^s$ назвемо *можливим напрямком*, якщо існує $\rho>0$ таке, що точка $\boldsymbol{x}^{s+1}$ задовольняє умову:

$$\boldsymbol{x}^{s+1}\in X \text{ або } f^i(\boldsymbol{x}^{s+1})\le 0,\ i=1,\dots,m. \qquad (11.2)$$

Оскільки $\boldsymbol{x}^s$ — допустима точка, тобто $f^i(\boldsymbol{x}^s)\le 0$, $i=1,\dots,m$, то умови (11.2) еквівалентні умовам

$$f^i(\boldsymbol{x}^{s+1})\le f^i(\boldsymbol{x}^s),\ i=1,\dots,m. \qquad (11.3)$$

Далі, оскільки тільки активні у точці $\boldsymbol{x}^s$ обмеження впливають на вибір можливого напрямку, то в умовах (11.3) слід вважати, що $i\in I_s$.

Будемо вимагати, щоб $\boldsymbol{r}^s$ був напрямком *можливим*, тоді він повинен визначатися умовами

$$f^i(\boldsymbol{x}^{s+1})\le f^i(\boldsymbol{x}^s),\ i\in I_s, \qquad (11.4)$$

і *підхожим*, тоді він повинен задовольняти нерівність

$$f^0(\boldsymbol{x}^{s+1})< f^0(\boldsymbol{x}^s) \qquad (11.5)$$

при досить малих $\rho>0$.

Як було доведено вище (див. Розділ 10), умова (11.5) буде виконуватись при досить малих $\rho>0$, тобто напрямок $\boldsymbol{r}^s$ буде підхожим, якщо похідна по напрямку $\boldsymbol{r}^s$ від функції $f^0(\boldsymbol{x})$ в точці $\boldsymbol{x}^s$ буде від'ємною:

$$D_{r^s} f^0(\boldsymbol{x}^s) = (\nabla f^0(\boldsymbol{x}^s), \boldsymbol{r}^s) < 0 . \qquad (11.6)$$

Аналогічно приходимо до висновку, що умови (11.4) будуть виконуватись при досить малих $\rho>0$ тільки для тих напрямків $\boldsymbol{r}^s$, для яких похідні $D_{r^s} f^i(\boldsymbol{x}^s)$, $i \in I_s$, недодатні:

$$D_{r^s} f^i(\boldsymbol{x}^s) = (\nabla f^i(\boldsymbol{x}^s), \boldsymbol{r}^s) \le 0 , \; i \in I_s . \qquad (11.7)$$

Для обмеження довжин векторів $\boldsymbol{r}^s$ до системи умов (11.6)(11.7), яка визначає всі можливі і підхожі напрямки, додають, як правило, яку-небудь умову нормування, наприклад, таку:

$$-1 \le \boldsymbol{r}_j^s \le 1 , \; j=1,...,n. \qquad (11.8)$$

Остаточно, для відшукання можливого і підхожого напрямку $\boldsymbol{r}^s$ отримуємо *задачу лінійного програмування*

$$D_{r^s} f^0(\boldsymbol{x}^s) = (\nabla f^0(\boldsymbol{x}^s), \boldsymbol{r}^s) \to \min,$$
$$D_{r^s} f^i(\boldsymbol{x}^s) = (\nabla f^i(\boldsymbol{x}^s), \boldsymbol{r}^s) \le 0, \; i \in I_s, \qquad (11.9)$$
$$-1 \le r_j^s \le 1, \; j=1,...,n.$$

Якщо оптимальне значення цієї задачі невід'ємне, то $\boldsymbol{x}^s$ - стаціонарна точка функції $f^0(\boldsymbol{x})$ за умов $f^i(\boldsymbol{x}) \le 0$, $i=1,...,m$, $\boldsymbol{x} \in E^n$; в іншому випадку вектор $\boldsymbol{r}^s$ визначає можливий і підхожий напрямок. Тоді у знайденому напрямку $\boldsymbol{r}^s$ будуємо промінь $\boldsymbol{x}(\rho) = \boldsymbol{x}^s + \rho \boldsymbol{r}^s$ $(\rho>0)$ і, підставляючи $\boldsymbol{x}(\rho)$ у всі неактивні обмеження, знаходимо число ε, яке обмежує крок ρ зверху: $0 < \rho \le \varepsilon$. Конкретне значення ρ_s визначаємо як

$$\rho_s = \arg\min_{\rho \in (0,\varepsilon]} f^0(\boldsymbol{x}^s + \rho \boldsymbol{r}^s)$$

за допомогою якої-небудь процедури одновимірної оптимізації.

Зауважимо, що, як і для градієнтних методів, метод можливих напрямків не гарантує нічого більшого, ніж збіжність $\boldsymbol{x}^s$ до стаціонарної точки функції $f^0(\boldsymbol{x})$.

До вибору початкового наближення $\boldsymbol{x}^0$. За $\boldsymbol{x}^0$ можна взяти довільну допустиму точку $\boldsymbol{x} \in X$. Якщо обмеження задачі (11.1) лінійні, то за $\boldsymbol{x}^0$ можна взяти довільний базисний розв'язок системи обмежень задачі (11.1).

Приклад 11.1. *Розв'язати методом можливих напрямків задачу*

$$f^0(x_1,x_2) = x_1^2 - 2x_1 + 1 + x_2^2 \to \max,$$
$$f^1(x_1,x_2) = x_1^2 + x_2 - 1 \le 0,$$
$$x_1 \ge 0, \; x_2 \ge 0,$$

вибираючи $\boldsymbol{x}^0 = (1/2, 0)$.

Перетворюємо задачу до потрібного вигляду:

$$f^0(x_1,x_2) = -(x_1-1)^2 - x_2^2 \to \min,$$
$$f^1(x_1,x_2) = x_1^2 + x_2 - 1 \le 0,$$
$$f^2(x_1,x_2) = -x_1 \le 0,$$
$$f^3(x_1,x_2) = -x_2 \le 0.$$

Обчислюємо градієнти:

$$\nabla f^0(x_1,x_2) = (-2x_1+2, -2x_2),$$
$$\nabla f^1(x_1,x_2) = (2x_1, 1),$$
$$\nabla f^2(x_1,x_2) = (-1, 0),$$
$$\nabla f^3(x_1,x_2) = (0, -1).$$

1-а ітерація. Визначаємо активні обмеження в точці $\boldsymbol{x}^0$:

$$f^1(1/2, 0) = -3/4 \ne 0,$$
$$f^2(1/2, 0) = -1/2 \ne 0,$$
$$f^3(1/2, 0) = 0. \quad \text{(Активне обмеження)}$$

Нехай $\boldsymbol{r}^0 = (r_1^0, r_2^0)$ □ вектор невідомого можливого і підхожого напрямку. Обчислюємо необхідні значення параметрів

$$\nabla f^0(1/2,0) = (1,0), \ \nabla f^3(1/2,0) = (0,-1),$$
$$D_{\boldsymbol{r}^0} f^0(\boldsymbol{x}^0) = (\nabla f^0(1/2,0), \boldsymbol{r}^0) = ((1,0), (r_1^0, r_2^0)) = r_1^0,$$
$$D_{\boldsymbol{r}^0} f^3(\boldsymbol{x}^0) = (\nabla f^3(1/2,0), \boldsymbol{r}^0) = ((0,-1), (r_1^0, r_2^0)) = -r_2^0,$$

та записуємо допоміжну задачу *ЛП* для визначення $\boldsymbol{r}$

$$L = D_{\boldsymbol{r}^0} f^0(\boldsymbol{x}^0) = r_1^0 \to \min,$$
$$D_{\boldsymbol{r}^0} f^3(\boldsymbol{x}^0) = -r_2^0 \le 0, \qquad (11.10)$$
$$-1 \le r_1^0 \le 1,$$
$$-1 \le r_2^0 \le 1.$$

Задача (11.10) легко розв'язується графічно, вона має альтернативний оптимум: $\boldsymbol{r}^* = (-1, r_2^*)$, $0 \le r_2^* \le 1$, $L^* = -1$. Оскільки її оптимальне значення рівне -1, то можливі і підхожі напрямки існують і задаються векторами $\boldsymbol{r}^* = (-1, r_2^*)$, $0 \le r_2^* \le 1$. За $\boldsymbol{r}^0$ приймаємо той з них, який має найбільшу довжину: $\boldsymbol{r}^0 = (-1,1)$.

Далі, будуємо промінь $\boldsymbol{x}(\rho)$, що виходить із точки $\boldsymbol{x}^0$ в напрямку $\boldsymbol{r}^0$

$$\boldsymbol{x}(\rho) = \boldsymbol{x}^0 + \rho \boldsymbol{r}^0 = (1/2, 0) + \rho(-1,1) = (1/2-\rho, \rho), \ \rho > 0.$$

За допомогою неактивних обмежень визначаємо обмеження для ρ зверху.

$$\begin{cases} f^1(1/2-\rho, \rho) = (1/2-\rho)^2 + \rho - 1 \le 0, \\ f^2(1/2-\rho, \rho) = \rho - 1/2 \le 0, \\ \rho > 0. \end{cases}$$

Розв'язавши цю систему, отримаємо: $0<\rho\le 1/2$. Отже, при $0<\rho\le 1/2$ точка $\boldsymbol{x}(\rho)$ залишиться в межах допустимої області.

Знаходимо

$$\min_{0<\rho\le 1/2} f^0(\boldsymbol{x}(\rho)) = \min_{0<\rho\le 1/2} f^0(1/2-\rho,\rho).$$

Маємо

$$\frac{d}{d\rho}f^0(\boldsymbol{x}(\rho)) = \frac{d}{d\rho}f^0(\boldsymbol{x}^0+\rho\boldsymbol{r}^0) = (\nabla f^0(\boldsymbol{x}^0+\rho\boldsymbol{r}^0),\boldsymbol{r}^0) =$$

$$= (\nabla f^0(1/2-\rho,\rho),(-1,1)) = ((-2(1/2-\rho)+2,-2\rho),(-1,1)) = -4\rho-1=0.$$

Звідси отримуємо: $\rho=-1/4<0$. Крім того маємо

$$\frac{d^2}{d\rho^2}f^0(\boldsymbol{x}(\rho)) = -4<0.$$

Отже, $\rho=-1/4$ – точка максимуму функції $g(\rho)=f^0(\boldsymbol{x}(\rho))=f^0(\boldsymbol{x}^0+\rho\boldsymbol{r}^0)$, яка є опуклою доверху. Тому мінімальне значення $f^0(\boldsymbol{x}^0+\rho\boldsymbol{r}^0)$ досягається на правому кінці відрізка $(0,1/2]$, тобто при $\rho_0 = 1/2$.

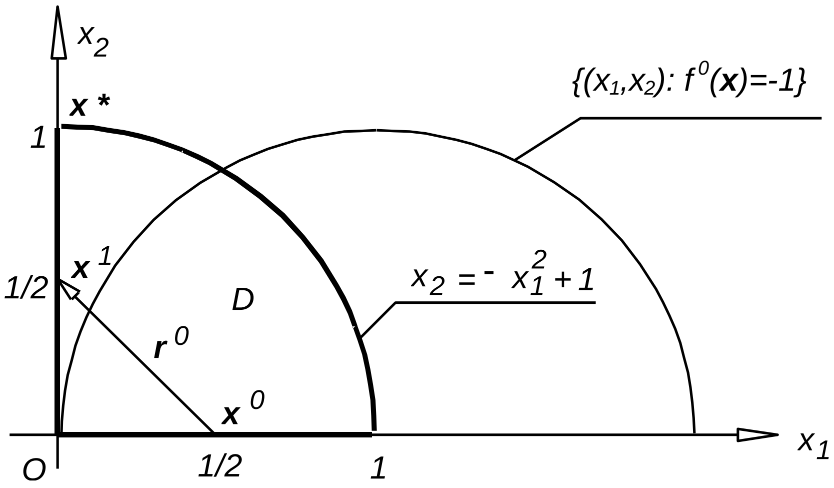


Рис. 11.1

Остаточно отримаємо (див. рис. 11.1)

$$\boldsymbol{x}^1 = \boldsymbol{x}^0+\rho_0\boldsymbol{r}^0 = (1/2,0)+1/2\ (-1,1) = (0,1/2).$$

Переходимо до наступної ітерації.

<u>*2-а ітерація*</u>. Визначаємо активні обмеження в точці $\boldsymbol{x}^1$:

$f^1(0,1/2) = -1 \neq 0$,

$f^2(0,1/2) = 0$, (Активне обмеження)

$f^3(0,1/2) = -1/2 \neq 0$.

Будуємо допоміжну оптимізаційну задачу для визначення можливого і підхожого напрямку $\boldsymbol{r}^1=(r_1^1,r_2^1)$:

$\nabla f^0(0,1/2) = (2,-1)$, $\nabla f^2(0,1/2) = (-1,0)$,

$D_{\boldsymbol{r}^1} f^0(\boldsymbol{x}^1) = (\nabla f^0(0,1/2),\boldsymbol{r}^1) = ((2,-1),(r_1^1,r_2^1)) = 2r_1^1 - r_2^1$,

$$D_{\boldsymbol{r}^1} f^2(\boldsymbol{x}^1) = (\nabla f^2(0,1/2), \boldsymbol{r}^1) = ((-1,0), (r_1^1, r_2^1)) = -r_1^1,$$

$$L = D_{\boldsymbol{r}^1} f^0(\boldsymbol{x}^1) = 2r_1^1 - r_2^1 \to \min,$$

$$D_{\boldsymbol{r}^1} f^2(\boldsymbol{x}^1) = -r_1^1 \le 0,$$

$$-1 \le r_1^1 \le 1,$$

$$-1 \le r_2^1 \le 1.$$

Задача має єдиний розв’язок: $\boldsymbol{r}^* = (0, 1)$, $L^* = -1$. Її оптимальне значення рівне -1, тому вектор $\boldsymbol{r}^1 = (0, 1)$ є вектором можливого і підхожого напрямку.

Будуємо промінь $\boldsymbol{x}(\rho)$, що виходить із точки $\boldsymbol{x}^1$ в напрямку $\boldsymbol{r}^1$

$$\boldsymbol{x}(\rho) = \boldsymbol{x}^1 + \rho \boldsymbol{r}^1 = (0,1/2) + \rho(0,1) = (0, \rho + 1/2), \ \rho > 0.$$

Визначаємо інтервал можливих значень для ρ, підставляючи $\boldsymbol{x}(\rho)$ у неактивні обмеження:

$$\begin{cases} f^1(0, \rho + 1/2) = \rho + 1/2 - 1 \le 0, \\ f^3(0, \rho + 1/2) = -(\rho + 1/2) \le 0, \\ \rho > 0, \end{cases}$$

Розв'язавши цю систему, отримаємо: $0 < \rho \le 1/2$. Отже, при $0 < \rho \le 1/2$ точка $\boldsymbol{x}(\rho)$ залишиться в межах допустимої області.

Знаходимо

$$\min_{0<\rho\le 1/2} f^0(\boldsymbol{x}(\rho)) = \min_{0<\rho\le 1/2} f^0(\rho, \rho + 1/2).$$

Покладемо $g(\rho) = f^0(0, \rho + 1/2) = -1 - (\rho + 1/2)^2$ та використаємо необхідні і достатні умови екстремуму для функції $g(\rho)$. Маємо: $g'(\rho) = -2\rho - 1 = 0$, звідки $\rho = -1/2 < 0$. Оскільки $g''(\rho) = -2 < 0$, то $\rho = -1/2$ □ точка максимуму функції $g(\rho) = f^0(\boldsymbol{x}(\rho)) = f^0(\boldsymbol{x}^1 + \rho \boldsymbol{r}^1)$, яка є опуклою доверху. Тому мінімальне значення $f^0(\boldsymbol{x}^1 + \rho \boldsymbol{r}^1)$ досягається на правому кінці відрізка $(0, 1/2]$, тобто при $\rho_1 = 1/2$.

Обчислюємо точку $\boldsymbol{x}^2$:

$$\boldsymbol{x}^2 = \boldsymbol{x}^1 + \rho_1 \boldsymbol{r}^1 = (0,1/2) + 1/2\ (0,1) = (0,1).$$

Переходимо до наступної ітерації.

<u>*3-а ітерація*</u>. Визначаємо активні обмеження в точці $\boldsymbol{x}^2$:

$f^1(0,1) = 0$, (Активне обмеження)

$f^2(0,1) = 0$, (Активне обмеження)

$f^3(0,1) = -1 \ne 0$.

Будуємо допоміжну оптимізаційну задачу для визначення можливого і підхожого напрямку $\boldsymbol{r}^2 = (r_1^2, r_2^2)$:

$$\nabla f^0(0,1) = (2,-2), \ \nabla f^1(0,1) = (0,1), \ \nabla f^2(0,1) = (-1,0),$$

$$D_{\boldsymbol{r}^2} f^0(\boldsymbol{x}^2) = (\nabla f^0(0,1), \boldsymbol{r}^2) = ((2,-2), (r_1^2, r_2^2)) = 2r_1^2 - 2r_2^2,$$

$$D_{\boldsymbol{r}^2} f^1(\boldsymbol{x}^2) = (\nabla f^1(0,1), \boldsymbol{r}^2) = ((0,1), (r_1^2, r_2^2)) = r_2^2,$$

$D_{\boldsymbol{r}^2} f^2(\boldsymbol{x}^2) = (\nabla f^2(0,1), \boldsymbol{r}^2) = ((-1,0), (r_1^2, r_2)) = -r_1^2$,

$L = D_{\boldsymbol{r}^2} f^0(\boldsymbol{x}^2) = 2\, r_1^2 - 2\, r_2^2 \to \min$,

$D_{\boldsymbol{r}^2} f^1(\boldsymbol{x}^2) = r_2^2 \le 0$,

$D_{\boldsymbol{r}^2} f^2(\boldsymbol{x}^2) = -r_1^2 \le 0$,

$-1 \le r_1^2 \le 1$,

$-1 \le r_2^2 \le 1$.

Задача має єдиний розв'язок: $\boldsymbol{r}^* = (0, 0)$, $L^* = 0$. Оскільки її оптимальне значення рівне 0, то можливих і підхожих напрямків в точці $\boldsymbol{x}^2$ не існує, а, отже, точка $\boldsymbol{x}^2 = \boldsymbol{x}^* = (0,1)$ і буде оптимальним розв'язком вихідної задачі. Значення $f^0(\boldsymbol{x}^*)$ рівне $f^0(0,1) = -2$..

Зауважимо, що можливий і підхожий напрямок r^s, знайдений методом Зойтендейка, не завжди співпадає з напрямком антиградієнта $-\nabla f^0(x^s)$, коли той є можливим. Це може привести до значного збільшення кількості ітерацій, а, значить, і часу розв'язування задачі. Тому для покращення збіжності методу можливих напрямків можна комбінувати його з методом найшвидшого спуску по антиградієнту тоді, коли той є можливим напрямком. Крім того, у випадку, коли на деякій ітерації при визначенні вектора можливого і підхожого напрямку r^s активними є тільки лінійні обмеження, то *ЗЛП* для визначення r^s будують за іншою, більш ефективною схемою.

Метод можливих напрямків для ЗНП з лінійними обмеженнями

Розглянемо алгоритм комбінованого методу на прикладі задачі з нелінійною цільовою функцією і лінійними обмеженнями

$$f(\boldsymbol{x}) \to \min,$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, i = 1,\dots,m,$$

$$x_j \ge 0, j = 1,\dots,n.$$

Нехай $f(\boldsymbol{x}) \in C^1$. Розглянемо $s+1$-у ітерацію, вважаючи, що

$$\boldsymbol{x}^s \in D = \left\{ \boldsymbol{x} \in E^n : \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = 1,\dots,m,\ x_j \ge 0,\ j = 1,\dots,n \right\}.$$

1. Обчислюємо градієнт $\nabla f(\boldsymbol{x}^s)$ і перевіряємо умову $\nabla f(\boldsymbol{x}^s) = 0$. Якщо умова виконується, то $\boldsymbol{x}^s$ □ стаціонарна точка і кінець обчислень. Якщо умова не виконується, то переходимо до наступного пункту.

2. Підставляємо $\boldsymbol{x}^s$ в усі обмеження задачі і формуємо множини індексів I_s та J_s активних обмежень в точці $\boldsymbol{x}^s$

$$I_S = \left\{ i : i = 1,\dots,m, \sum_{j=1}^{n} a_{ij} x_j = b_i \right\},$$

$$J_S = \{ j : j = 1,\dots,n,\ x_j^S = 0 \}.$$

Переходимо до наступного пункту.

3. Перевіряємо, чи буде антиградієнт $-\nabla f(\boldsymbol{x}^S)$ можливим напрямком. Для цього будуємо промінь $\boldsymbol{x}(\rho) = \boldsymbol{x}^S - \rho \nabla f(\boldsymbol{x}^S)$, $\rho > 0$, який виходить із точки $\boldsymbol{x}^S$ в напрямі антиградієнта $-\nabla f(\boldsymbol{x}^S)$ і підставляємо $\boldsymbol{x}(\rho)$ в активні обмеження. Отримуємо систему лінійних нерівностей відносно ρ. Розв'язуємо її. Якщо $\rho > 0$, то антиградієнт $-\nabla f(\boldsymbol{x}^S)$ є можливим напрямком, переходимо до наступного пункту. Якщо $\rho \le 0$, то антиградієнт $-\nabla f(\boldsymbol{x}^S)$ не є можливим напрямком, переходимо до визначення можливих і підхожих напрямків, тобто до пункту 7.

4. Визначаємо множину значень кроку ρ, для яких точка променя $\boldsymbol{x}(\rho)$ залишається допустимою. З цією метою підставляємо $\boldsymbol{x}(\rho) = \boldsymbol{x}^S - \rho \nabla f(\boldsymbol{x}^S)$ в усі неактивні обмеження. Отримуємо систему лінійних нерівностей відносно ρ. Вона визначає шуканий інтервал $(0, \varepsilon]$ для кроку ρ. Переходимо до наступного пункту.

5. Знаходимо

$$\rho_S = \arg\min_{\rho \in (0,\varepsilon]} f(\boldsymbol{x}^S - \rho \nabla f(\boldsymbol{x}^S)).$$

Зауважимо, що процедура відшукання ρ_S є процедурою одновимірної оптимізації. Вона реалізується або за допомогою необхідних і достатніх умов екстремуму функції однієї змінної, або за допомогою чисельних методів типу дихотомії, золотого перерізу, Фібоначчі. Переходимо до наступного пункту.

6. Обчислюємо

$$\boldsymbol{x}^{S+1} = \boldsymbol{x}^S - \rho_S \nabla f(\boldsymbol{x}^S).$$

Переходимо до пункту 1.

7. Будуємо задачу *ЛП* для відшукання підхожих і можливих напрямків. Нехай $\boldsymbol{r}^S = (r_1^S, \dots, r_n^S)$ поки-що невідомий вектор можливого напрямку. Будуємо промінь $\boldsymbol{x}'(\rho)$, який виходить із точки $\boldsymbol{x}^S$ в напрямку $\boldsymbol{r}^S$:

$$\boldsymbol{x}'(\rho) = \boldsymbol{x}^S - \rho \boldsymbol{r}^S,\ \rho > 0.$$

Підставляємо $\boldsymbol{x}'(\rho)$ в активні обмеження. Отримуємо систему

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j^S + \rho \sum_{j=1}^{n} a_{ij} r_j^S \le b_i, & i \in I_S, \\ x_j^S + \rho\, r_j^S, & j \in J_S. \end{cases}$$

Звідки, враховуючи умови

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j^S = b_i, & i \in I_S, \\ x_j^S = 0, & j \in J_S, \end{cases}$$

та $\rho > 0$ маємо

$$\begin{cases} \sum_{j=1}^{n} a_{ij} r_j^s \le 0, & i \in I_s, \\ r_j^s \ge 0, & j \in J_s. \end{cases} \quad (11.11)$$

До останньої системи додаємо умову нормування, яка обмежує довжину вектора $\boldsymbol{r}^s$

$$-1 \le r_j^s \le 1,\ j=1,\ldots,n. \quad (11.12)$$

Серед всіх можливих напрямків знаходимо напрямок $\boldsymbol{r}^s$, який мінімізує похідну по напрямку $\boldsymbol{r}^s$ від функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^s$. Він є оптимальним розв'язком задачі мінімізації

$$D_{\boldsymbol{r}^s} f(\boldsymbol{x}^s) = (\nabla f(\boldsymbol{x}^s), \boldsymbol{r}^s) \to \min, \quad (11.13)$$

з обмеженнями (11.11) і (11.12).

Розв'язавши задачу (11.11)(11.13), перевіряємо умову:

- якщо оптимальне значення функції (11.13) невід'ємне, то кінець обчислень, оскільки підхожих напрямків в точці $\boldsymbol{x}^s$ не існує; $\boldsymbol{x}^s$ стаціонарна точка функції $f(\boldsymbol{x})$ в області D;
- якщо оптимальне значення функції (11.13) від'ємне, то виконуємо дії пунктів 4, 5, 6, замінивши антиградієнт $-\nabla f(\boldsymbol{x}^s)$ вектором знайденого можливого і підхожого напрямку $\boldsymbol{r}^s$.

Приклад 11.2. *Розв'язати задачу комбінованим методом*

$$f(x_1,x_2) = (x_1 - 6)^2 + (x_2 - 4)^2 \to \min,$$
$$x_1 + x_2 \le 8,$$
$$x_1 + 3x_2 \le 18,$$
$$x_1 \ge 0,\ x_2 \ge 0,$$

вибираючи $\boldsymbol{x}^0 = (2,4)$.

Розв'язування. Обчислимо градієнт $\nabla f(x_1,x_2)$

$$\nabla f(x_1,x_2) = (2x_1 - 12,\ 2x_2 - 8).$$

1-а ітерація.

$$\nabla f(\boldsymbol{x}^0) = (-8,\ 0) \ne 0.$$

Визначаємо активні обмеження в точці $\boldsymbol{x}^0$:

$$2 + 4 = 6 < 8,$$
$$2 + 3 \times 4 = 14 < 18,$$
$$2 > 0,$$
$$4 > 0.$$

Активних обмежень немає. Точка $\boldsymbol{x}^0$ є внутрішньою точкою допустимої області задачі, тому напрямок антиградієнта $-\nabla f(\boldsymbol{x}^0)$ є можливим. Будуємо

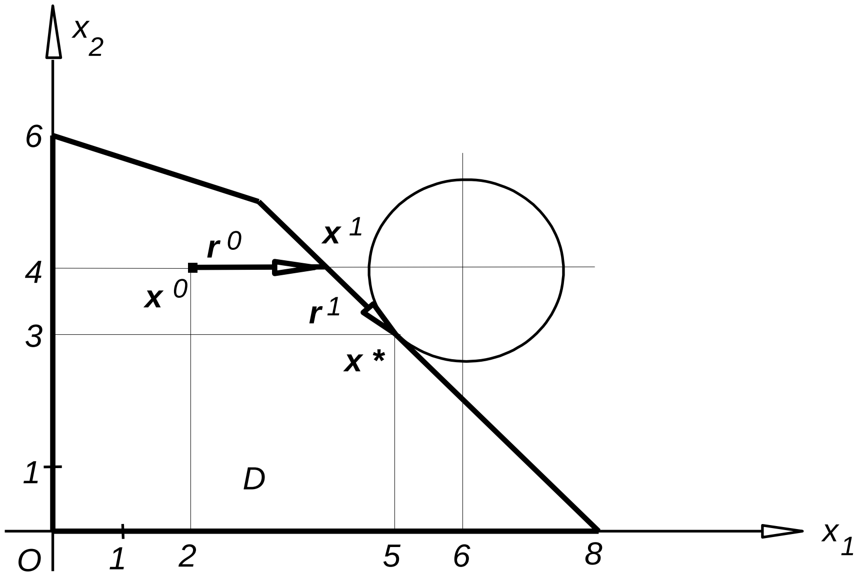


Рис. 11.2

промінь $\boldsymbol{x}(\rho)$, що виходить із точки $\boldsymbol{x}^0$ в напрямі антиградієнта $-\nabla f(\boldsymbol{x}^0)$:

$$\boldsymbol{x}(\rho) = \boldsymbol{x}^0 - \rho\nabla f(\boldsymbol{x}^0) = (2,4) - \rho(-8,0) = (2+8\rho, 4), \ \rho > 0.$$

Визначаємо інтервал допустимих значень ρ:

Звідки маємо $\rho \in (0, 1/4]$.

Знаходимо найменше значення функції $g(\rho) = f(\boldsymbol{x}^0 - \rho\nabla f(\boldsymbol{x}^0))$ на відрізку $(0, 1/4]$. Скористаємось необхідними умовами екстремуму функції однієї змінної:

$$\frac{dg(\rho)}{d\rho} = \frac{d}{d\rho} f(\boldsymbol{x}^0 - \rho\nabla f(\boldsymbol{x}^0)) = (-\nabla f(\boldsymbol{x}^0 - \rho\nabla f(\boldsymbol{x}^0)), \nabla f(\boldsymbol{x}^0)) =$$

$$= -((2(2+8\rho)-12, 0), (-8, 0)) = 8(16\rho - 8) = 0,$$

звідки маємо $\rho = 1/2$.

Так як $\dfrac{d^2 g(\rho)}{d\rho^2} = 128 > 0$, то $\rho = 1/2$ є точкою мінімуму функції $g(\rho)$ для всіх ρ. Оскільки ця точка лежить праворуч від допустимого інтервалу, а функція $g(\rho)$ опукла донизу, то мінімальне значення функції $g(\rho)$ досягається на правому кінці допустимого інтервалу для ρ. Отже, $\rho_0 = 1/4$.

Обчислюємо $\boldsymbol{x}^1 = \boldsymbol{x}^0 - \rho_0\nabla f(\boldsymbol{x}^0) = (2,4) - (1/4)(-8,0) = (4,4)$.

<u>*2-а ітерація*</u>.

$$\nabla f(\boldsymbol{x}^1) = (-4, 0) \neq 0.$$

Визначаємо активні обмеження в точці $\boldsymbol{x}^1$:

$4 + 4 = 8,$ (Активне обмеження)
$4 + 3 \times 4 = 16 < 18,$
$4 > 0,$
$4 > 0.$

Перевіряємо, чи є антиградієнт $-\nabla f(\boldsymbol{x}^1)$ можливим напрямком. Будуємо промінь $\boldsymbol{x}(\rho)$, що виходить з точки $\boldsymbol{x}^1$ в напрямі антиградієнта $-\nabla f(\boldsymbol{x}^1)$:

$$\boldsymbol{x}(\rho) = \boldsymbol{x}^1 - \rho\nabla f(\boldsymbol{x}^1) = (4,4) - \rho(-4,0) = (4+4\rho, 4), \ \rho > 0.$$

Підставляємо точку $\boldsymbol{x}^1$ в активне обмеження: $4+4\rho+4\le 8$, звідки $\rho\le 0$. Водночас повинно бути $\rho>0$. Тому $-\nabla f(\boldsymbol{x}^1)$ не є можливим напрямком.

Визначимо можливий напрямок. Нехай $\boldsymbol{r}^1=(r_1^1, r_2^1)$ □ вектор невідомого можливого напрямку. Будуємо промінь $\boldsymbol{x}'(\rho)$, що виходить із точки $\boldsymbol{x}^1$ в напрямі $\boldsymbol{r}^1$:

$$\boldsymbol{x}'(\rho) = \boldsymbol{x}^1+\rho\boldsymbol{r}^1 = (4,4)+\rho(r_1^1, r_2^1) = (4+\rho r_1^1, 4+\rho r_2^1),\ \rho>0.$$

Підставляючи точку $\boldsymbol{x}'(\rho)$ в активне обмеження, отримаємо

$$4+\rho r_1^1+4+\rho r_2^1 \le 8,$$

або

$$r_1^1+r_2^1 \le 0.$$

Додаємо до останньої нерівності умову нормування

$$-1 \le r_1^1 \le 1,$$
$$-1 \le r_2^1 \le 1.$$

Знаходимо похідну за напрямком $\boldsymbol{r}^1$ від функції $f(\boldsymbol{x})$ в точці $\boldsymbol{x}^1$:

$$D_{\boldsymbol{r}^1} f(\boldsymbol{x}^1) = (\nabla f(\boldsymbol{x}^1), \boldsymbol{r}^1) = ((-4,\ 0), (r_1^1, r_2^1)) = -4\ r_1^1.$$

Остаточно отримуємо задачу *ЛП* для визначення підхожого і можливого напрямку

$$-4\ r_1^1 \to \min,$$
$$r_1^1+r_2^1 \le 0,$$
$$-1 \le r_1^1 \le 1,$$
$$-1 \le r_2^1 \le 1.$$

Оптимальним розв'язком цієї задачі є вектор $\boldsymbol{r}^1=(1,-1)$. Відповідні значення r_1^1, r_2^1 дорівнюють $r_1^1=1$, $r_2^1=-1$. Значення похідної за напрямком $\boldsymbol{r}^1$

$$D_{\boldsymbol{r}^1} f(\boldsymbol{x}^1) = -4\ r_1^1 = -4<0.$$

Отже, напрямок $\boldsymbol{r}^1=(1,-1)$ є можливим і підхожим. Будуємо промінь $\boldsymbol{x}'(\rho)$, що виходить із точки $\boldsymbol{x}^1$ в напрямку $\boldsymbol{r}^1$:

$$\boldsymbol{x}'(\rho) = \boldsymbol{x}^1+\rho\boldsymbol{r}^1 = (4,4)+\rho(1,-1) = (4+\rho, 4-\rho),\ \rho>0.$$

Підставляючи $\boldsymbol{x}'(\rho)$ у неактивні обмеження, визначаємо допустимий інтервал для ρ:

$$\begin{cases} 4+\rho+3(4-\rho)\le 18, \\ 4+\rho \ge 0, \\ 4-\rho \ge 0, \\ \rho>0. \end{cases}$$

Розв'язуючи систему, отримаємо $0<\rho\le 4$.

Обчислимо найменше значення функції $g(\rho)=f(\boldsymbol{x}^1+\rho\boldsymbol{r}^1)$ на відрізку $(0,4]$. Маємо

$$\frac{dg(\rho)}{d\rho} = \frac{d}{d\rho} f(\boldsymbol{x}^1+\rho\, r^1) = (\nabla f(\boldsymbol{x}^1+\rho\boldsymbol{r}^1), \boldsymbol{r}^1) =$$
$$= ((2(4+\rho)-12, 2(4-\rho)-8), (1,-1)) = 4\rho-4=0,$$

звідки знаходимо $\rho=1.$

Так як $\frac{d^2 g(\rho)}{d\rho^2}=4>0$, то $\rho=1$ є точкою мінімуму опуклої донизу функції $g(\rho)$. Точка $\rho=1$ належить допустимому інтервалу для ρ. Тому $\rho_1=1$ і

$$\boldsymbol{x}^2=\boldsymbol{x}^1+\rho_1\boldsymbol{r}^1=(4,4)+(1,-1)=(5,3).$$

<u>*3-я ітерація*</u>. Обчислюємо

$$\nabla f(\boldsymbol{x}^2)=(-2,-2)\neq 0.$$

Визначаємо активні обмеження в точці $\boldsymbol{x}^2$:

$5+3=8,$ (Активне обмеження)
$5+3\times 3=14<18,$
$5>0,$
$3>0.$

Перевіряємо, чи буде антиградієнт $-\nabla f(\boldsymbol{x}^2)$ можливим напрямком. Будуємо промінь, що виходить з точки $\boldsymbol{x}^2$ в напрямі антиградієнта $-\nabla f(\boldsymbol{x}^2)$:

$$\boldsymbol{x}(\rho)=\boldsymbol{x}^2-\rho\nabla f(\boldsymbol{x}^2)=(5,3)-\rho(-2,-2)=(5+2\rho,3+2\rho),\ \rho>0.$$

Підставляємо точку $\boldsymbol{x}(\rho)$ в активне обмеження: $5+2\rho+3+2\rho\leq 8,$ звідки $\rho\leq 0.$ В той же час $\rho>0.$ Тому $-\nabla f(\boldsymbol{x}^2)$ не є можливим напрямком.

Будуємо допоміжну *ЗЛП* для відшукання можливого напрямку. Нехай $\boldsymbol{r}^2=(r_1^2,r_2^2)$ □ вектор невідомого можливого напрямку, $\boldsymbol{x}'(\rho)$ □ промінь у напрямку $\boldsymbol{r}^2$. Маємо

$$\boldsymbol{x}'(\rho)=\boldsymbol{x}^2+\rho\boldsymbol{r}^2=(5,3)+\rho(r_1^2,r_2^2)=(5+\rho r_1^2,3+\rho r_2^2),\ \rho>0.$$

Підстановка $\boldsymbol{x}'(\rho)$ в активне обмеження дає нам

$$5+\rho r_1^2+3+\rho r_2^2\leq 8,\ \rho>0,$$

або з умовами нормування

$$r_1^2+r_2^2\leq 0,$$
$$-1\leq r_1^2\leq 1,$$
$$-1\leq r_2^2\leq 1.$$

За цих умов ми повинні мінімізувати функцію

$$D_{\boldsymbol{r}^2}f(\boldsymbol{x}^2)=(\nabla f(\boldsymbol{x}^2),\boldsymbol{r}^2)=((-2,-2),(r_1^2,r_2^2))=-2r_1^2-2r_2^2.$$

Оскільки $r_1^2+r_2^2\leq 0$, то $-2(r_1^2+r_2^2)\geq 0$, тобто підхожих напрямків не існує. Отже, точка $\boldsymbol{x}^*=\boldsymbol{x}^2=(5,3)$ є оптимальним розв'язком задачі, $f(\boldsymbol{x}^*)=f(5,3)=2$.

Лекція 34. *Методи штрафних та бар'єрних функцій.– 2год. [1]*

Завдання для самостійної роботи. *Суть методу штрафних функцій. Відмінності від методу бар'єрних функцій. Означення штрафної функції. Вибір штрафних функцій. Теорема про збіжність методу штрафних функцій. -4год. [1-3.]*

Методи штрафних та бар'єрних функцій

Ідея методів штрафних та бар'єрних функцій полягає в зведенні *ЗНЛП* з обмеженнями до спеціальної задачі безумовної оптимізації. При цьому обмеження вихідної задачі включаються в цільову функцію цієї допоміжної оптимізаційної задачі. Зауважимо, що ця ідея вже використовувалась при розв'язуванні класичної задачі пошуку умовного екстремуму методом Лагранжа, а також в задачі опуклого програмування (теорема Куна-Таккера).

Нехай маємо задачу

$$\min \{f^0(\boldsymbol{x}) : f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m,\ \boldsymbol{x} \in E^n\}. \qquad (12.1)$$

Введемо до розгляду функцію

$$P(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \in D = \{\boldsymbol{x} \in E^n : f^i(\boldsymbol{x}) \le 0,\ i=1,\dots,m\}, \\ \infty, & \boldsymbol{x} \notin D. \end{cases} \qquad (12.2)$$

Замінимо задачу умовної оптимізації (12.1) задачею безумовної оптимізації

$$\min \{F(\boldsymbol{x}) : F(\boldsymbol{x}) = f^0(\boldsymbol{x}) + P(\boldsymbol{x}),\ \boldsymbol{x} \in E^n\}. \qquad (12.3)$$

Функцію $P(\boldsymbol{x})$ називають *штрафною*, оскільки вона викликає нескінченний штраф за вихід з області *D*.

Очевидно, що оптимальні розв'язки задач (12.1) і (12.3) однакові.

Геометрична інтерпретація викладеного для функції однієї змінної приведена нижче (див. рис. 12.1).

Очевидно також і те, що для так введеної функції штрафу $P(\boldsymbol{x})$ (див. співвідношення (12.2)) задача (12.3) розв'язуванню не піддається. Тому в дійсності замість вказаної $P(\boldsymbol{x})$ розглядається послідовність штрафних функцій, які апроксимують $P(\boldsymbol{x})$ і мають хороші аналітичні властивості.

Означення 12.1. *Функція* $P(f^1(\boldsymbol{x}),\dots,f^m(\boldsymbol{x}),\boldsymbol{r})$, *де* $\boldsymbol{r} = (r_1,\dots,r_m)$ — *деякий вектор керуючих параметрів*, *а* $f^i(\boldsymbol{x})$, *i=1,…,m,* — *функції обмежень задачі (12.1), називається штрафною, якщо*

$$\lim_{r_i \to \infty\ (i=1,\dots,m)} P(y_1,\dots,y_m,\boldsymbol{r}) = \begin{cases} 0, & y_i \le 0,\ i=1,\dots,m, \\ \infty, & \text{в іншому випадку}. \end{cases}$$

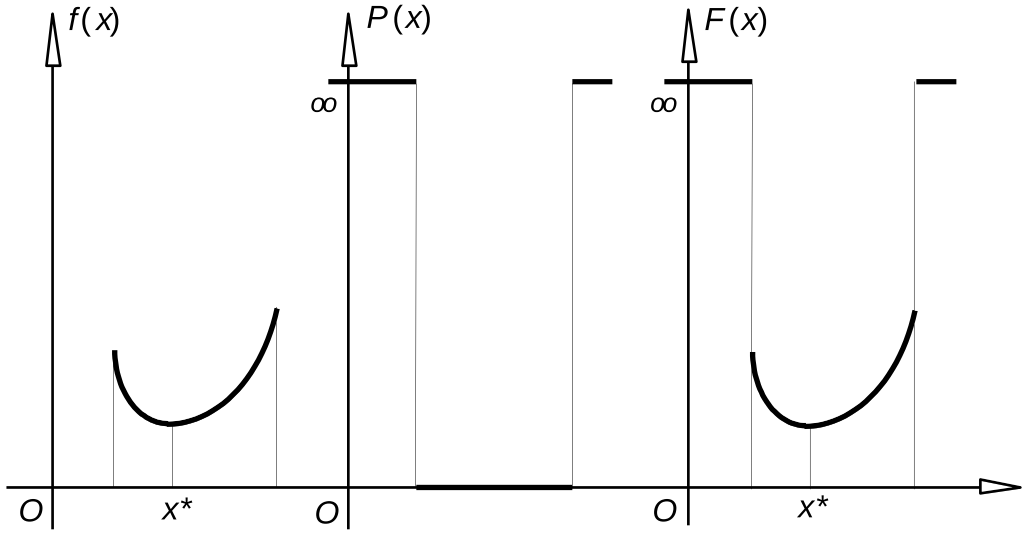


Рис. 12.1

Замість задачі (12.3) введемо до розгляду послідовність задач безумовної оптимізації

$$\min\{F(\boldsymbol{x},\boldsymbol{r}) = f^0(\boldsymbol{x}) + P(f^1(\boldsymbol{x}),...,f^m(\boldsymbol{x}),\boldsymbol{r})\}. \qquad (12.4)$$

Метод штрафних функцій полягає в переході від задачі (12.1) до послідовності задач безумовної оптимізації (12.4) з наступним їх розв'язуванням яким-небудь методом безумовної оптимізації.

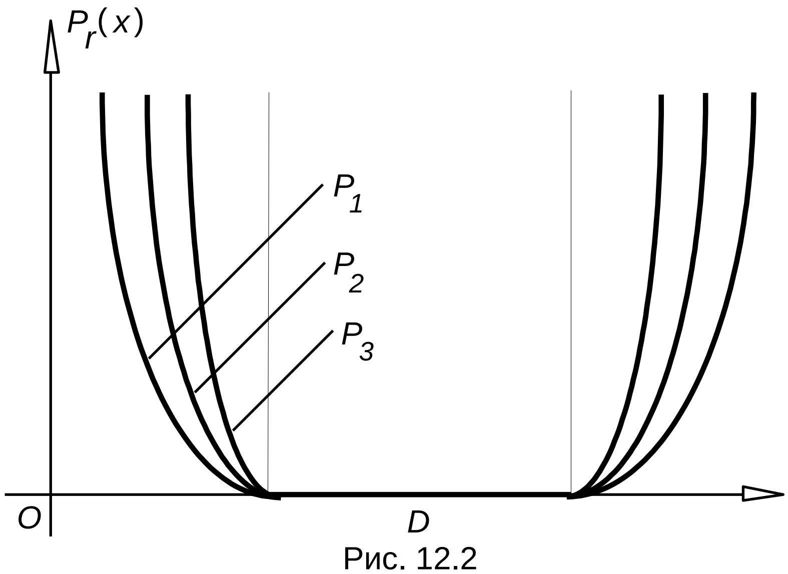


Рис. 12.2

При вдалому виборі штрафних функцій $P(f^1(\boldsymbol{x}),...,f^m(\boldsymbol{x}),\boldsymbol{r})$ можна очікувати, що послідовність розв'язків задач (12.4) буде збігатися до розв'язку задачі (12.1).

Вкажемо на відмінність *методу бар'єрних функцій* від методу штрафних функцій. В методі штрафних функцій апроксимація функції $P(\boldsymbol{x})$ здійснюється зовні області D (див. рис. 12.2), а в методі бар'єрних функцій функції $B(f^1(\boldsymbol{x}),...,f^m(\boldsymbol{x}),\boldsymbol{r})$ наближають $P(\boldsymbol{x})$ зсередини області D (див. рис. 12.3).

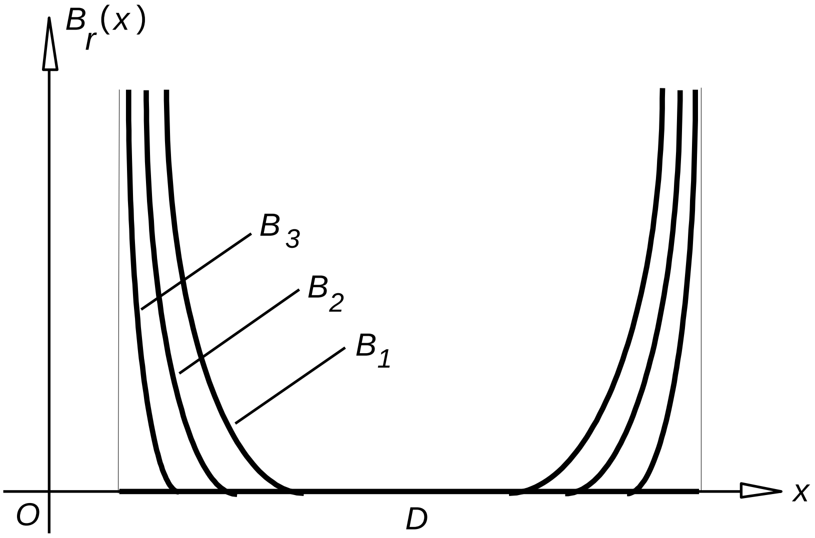


Рис. 12.3

Найчастіше за штрафну вибирають функцію такого вигляду

$$P(f^1(\boldsymbol{x}),...,f^m(\boldsymbol{x}),r) = \frac{1}{r}P(\boldsymbol{x}) = \frac{1}{r}\sum_{i=1}^{m} P_i(f^i(\boldsymbol{x})), \tag{12.5}$$

де r — керуючий параметр, а функції $P_i(y)$ визначені, неперервні і невід'ємні для довільного y, причому: $P_i(y)=0$ при $y \le 0$ і $P_i(y) \to \infty$ монотонно при $y \to \infty$.

Прикладами функцій $P_i(f^i(\boldsymbol{x}))$ можуть бути такі функції:

$$P_i^1(f^i(\boldsymbol{x})) = \max\{0, f^i(\boldsymbol{x})\}, \tag{12.6}$$

$$P_i^2(f^i(\boldsymbol{x})) = [\max\{0, f^i(\boldsymbol{x})\}]^2. \tag{12.7}$$

Зауважимо, що функція (12.6) в загальному випадку недиференційовна, навіть, якщо $f^i(\boldsymbol{x}) \in C^1$, в той час як функція (12.7) — диференційовна, якщо диференційовна $f^i(\boldsymbol{x})$.

Якщо ж обмеження задачі (12.1) мають вигляд рівнянь $f^i(\boldsymbol{x})=0$, $i=1,\ldots,m$, то функції $P_i(y)$ у (12.5) повинні задовольняти такі умови: $P_i(y)=0$ при $y=0$ і $P_i(y) \to \infty$ монотонно при $|y| \to \infty$. Для цього випадку функції $P_i(f^i(\boldsymbol{x}))$ найдоцільніше вибирати у вигляді

$$P_i(f^i(\boldsymbol{x})) = [f^i(\boldsymbol{x})]^2.$$

Теорема 12.1 (про збіжність методу штрафних функцій). *Нехай* $\boldsymbol{x}^*$ *є розв'язком задачі* (12.1)

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in D} f^0(\boldsymbol{x}), \tag{12.8}$$

а $\boldsymbol{x}^*(r)$ $\forall r>0$ *є розв'язком задачі* (12.4)

$$\boldsymbol{x}^*(r) = \arg\min_{\boldsymbol{x} \in E^n} F(\boldsymbol{x}, r). \tag{12.9}$$

Якщо штрафна функція визначається рівністю (12.5), *функції* $f^i(\boldsymbol{x})$, *i=1,...,m, неперервні і існує замкнена множина* $Y \subset E^n$ *така, що* $\boldsymbol{x}^*(r) \in Y \ \forall r>0$, *то*

$$\lim_{r \to 0} F(\boldsymbol{x}^*(r), r) = f^0(\boldsymbol{x}^*). \qquad (12.10)$$

<u>Доведення</u>. Доведемо, що одночасно мають місце дві нерівності

$$\lim_{r \to 0} F(\boldsymbol{x}^*(r), r) \le f^0(\boldsymbol{x}^*). \qquad (12.11)$$

$$\lim_{r \to 0} F(\boldsymbol{x}^*(r), r) \ge f^0(\boldsymbol{x}^*). \qquad (12.12)$$

Це і буде означати, що виконується рівність (12.10).

Доведемо нерівність (12.11). Маємо $\forall r_2 > 0$ в силу (12.9)

$$F(\boldsymbol{x}^*(r_1), r_1) = \min_{\boldsymbol{x} \in E^n} F(\boldsymbol{x}, r_1) \le F(\boldsymbol{x}^*(r_2), r_1). \qquad (12.13)$$

Нехай $r_1 \ge r_2 > 0$. Оскільки за означенням (12.5) $P(\boldsymbol{x}) \ge 0$, то

$$(1/r_1)\, P(\boldsymbol{x}^*(r_2)) \le (1/r_2)\, P(\boldsymbol{x}^*(r_2)).$$

Тоді

$$F(\boldsymbol{x}^*(r_1), r_1) \le F(\boldsymbol{x}^*(r_2), r_1) = f^0(\boldsymbol{x}^*(r_2)) + (1/r_1)\, P(\boldsymbol{x}^*(r_2)) \le$$

$$\le f^0(\boldsymbol{x}^*(r_2)) + (1/r_2)\, P(\boldsymbol{x}^*(r_2)) = F(\boldsymbol{x}^*(r_2), r_2).$$

Отже при $r_1 \ge r_2 > 0$

$$F(\boldsymbol{x}^*(r_1), r_1) \le F(\boldsymbol{x}^*(r_2), r_2), \qquad (12.14)$$

тобто при спаданні r функція $F(\boldsymbol{x}^*(r), r)$ не спадає.

За означенням (12.5) $P(\boldsymbol{x})=0$ при $\boldsymbol{x} \in D$. Тому $P(\boldsymbol{x}^*)=0$, оскільки $\boldsymbol{x}^* \in D$, і

$$F(\boldsymbol{x}^*(r), r) = \min_{\boldsymbol{x} \in E^n} F(\boldsymbol{x}, r) \le F(\boldsymbol{x}^*, r) = f^0(\boldsymbol{x}^*) + (1/r)P(\boldsymbol{x}^*) = f^0(\boldsymbol{x}^*). \qquad (12.13)$$

Отже, $\forall r > 0$

$$F(\boldsymbol{x}^*(r), r) \le f^0(\boldsymbol{x}^*). \qquad (12.15)$$

Тоді існує границя

$$\lim_{r \to 0} F(\boldsymbol{x}^*(r), r) = F^* \qquad (12.16)$$

і виконується нерівність

$$F^* = \lim_{r \to 0} F(\boldsymbol{x}^*(r), r) \le f^0(\boldsymbol{x}^*). \qquad (12.17)$$

Доведемо тепер нерівність (12.12). Маємо $\forall r_1 > 0$

$$F(\boldsymbol{x}^*(r_2), r_2) = \min_{\boldsymbol{x} \in E^n} F(\boldsymbol{x}, r_2) \le F(\boldsymbol{x}^*(r_1), r_2). \qquad (12.18)$$

Розпишемо (12.13) і (12.18). Маємо $\forall r_1 > 0$ та $\forall r_2 > 0$

$$f^0(\boldsymbol{x}^*(r_1)) + (1/r_1)\, P(\boldsymbol{x}^*(r_1)) \le f^0(\boldsymbol{x}^*(r_2)) + (1/r_1)\, P(\boldsymbol{x}^*(r_2)), \qquad (12.19)$$

$$f^0(\boldsymbol{x}^*(r_2)) + (1/r_2)\, P(\boldsymbol{x}^*(r_2)) \le f^0(\boldsymbol{x}^*(r_1)) + (1/r_1)\, P(\boldsymbol{x}^*(r_1)). \qquad (12.20)$$

Додамо нерівності (12.19), (12.20). Отримаємо

$$(1/r_2)(P(\boldsymbol{x}^*(r_2)) - P(\boldsymbol{x}^*(r_1))) \le (1/r_1)(P(\boldsymbol{x}^*(r_2)) - P(\boldsymbol{x}^*(r_1))) \quad (12.21)$$

або

$$(1/r_1 - 1/r_2)(P(\boldsymbol{x}^*(r_2)) - P(\boldsymbol{x}^*(r_1))) \ge 0. \quad (12.22)$$

Нехай $r_1 \ge r_2 > 0$. Тоді $1/r_1 \le 1/r_2$ або $(1/r_1 - 1/r_2) \le 0$ і, як наслідок, із нерівності (12.22) отримаємо

$$(P(\boldsymbol{x}^*(r_2)) - P(\boldsymbol{x}^*(r_1))) \le 0. \quad (12.23)$$

Із нерівності (12.19) маємо

$$f^0(\boldsymbol{x}^*(r_1)) - f^0(\boldsymbol{x}^*(r_2)) \le (1/r_1)\,(P(\boldsymbol{x}^*(r_2)) - P(\boldsymbol{x}^*(r_1)))$$

або, враховуючи (12.23) та умову $r_1 > 0$,

$$f^0(\boldsymbol{x}^*(r_1)) - f^0(\boldsymbol{x}^*(r_2)) \le 0.$$

Отже, при $r_1 \ge r_2 > 0$

$$f^0(\boldsymbol{x}^*(r_1)) \le f^0(\boldsymbol{x}^*(r_2)), \quad (12.24)$$

тобто функція $f^0(\boldsymbol{x}^*(r))$ не спадає при спаданні r.

Далі, за означенням (12.5) $P(\boldsymbol{x}) \ge 0$, тому $\forall r > 0$

$$f^0(\boldsymbol{x}^*(r)) \le f^0(\boldsymbol{x}^*(r)) + (1/r)\,P(\boldsymbol{x}^*(r)) = F(\boldsymbol{x}^*(r), r). \quad (12.25)$$

В свою чергу функція $F(\boldsymbol{x}^*(r), r)$ (згідно співвідношень (12.14), (12.16)) також не спадає при $r \to 0$ і має границю, рівну F^*. Тому $\forall r > 0$

$$f^0(\boldsymbol{x}^*(r)) \le F^*,$$

тобто $f^0(\boldsymbol{x}^*(r))$ обмежена зверху. Оскільки при $r \to 0$ послідовність $f^0(\boldsymbol{x}^*(r))$ неспадна, то вона має границю при $r \to 0$, причому

$$\lim_{r \to 0} f^0(\boldsymbol{x}^*(r)) \le F^* = \lim_{r \to 0} F(\boldsymbol{x}^*(r), r). \quad (12.26)$$

Покажемо, що

$$\lim_{r \to 0} f^0(\boldsymbol{x}^*(r)) = f^0(\boldsymbol{x}^*).$$

Від супротивного припустимо, що існують $\delta > 0$ та послідовність $r_k \to 0$ при $k \to \infty$ такі, що $\forall k$ виконується

$$|f^0(\boldsymbol{x}^*(r_k)) - f^0(\boldsymbol{x}^*)| > \delta. \quad (12.27)$$

За умовами теореми $\forall r > 0$ $\boldsymbol{x}^*(r) \in Y$, де Y — замкнена множина. Тоді із послідовності $\boldsymbol{x}^*(r_k)$ завжди можна вибрати збіжну підпослідовність. Не обмежуючи загальності, будемо вважати, що

$$\lim_{k \to \infty} \boldsymbol{x}^*(r_k) = \bar{\boldsymbol{x}}.$$

Оскільки

$$(F(\boldsymbol{x}^*(r_k), r_k) = f^0(\boldsymbol{x}^*(r_k)) + (1/r_k)\,P(\boldsymbol{x}^*(r_k)),$$

то

$$P(\boldsymbol{x}^*(r_k)) = r_k\,(F(\boldsymbol{x}^*(r_k), r_k) - f^0(\boldsymbol{x}^*(r_k))).$$

Функція $P(\boldsymbol{x})$ неперервна, тому, переходячи до границі при $k \to \infty$, отримаємо

$$\lim_{k\to\infty} P(\boldsymbol{x}^*(r_k)) = 0 = P(\bar{\boldsymbol{x}}).$$

Оскільки $P(\boldsymbol{x}) = 0 \ \forall \boldsymbol{x} \in D$, то

$$\bar{\boldsymbol{x}} \in D. \tag{12.28}$$

Ми довели (див. нерівність (12.17)), що

$$\lim_{r\to 0} F(\boldsymbol{x}^*(r), r) \le f^0(\boldsymbol{x}^*).$$

Також було доведено (див. нерівність (12.26)), що

$$\lim_{r\to 0} f^0(\boldsymbol{x}^*(r)) \le \lim_{r\to 0} F(\boldsymbol{x}^*(r), r).$$

Тому

$$\lim_{r\to 0} f^0(\boldsymbol{x}^*(r)) \le f^0(\boldsymbol{x}^*),$$

тим більше буде

$$\lim_{k\to\infty} f^0(\boldsymbol{x}^*(r_k)) \le f^0(\boldsymbol{x}^*).$$

Оскільки $f^0(\boldsymbol{x})$ — неперервна функція, то із останньої нерівності отримуємо при $k \to \infty$

$$f^0(\bar{\boldsymbol{x}}) \le f^0(\boldsymbol{x}^*). \tag{12.29}$$

Так як $\bar{\boldsymbol{x}} \in D$, а

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in D} f^0(\boldsymbol{x}^*)$$

за умовами теореми, то отримана нерівність суперечить умовам теореми, тобто наше припущення було невірним і тому

$$\lim_{r\to 0} f^0(\boldsymbol{x}^*(r)) = f^0(\boldsymbol{x}^*).$$

Теорема доведена.

Приклад 12.1. *Розв'язати задачу методом штрафних функцій*

$$\min \{ x^2 - 10x : x \le 1 \}.$$

Покладемо

$$P(f^1(x), r) = (1/r)(\max\{0, f^1(x)\})^2 = (1/r)(\max\{0, x-1\})^2,$$

тоді

$$F(x, r) = f^0(x) + P(f^1(x), r) = x^2 - 10x + (1/r)(\max\{0, x-1\})^2.$$

Схематичні графіки функцій $f^0(x)$, $P(f^1(x), r)$, $F(x, r)$ наведені на рис. 12.4. Функція $F(x, r)$ опукла і диференційовна по x при довільних додатних значеннях параметра r.

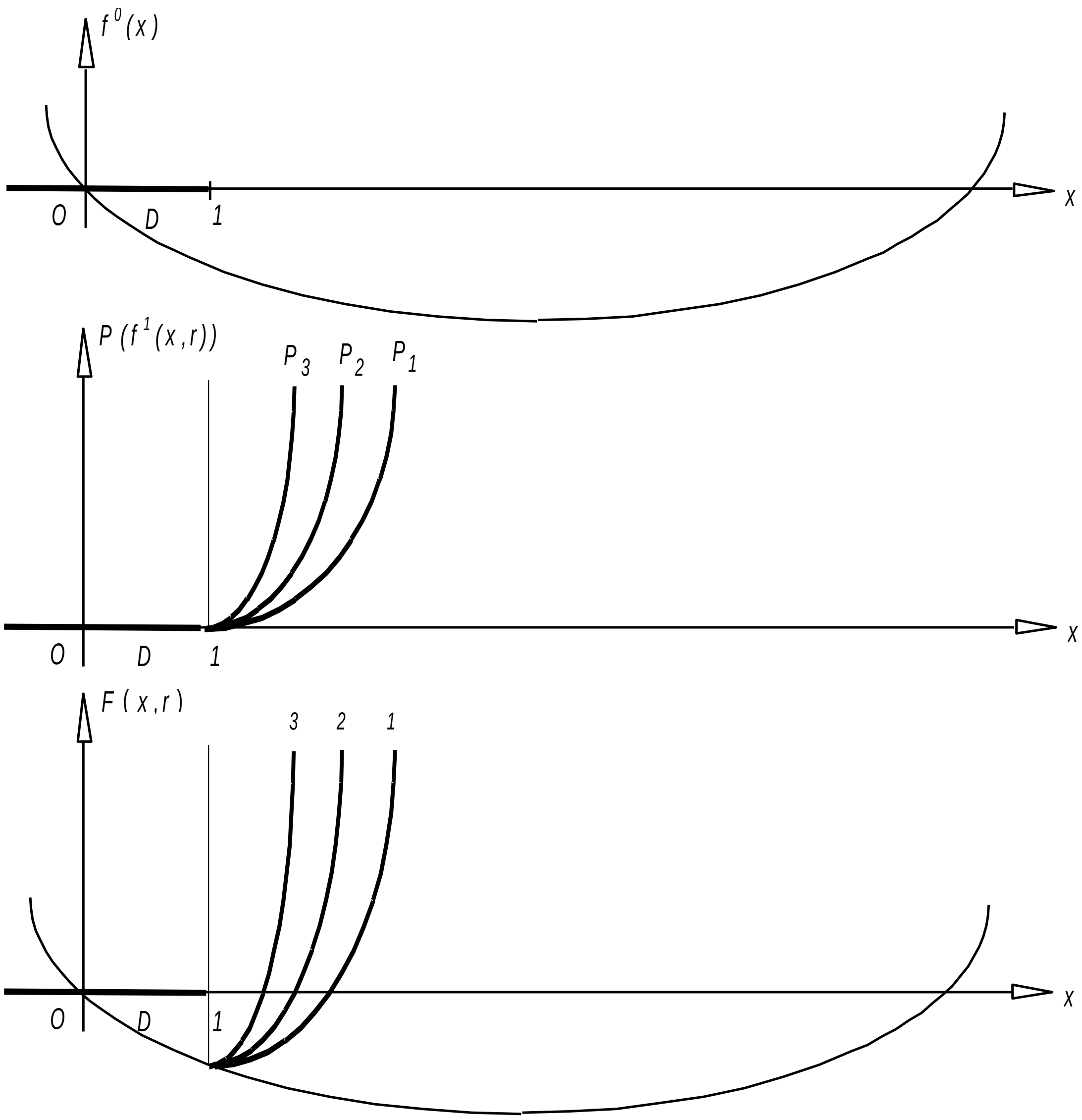


Рис. 12.4

Маємо:

а) всередині допустимої області (при $x<1$)

$F(x,r) = x^2 - 10x$,

б) на межі та зовні допустимої області (при $x \geq 1$)

$F(x,r) = x^2 - 10x + (1/r)(x-1)^2$.

Скористаємось необхідними умовами екстремуму:

а) $\begin{cases} dF/dx = 2x - 10 = 0, \\ x < 1, \end{cases}$ або $\begin{cases} x = 5, \\ x < 1. \end{cases}$

Отримана система не має розв'язків всередині допустимої області.

б) $\begin{cases} dF/dx = 2x - 10 + (2/r)(x-1) = 0, \\ x \geq 1, \end{cases}$ або $\begin{cases} x^*(r) = (5r+1)/(r+1), \\ x \geq 1. \end{cases}$

Остання система дає стаціонарну точку $x^*(r)=(5r+1)/(r+1)$, в якій функція $F(x,r)$, як опукла донизу по x, має мінімум. Оптимальний розв'язок вихідної задачі

$$x^* = \lim_{r\to 0} x^*(r) = \lim_{r\to 0} ((5r+1)/(r+1)) = 1 .$$

Слід зауважити, що розглянутий приклад розв'язаний аналітично в силу його простоти.

Приклад 12.2. *Розв'язати задачу методом штрафних функцій*

$$\min \{ x^2 + xy + y^2 : \ x + y = 2 \} .$$

Маємо для цієї задачі

$$P(f^1(x,y),r) = (1/r)(f^1(x,y))^2 = (1/r)(x + y - 2)^2,$$

$$F(x,y,r) = f^0(x) + P(f^1(x,y),r) = x^2 + xy + y^2 + (1/r)(x + y - 2)^2 .$$

Для довільного додатного значення параметра r функція $F(x,y,r)$ опукла і диференційовна по x та y. Скориставшись необхідними умовами екстремуму, отримаємо систему:

$$\begin{cases} \frac{\partial}{\partial x} F(x,y,r) = 2x + y + (2/r)(x + y - 2) = 0, \\ \frac{\partial}{\partial y} F(x,y,r) = x + 2y + (2/r)(x + y - 2) = 0, \end{cases}$$

розв'язавши яку, матимемо:

$$x^*(r) = y^*(r) = (4/r)/(3 + 4/r) = 4/(3r + 4),$$

звідки

$$x^* = y^* = \lim_{r\to 0} 4/(3r+4) = 1.$$

ЗАВДАННЯ МОДУЛЬНОЇ КОНТРОЛЬНОЇ РОБОТИ №6

Приклад типового завдання.

1. Звести дану задачу до задачі мінімізації.
2. Перевірити цільову функцію на опуклість.
3. Перевірити допустиму область задачі на опуклість.
4. За допомогою необхідних та достатних умов дослідити цільову функцію задачі на безумовний екстремум. Знайти точку її безумовного абсолютного мінімуму.
5. Обчислити найближчу вершину допустимої області до точки безумовного абсолютного мінімуму цільової функції.
6. Вибрати знайдену вершину за початкову та виконати одну ітерацію методу найшвидшого спуску для безумовної мінімізації цільової функції. Перевірити, чи задовольняє цільова функція умови теореми про збіжність методу найшвидшого спуску.
7. Вибрати знайдену вершину за початкову та розв'язати задачу методом можливих напрямків.
8. Побудувати допустиму область задачі графічно та проінтерпретувати графічно метод можливих напрямків.
9. Розв'язати задачу аналітично, використавши представлення кутового коефіцієнта дотичної до лінії рівня цільової функції як функції координат точки площини.
10. На основі умов оптимальності Куна-Таккера записати задачу лінійного програмування для визначення сідлової точки функції Лагранжа задачі та виконати одну ітерацію квадратичного симлекс-методу.

Контрольні запитання до змістового модуля 6.

1. Нелінійне програмування, класифікація задач.
2. Основні методи розв'язування задач НП, порівняння з методами розв'язування задач ЛП.
3. Геометрична інтерпретація задачі НП у порівнянні з геометричною інтерпретацією задачі ЛП.
4. Основні відмінності задач НП від задач ЛП (на прикладах).
5. Класичні задачі оптимізації.
6. Унімодальні функції. Методи одновимірної оптимізації.
7. Гіперповерхня рівня функції.
8. Основні локальні властивості функцій багатьох змінних. Неперервність, диференційовність функції.
9. Градієнт, гессіан, похідна за напрямком. Геометричний зміст градієнта.
10. Глобальні та локальні екстремуми функції.
11. Означення опуклої множини.
12. Означення проекції точки на опуклу множину.
13. Теорема про проекцію точки на опуклу множину.
14. Теорема про відділяючу гіперплощину.
15. Означення опорної гіперплощини для множини.
16. Теорема про опорну гіперплощину.
17. Теорема про гіперплощину, що розділяє дві опуклі множини.
18. Означення опуклої (донизу) та увігнутої (опуклої доверху) функцій.
19. Приклади опуклих функцій.
20. Основні властивості опуклих функцій.
21. Теорема про опуклість множини, що визначається опуклою нерівністю.
22. Нерівність Ієнсена.
23. Властивість неперервності.
24. Існування похідної за напрямком.
25. Критерій опуклості диференційовної функції.
26. Субградієнт функції.
27. Теорема про існування субградієнта.
28. Теорема про властивості множини субградієнтів.
29. Теорема про обчислення похідної за напрямком.
30. Геометричний зміст субградієнта.
31. Теорема про глобальний мінімум опуклої функції.
32. Критерій глобального мінімуму опуклої функції.

33. Критерій глобального мінімуму опуклої диференційовної функції.
34. Задача опуклого програмування.
35. Достатні умови оптимальності в задачах НП.
36. Умови регулярності.
37. Теорема Куна-Таккера.
38. Умови оптимальності в задачах ОП (випадок диференційовності всіх функцій задачі), їх застосування до розв'язування задач квадратичного програмування.
39. Активні обмеження.
40. Можливі напрямки.
41. Допоміжна задача для відшукання можливих і підхожих напрямків.
42. Обчислення кроку методу можливих напрямків. Вибір початкового наближення в методі можливих напрямків.
43. Суть методу штрафних функцій. Відмінності від методу бар'єрних функцій.
44. Означення штрафної функції.
45. Вибір штрафних функцій.
46. Теорема про збіжність методу штрафних функцій.
47. Субградієнтний метод безумовної мінімізації опуклих функцій.
48. Метод проектування субградієнтів.
49. Теорема про збіжність методу проектування субградієнтів.
50. Розв'язування мінімаксних задач.

Перелік запитань до іспиту складається з контрольних запитань до змістових модулів 1-6.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА:

Основна:

1. Ю.Д.Попов, В.І.Тюптя, В.І.Шевченко “Методи оптимізації”, К.,2000.
2. Ю.М.Ермольев и др. “Математические методы исследования операций”, К.1977.
3. И.Н.Ляшенко и др. “Линейное и нелинейное программирование”, К.,1978.
4. И.А.Калихман, «Сборник задач по математическому программированию», М., 1975.
5. В.Ф.Капустин. Практические занятия по курсу математического программирования. Издательство Ленинградского университета, 1976.

Додаткова:

6. Ю.Д.Попов, В.І.Тюптя, В.І.Шевченко “Методичні вказівки до виконання лабораторних робіт з методів оптимізації”, К.1995, 1998, 2000.
7. Ю.П.Зайченко, “Исследование операций”, К.,1988г.
8. Ю.П.Зайченко, С.А.Шумилова “Исследование операций”, зб.задач, К.,1984г.
9. В.Г.Карманов. “Математическое программирование”, М.1975.