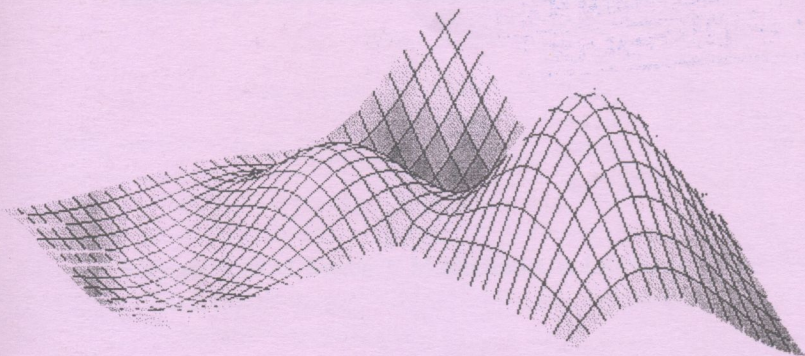


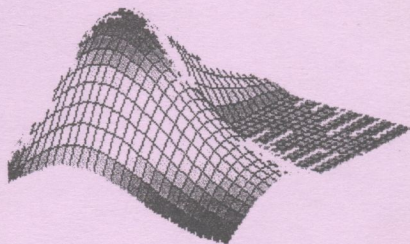
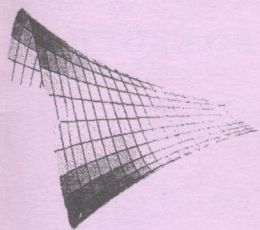
517.9(075)

І.В. КУЗЬМІН М.М. БИКОВ С.М. МОСКВІНА А.І. КУЗЬМІН

М54



МЕТОДИ  
ОПТИМІЗАЦІЇ  
СКЛАДНИХ  
СИСТЕМ



3482-13

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Вінницький державний технічний університет**

І.В.Кузьмін, М.М. Биков, С.М.Москвіна, А.І.Кузьмін

## **МЕТОДИ ОПТИМІЗАЦІЇ СКЛАДНИХ СИСТЕМ**

**НТБ ВНТУ**



**3482-13**

**517.9(075) M 54 2003**

Методи оптимізації складних систем

Затверджено Ученою радою Вінницького державного технічного університету як навчальний посібник для студентів спеціальності "Комп'ютеризовані системи управління і автоматики". Протокол № 4 від 28 листопада 2002 р.

Вінниця ВДТУ 2003

УДК 517.9

К 89

Рецензенти:

*Квєтний Р.Н.*, доктор технічних наук, професор

*Рудомін Г.А.*, кандидат технічних наук

*Юхимчук С.В.*, доктор технічних наук, професор

Рекомендовано до видання Ученою радою Вінницького державного технічного університету Міністерства освіти і науки України

**І. В. Кузьмін, М. М. Биков, С. М. Москвіна, А. І. Кузьмін.**

**К 89 Методи оптимізації складних систем.** Навчальний посібник. – Вінниця: ВДТУ, 2003. – 165 с.

Навчальний посібник стане в нагоді для студентів спеціальностей "Комп'ютеризовані системи управління і автоматики" та "Інтелектуальні системи" при вивченні дисциплін "Методи оптимізації", "Обчислювальні методи та застосування ЕОМ", "Основи проектування інтелектуальних систем", а також для широкого кола спеціалістів, що займаються проектуванням систем керування та прийняття рішень.



УДК 517.9

© І. В. Кузьмін, М. М. Биков, С. М. Москвіна, А. І. Кузьмін, 2003

## ВСТУП

В світі не відбувається нічого, в чому не було видно змісту якогось максимуму чи мінімуму.

**Леонардо Ейлер**

Серед великої кількості проблем розвитку людства важливою є проблема вдосконалення керування в усіх областях прогресу на основі методів пошуку найкращих рішень (методів оптимізації).

Цей посібник присвячений проблемам оптимізації. Біля її витоків стояли такі вчені, як Ферма, Ньютон, Лейбніц, Бернуллі. В даний час оптимізація стала класичним розділом математики, без якої сьогодні не обходиться ні один курс математичного аналізу.

Чому ж інтерес до оптимізації не тільки не слабшає, але постійно зростає? Можна вказати, щонайменше, дві основні причини. Оптимізація – це вибір, тобто те, чим постійно доводиться нам усім займатися у повсякденному житті, яке породжує все нові і нові задачі пошуку найкращих рішень, при чому складність їх зростає. Необхідні нові математичні моделі і методи, які враховують наявність багатьох критеріїв, проводять глобальний пошук оптимуму і т. ін. Іншими словами, життя змушує розвивати математичний апарат оптимізації і кінця цьому процесу поки що не видно. Друга причина полягає в тому, що реальні прикладні задачі оптимізації настільки складні, що їх не можливо розв'язати за допомогою олівця і паперу, а доводиться використовувати найсучаснішу обчислювальну техніку. При цьому методи оптимізації, розроблені математиками, ще необхідно перетворити у обчислювальні алгоритми. Тому найбільш актуальними методами оптимізації є методи математичної оптимізації, основані на аналізі і синтезі математичних моделей процесів функціонування досліджуваних систем з урахуванням їх складності, цільового призначення, сукупності обмежень.

В посібнику ставиться мета осмисленого перетворення практичної ідеї в теоретичну, в алгоритми і програми, що можуть бути реалізовані на ЕОМ, або використання пакетів програм, таких як MathCad та MathLab. Тому посібник розроблявся як базовий для дисциплін “Математичні методи оптимізації” та “Обчислювальні методи та застосування ЕОМ”, які читаються для студентів спеціальності 7091401. Крім того, окремі розділи посібника можуть бути використані в дисциплінах “Моделювання процесів в управлінні”, “Основи теорії складних систем”.

Над створенням посібника, який ви тримаєте в руках, працювала велика команда. Але особливо автори вдячні Шевчук О. і Заржецькій О., які допомагали підготувати посібник до друку в якості перекладачів, редакторів, секретарів-операторів ЕОМ.

# І СУТНІСТЬ МЕТОДІВ МАТЕМАТИЧНОЇ ОПТИМІЗАЦІЇ І ЇХ КЛАСИФІКАЦІЯ

## 1.1 Основні поняття і положення оптимізації

В практиці створення і застосування складних систем зустрічається ряд задач, в яких потрібно визначити максимальні чи мінімальні значення показників або параметрів. Методи визначення максимумів і мінімумів називають *екстремальними методами*.

До задач оптимізації можна віднести задачі визначення максимальної ефективності і продуктивності, максимальної вірогідності досягнення мети і швидкодії, пропускної спроможності, мінімальної вартості і собівартості, мінімальної маси, розмірності і енергоємності, та інші.

Для розв'язання подібних задач широко застосовується диференціальне числення, теорія ймовірностей та інформації, теорія масового обслуговування та надійності, теорія прийняття рішень та ін.

В теорії оптимізації (екстремалізації) розрізняють методи оптимізації функцій однієї змінної  $f(x)$  і функцій багатьох змінних  $f(x_1, x_2, \dots, x_n)$  [1].

При цьому передбачається, що ці функції неперервні і неперервні їх похідні. Графічна модель функції  $f(x)$  показана на рисунку 1.1.

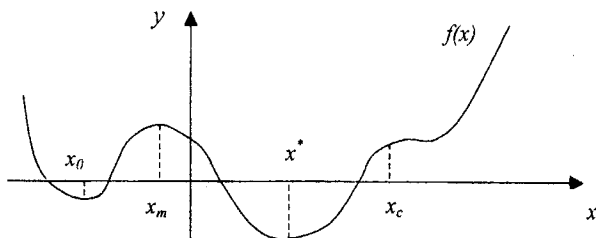


Рисунок 1.1 – Графічна модель функції

Функція  $f(x)$  має локальний та глобальний мінімум (максимум) в таких випадках:

- якщо існує деяка позитивна величина  $\delta$ , така що якщо  $|x - x_0| < \delta$ , то  $f(x) \leq f(x_0)$ , тобто якщо існує околі точки  $x_0$ , такий, що для всіх значень  $x$  в цьому околі  $f(x)$  більше  $f(x_0)$ , то функція  $f(x)$  має локальний мінімум.
- функція  $f(x)$  має глобальний мінімум в точці  $x^*$ , якщо для всіх  $x$  справедлива нерівність  $f(x) \geq f(x^*)$ .

В точках  $x_0$  і  $x^*$  похідна  $f'(x_0)$  (градієнт функції) дорівнює нулю. Отже,  $x_0$  і  $x^*$  будуть розв'язками рівняння

$$f'(x) = 0. \quad (1.1)$$

Точка  $x_m$ , в якій досягається локальний максимум, і точка  $x_c$ , в якій проходить горизонтальний перегин функції, також задовольняють це рівняння. Звідси випливає, що рівняння 1.1 є тільки *необхідною умовою* наявності мінімуму, чи максимуму.

*Достатню умову* можна сформулювати, розглядаючи рисунок 1.2 а, б, на якому зображені *опукла крива* (рис. 1.2, а), що відповідає максимуму функції  $f(x)$ , і *вгнута крива* (рис. 1.2, б), що відповідає мінімуму функції  $f(x)$ .

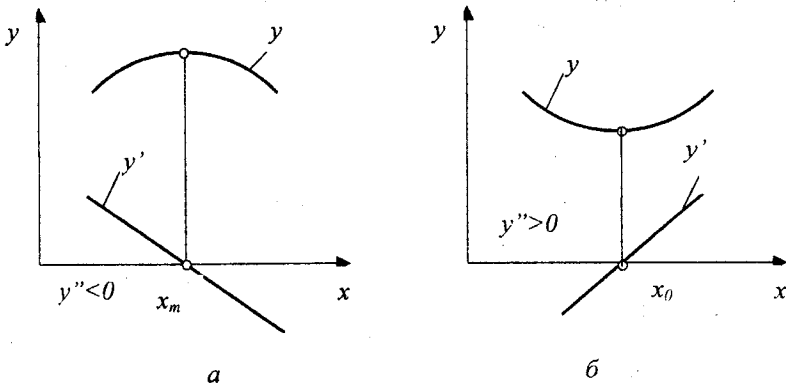


Рисунок 1.2 – Графічна ілюстрація локалізації екстремумів

З рисунку 1.2. видно, що функція  $f(x)$  неперервна і при значеннях аргументу  $x_m$  має максимум, при  $x_0$  — мінімум, при збільшенні  $x$  до значення  $x_m$  функція  $f(x)$  зростає, після  $x_m$  спадає. Якщо функція має мінімум, то, навпаки, спочатку функція спадає, а потім, після  $x_0$ , зростає.

При максимумі  $f(x)$  і значеннях аргументу, менших  $x_m$ , похідна від функції — додатна, а для значень більших за  $x_m$  — від'ємна, друга похідна при цьому менша нуля,  $f''(x) < 0$ . Для вгнутої кривої, навпаки, для значень аргументу, менших за  $x_0$ , похідна буде від'ємною, а для більших — додатною, друга похідна більша нуля  $f''(x) > 0$ .

З наведених міркувань випливає правило:

- якщо знайдено таке  $x_m$ , при якому  $f'(x) = 0$ ,  $f''(x) < 0$ , то функція  $f(x)$  має максимум;
- якщо знайдено таке  $x_0$ , при якому  $f'(x) = 0$ ,  $f''(x) > 0$ , то функція  $f(x)$  має мінімум;
- якщо друга похідна дорівнює нулю, то диференціювання виконується до тих пір, поки похідна не стане додатною або

від'ємною;

- якщо перша, третя і т.д. похідні до деякої похідної непарного порядку, дорівнюють нулю, а наступна похідна парного порядку додатна, то функція має мінімум, якщо вона від'ємна — функція має максимум.

Для функції двох змінних  $f(x_1, x_2)$ , необхідно знайти похідні  $f'(x_1)$  і  $f'(x_2)$ , прирівняти їх до нуля, а потім одночасно розв'язати рівняння  $f'(x_1)=0$  і  $f'(x_2)=0$ . Отримані значення  $x_1$  і  $x_2$  будуть відповідати максимуму чи мінімуму функції  $f(x_1, x_2)$ .

### Приклад 1.1

Дослідити характер точок перегину функції  $f(x) = x^3 - 2x^2 + x + 1$ .

Похідна  $f'(x) = 3x^2 - 4x + 1$ , або  $(3x - 1)(x - 1) = 0$ , тобто  $x = 1/3$  або  $x = 1$ .

При  $x = 1/3$  похідна  $f'(x)$  змінює знак з додатного на від'ємний, а при  $x = 1$  — з від'ємного на додатний. Отже в точці  $x = 1/3$  досягається максимум, а в точці  $x = 1$  — мінімум.

Цей приклад можна розв'язати простішим способом, якщо обчислити другу похідну  $f''(x) = 6x - 4$ ;

$f''(1/3) = -2$ , тобто від'ємна, і при  $x = 1/3$  досягається максимум;

$f''(1) = 2$ , тобто додатна, і при  $x = 1$  досягається мінімум.

### 1.2 Класифікація математичних методів оптимізації

Існують різні принципи класифікації математичних методів оптимізації [2]. Найбільш цікавим принципом можна вважати класифікацію за математичним виглядом функції мети  $f(x)$  і функцій обмежень  $\varphi_j(x), \psi_\xi(x)$ , де  $j = 1, m$  і  $\xi = 1, p$ , рисунок 1.3.

1. *Статичними моделями оптимізації* називаються такі моделі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  є алгебраїчними або трансцендентними рівняннями, а розділ математичного програмування називається *статистичним програмуванням*.

2. *Динамічними моделями оптимізації* називають такі моделі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  є диференціальними або інтегро-диференціальними рівняннями. Ці моделі реалізуються *динамічним програмуванням*.

3. *Лінійними моделями* називаються такі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  є лінійними. Ці моделі визначають метод *лінійного програмування*.

4. *Квадратичними* називаються задачі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  визначаються квадратичними многочленами. Ці задачі розв'язуються методом *квадратичного програмування*.

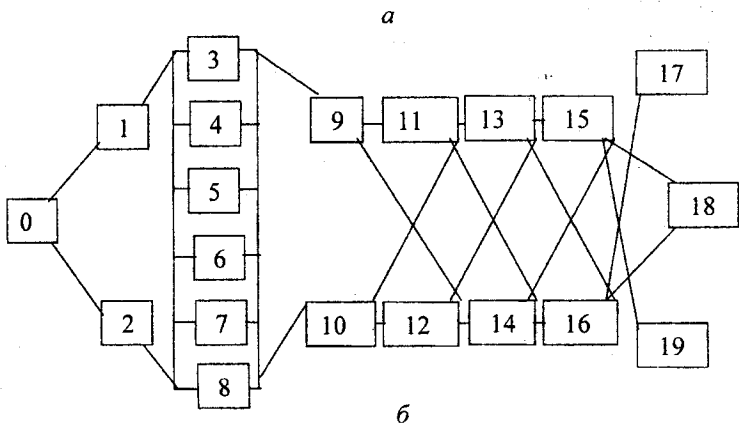
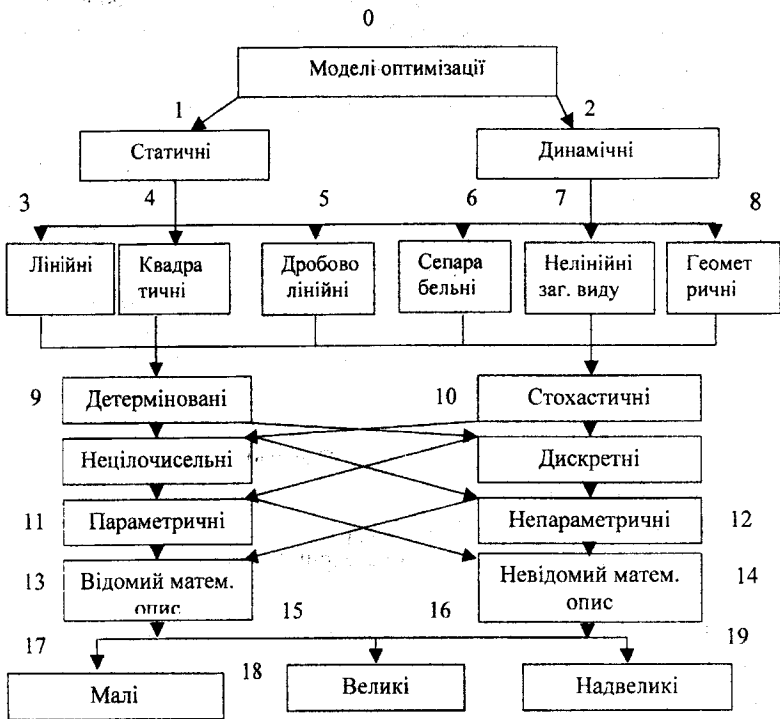


Рисунок 1.3 – Класифікація математичних методів оптимізації: а – класифікаційна схема; б – граф-схема моделей оптимізації



5. *Дробово-лінійними моделями оптимізації* називають моделі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  є дробово-лінійними: розрізняють дробово-лінійні моделі однорідні, в яких вільні члени дорівнюють нулю, інакше — неоднорідні. Метод, який реалізує ці моделі, називається *гіперболічним програмуванням*.

6. *Нелінійними моделями оптимізації* називаються моделі, в яких будь-яка з функцій  $f(x), \varphi_j(x), \psi_\xi(x)$  нелінійна. Ці моделі реалізуються *методом нелінійного програмування*.

7. *Дискретними моделями оптимізації* називаються моделі, в яких змінні  $\bar{x}$  всі або частково приймають цілочислові значення. Якщо змінні приймають значення 0 або 1, то оптимізація буде булева, що відноситься до *комбінаторних задач*.

8. *Сепарабельні моделі оптимізації* — це моделі оптимізації, в яких функція  $f(\bar{x}) = f_1(x_1) + f_2(x_2) + \dots + f_n(x_n)$ . Сепарабельні задачі оптимізації розв'язуються *сепарабельним програмуванням*.

9. *Геометричними моделями* називаються моделі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  інтерпретуються геометрично. Розв'язуються такі задачі *геометричним програмуванням*.

10. *Стохастичними моделями оптимізації* називаються моделі, в яких функції  $f(x), \varphi_j(x), \psi_\xi(x)$  змінюються випадково. Стохастичні задачі розв'язуються *стохастичним програмуванням* з урахуванням двох аспектів, в ситуації *ризик* і в ситуації *невизначеності*. Якщо числові характеристики випадкових величин відомі, то це ситуація *ризик*, задачі розв'язуються *стохастичним програмуванням*. Якщо числові характеристики випадкових величин невідомі, або відомі частково, то задачі розв'язуються в рамках *теорії ігор*.

11. Якщо вхідні дані, що входять до функцій  $f(x), \varphi_j(x), \psi_\xi(x)$ , являють собою відомі функції параметрів, то задачу відносять до *параметричних моделей оптимізації*.

12. До *малих моделей оптимізації* відносять моделі, вхідні дані яких разом з алгоритмом розв'язку вміщуються в робочому полі оперативної пам'яті ЕОМ.

13. До *великих моделей оптимізації* відносять моделі, вхідні дані яких разом з алгоритмом розв'язку вміщуються в робочому полі зовнішньої пам'яті ЕОМ.

Інакше задачі відносять до *надвеликих*.

14. *Робочим полем* називається частина пам'яті ЕОМ, що залишається після введення БСП, трансляторів та інших службових програм.

15. Будь-яку модель оптимізації, в якій відсутні величини, що змінюються випадково, називають *детермінованими моделями оптимізації*.

На практиці об'єкт оптимізації не завжди можна описати аналітично, в таких випадках задачі поділяються на задачі з аналітичним описом об'єкта і задачі без математичного опису.

Задача оптимізації може бути зображена одним з можливих шляхів в показаній на рисунку 1.3 граф-схемі, наприклад шлях (1, 3, 9, 11, 14, 15, 17) являє собою просту задачу лінійного програмування: нецілочислова, детермінована і т.д. Клас оптимізаційних задач — одна з гілок даної граф-схеми.

### 1.3 Векторні математичні моделі оптимізації

Моделі векторної оптимізації – це такі моделі, коли функції  $f(\bar{x})$ ,  $\varphi_j(\bar{x})$ ,  $\psi_\xi(\bar{x})$  подаються в вигляді векторів-функцій  $\overline{f(\bar{x})}$ ,  $\overline{\varphi_j(\bar{x})}$ ,  $\overline{\psi_\xi(\bar{x})}$  [3]. Векторні або багатокритеріальні моделі застосовуються в сфері планування і організації в таких випадках:

- коли рішення визначають спільні дії кількох об'єктів, ефективність кожного з яких оцінюється окремим критерієм;
- коли якість рішень необхідно використовувати для кількох варіантів умов і для кожного варіанту вводиться своя оцінка;
- коли рішення оцінюється в динаміці, або поетапно, а для оцінки рішення на кожному етапі вводиться самостійний критерій;
- коли якість рішення необхідно оцінювати з декількох точок зору за окремими компонентами якості.

Для задач економіки характерні такі цільові функції: максимальний прибуток, максимальна рентабельність, максимальна кількість продукції, мінімальна собівартість, максимальний коефіцієнт використання обладнання, максимальна продуктивність праці та інші.

При побудові векторної моделі вся множина залежних параметрів розбивається на дві підмножини: множина цільових функцій і множина допустимих розв'язків.

Цільові функції в векторному зображенні критеріїв можуть зводитись до пошуку максимуму чи мінімуму. Оптимальні розв'язки для різних критеріїв можуть не збігатися. У випадку жорсткого пріоритету критеріїв спочатку проводиться послідовна оптимізація критеріїв, при цьому розв'язується ряд однокритеріальних моделей, після чого вибирається кращий розв'язок. Це можна зробити, якщо кожному надати визначену вагу, розв'язувати задачу оптимізації на двох рівнях ієрархії: на нижньому рівні визначаються оптимальні розв'язки та значення локальних критеріїв, на верхньому формується узагальнений критерій, що визначає компромісний розв'язок з урахуванням ваг. При цьому враховується, що узагальнений (компромісний) критерій виводиться із частинних (локальних) критеріїв, які мають різну розмірність і є величинами різного

походження. Для цього виконується нормалізація частинних критеріїв, що характеризують якісні показники складних систем [4].

Нехай в задачі

$$\min(\max) f(x), \quad x \in D \subset E^*, \quad (1.2)$$

де  $E^*$  — евклідовий простір розмірності;

$D$  — множина допустимих розв'язків,

і задані всі локальні критерії  $\alpha_l$ , за звичай  $\sum_{l=1}^k \alpha_l = 1$ ,  $l = \overline{1, k}$ .

Потім всі локальні критерії мінімізуються і розв'язуються всі задачі

$$\min(\max) f_l(x), \quad x \in \alpha \subset E^*. \quad (1.3)$$

За норми локальних критеріїв беруться оптимальні (потенціальні) значення

$$H f_l(x) \min_D f_l(x). \quad (1.4)$$

Адитивний узагальнений критерій можна записати у вигляді

$$\min(\max) \left\{ F(x) = \sum_{l=1}^k \alpha_l \frac{f_l(x)}{f_l(x)}, \quad x \in D \subset E^*. \quad (1.5) \right.$$

При максимумі локальних критеріїв узагальнений критерій отримується аналогічно.

З урахуванням втрат оптимальності задачу на другому рівні можна поставити так: необхідно, щоб в точці компромісних розв'язків сума втрат оптимальності з урахуванням вагових коефіцієнтів по всіх локальних критеріях була мінімальною.

На першому етапі знаходиться оптимальний розв'язок локальних критеріїв

$$\min_D f_l(x) = f_l^{\min}. \quad (1.6)$$

В кожній з точок  $x_l^*$  розраховуються значення інших критеріїв. Для кожного з критеріїв вибирається найбільше його значення

$$f_l^{max} = \max_D f_l(x). \quad (1.7)$$

Максимальна втрата оптимальності для  $l$ -го локального критерію

$$\Delta f_l^{max} = f_l^{max}(x) - f_l^{min}(x). \quad (1.8)$$

Для довільної точки  $x$  величина втрати оптимальності

$$\Delta f_l(x) = \Delta f_l(x) - f_l^{min}(x). \quad (1.9)$$

З урахуванням вагових коефіцієнтів компромісний критерій

$$F_l(x) = \alpha_1 \frac{\Delta f_1(x)}{f_1^{max} - f_1^{min}} + \dots + \alpha_k \frac{\Delta f_k(x)}{f_k^{max} - f_k^{min}},$$

або

$$\min \left\{ F_l(x) = \sum_{l=1}^k \alpha_l \frac{f_l(x) - f_l^{min}}{f_l^{max} - f_l^{min}}, \quad x \in D \subset E^*. \right. \quad (1.10)$$

Якщо всі локальні критерії були лінійні, то і критерій 1.10 зберігають лінійність.

Для формування квадратичних критеріїв необхідно, щоб сума відхилень локальних критеріїв від оптимального значення була мінімальною.

Критерій має вигляд

$$\min \left\{ F_l(x) = \sum_{l=1}^k \left[ \frac{f_l(x) - f_l^{\min}}{f_l^{\max} - f_l^{\min}} \right]^2 \right\}, \quad x \in D \subset E^*. \quad (1.11)$$

Для векторної лінійної моделі оптимізації за таким критерієм кінцевий розв'язок може бути отриманий всередині допустимої області, але якщо узагальнений критерій лінійний, то цього не буде.

#### 1.4 Моделі оптимізації методом дослідження операцій

Метод *дослідження операцій* є розділом економіко-математичної теорії прийняття оптимальних рішень. Дослідження операцій використовується при розв'язанні транспортних задач, задач теорії масового обслуговування і сіткового планування.

Під *операцією* розуміють захід або систему заходів, об'єднаних єдиним замислом і направлених на досягнення визначеної мети.

Дослідження операцій і планування починається з оцінки координат чи параметрів об'єктів оптимізації і закінчується обчисленням значень тих чи інших подій.

*Керування операцією* складається з комплексної дії на людей в процесі функціонування системи (адміністративні, економічні, правові, соціологічні, психологічні, військові та інші).

*Прийняття рішень* при цьому не входить в задачу дослідження операцій, кінцеве рішення приймає керівник організації з урахуванням даних, не врахованих при розрахунках.

*Планування операцій* дає керівнику допомогу в прийнятті кінцевого рішення і виступає в ролі "радника".

Дослідження операцій опирається на *експеримент і індукцію* (розмірковування від окремих факторів і положень до загальних висновків).

В задачі дослідження операцій використовуються детерміновані, ймовірнісні і індетерміновані системи.

До детермінованих систем відносяться системи, в яких процеси взаємозв'язані і в статичному режимі описуються алгебраїчними, а в динамічному — диференціальними чи інтегральними рівняннями.

До ймовірнісних (стохастичних) відносять такі системи, в яких відношення між входами і виходами визначається теорією ймовірностей і теорією надійності, теорією масового обслуговування.

До індетермінованих систем відносять системи, в яких включаються детерміновані підсистеми з випадковими зв'язками між ними.

Для кількісної оцінки ефективності операцій використовуються цільові функції і критерії. В дослідженні операцій критерії є

ймовірнісними, спочатку оцінюють їх математичне очікування, а потім їх максимум чи мінімум.

Для множини критеріїв, в залежності від важливості критеріїв, вводять різні вагові коефіцієнти шляхом опитування експертів.

В будь-яких кількісних методах дослідження операцій спочатку будується математична модель об'єкта, потім знаходиться максимум чи мінімум показника ефективності.

### 1.5 Моделі оптимізації задач обчислювальної математики

#### А) Розв'язання систем алгебраїчних та трансцендентних рівнянь

Припустимо, що існує система з  $n$  рівнянь відносно  $n$  координат

$$\begin{cases} \varphi_1(x_1, x_2, \dots, x_n) = 0, \\ \varphi_2(x_1, x_2, \dots, x_n) = 0, \\ \dots\dots\dots \\ \varphi_n(x_1, x_2, \dots, x_n) = 0. \end{cases} \quad (1.12)$$

Задачу пошуку коренів системи (1.12) можна сформулювати в вигляді такої оптимізаційної задачі:

$$\min \left\{ f(x) = \varphi_1(x) + \dots + \varphi_n(x) = \sum_{i=1}^n \varphi_i(x) \right\}. \quad (1.13)$$

Якщо є можливість вказати границі зміни коренів, то модель оптимізації буде

$$\min \left\{ f(x) = \sum_{i=1}^n \varphi_i(x), \quad x \in X \in D \subset R^n \right\}, \quad (1.14)$$

де

$$D = (x_i^0 \leq x_i \leq x_i^{00}).$$

#### Б) Задача про найкраще наближення

Ця задача є класичною в статистиці і в плануванні експерименту, а також там, де стоїть питання апроксимації чи екстраполяції функцій однієї або багатьох змінних.

Нехай  $f(x)$  задана графічно або таблицею. Необхідно отримати аналітичний вираз  $f(x) = P(x)$ , де поліном

$$P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n. \quad (1.15)$$

Очевидно, задача найкращого наближення  $f(x)$  зводиться до знаходження таких коефіцієнтів полінома, при яких заштрихована область на рис. 1.4 була б мінімальною. Функція мети в моделі оптимізації задачі найкращого наближення може бути різною, наприклад потрібно знайти

$$\min f_1(x) = \max [f(x) - P(x)], \quad (1.16)$$

де  $x \in [x_0, x_n]$

$$\min f_2(x) = \int_{x_0}^{x_n} [f(x) - P(x)] dx. \quad (1.17)$$

Обмеженнями можуть бути :

– границі зміни апріорно відомих коефіцієнтів

$$a_i^0 \leq a_i \leq a_i^{00}, \quad i \in n;$$

– значення функції в деяких точках

$$P(x_k, \bar{a}) = f(x_k) = A_k;$$

– наявність вигинів в графіку функції.

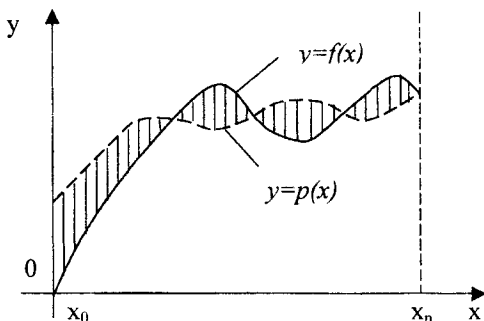


Рисунок 1.4 – Апроксимація функції мети в задачі оптимізації

## 1.6 Економіко-математичні моделі

### А) Задача про оптимальні призначення

Припустимо, що існує  $m$  робіт і  $n$  робітників (люди, станки, машини). Кожному  $i$ -му робітнику, назначеному на  $j$ -ту роботу поставим в відповідність деякий коефіцієнт ефективності  $C_{ij}$ . Потрібно визначити таке призначення робітників на роботу, при якому сумарний ефект від призначення був би максимальним.

Нехай  $X = \|x_{ij}\|_n^m$  — матриця призначень,

де

$$x_{ij} = \begin{cases} 0, & \text{якщо } i\text{-ий робітник не призначається на } j\text{-ту роботу} \\ 1, & \text{якщо призначається.} \end{cases}$$

Можливі два варіанти:

**Варіант 1.** На одну роботу може бути призначений тільки один робітник, і один робітник може виконувати тільки одну роботу.

Модель оптимізації в цьому випадку буде такою

$$f(x) = \sum_{i=1}^n \sum_{j=1}^m C_{ij} \cdot x_{ij} \rightarrow \max, \quad (1.18)$$

якщо  $\sum_{j=1}^m x_{ij} = 1, \quad \sum_{i=1}^n x_{ij} = 1.$

Дана модель є статистичною, лінійною, детермінованою, дискретною, непараметричною з відомим математичним описом.

**Варіант 2.** На одну роботу може бути призначено декілька робітників, або декілька робіт виконує один робітник. Якщо позначити через  $M_i$  — допустиме число робіт, а через  $N_j$  — допустиме число робітників, що виконують одну роботу, то отримаємо розподіл робітників по роботах з урахуванням трудомісткості робіт і нормативів на одного робітника. Подібна задача актуальна при розподілі річного кафедрального навантаження у ВНЗі.



Якщо  $a_i^*$  — допустима максимальна норма на одного  $i$ -го робітника,  $a_j^{**}$  — допустима максимальна норма на одного  $j$ -го робітника,  $b_j$  — трудомісткість  $j$ -ї роботи (в годинах), то модель оптимізації має вигляд

$$f(x) = \sum_{i=1}^n \sum_{j=1}^m C_{ij} X_{ij} \rightarrow \max, \quad (1.19)$$

якщо  $a_i^* = \sum_{j=1}^m x_{ij} b_j \leq a_i^{**}$  і  $\sum_{i=1}^n x_{ij} \leq N_j$ .

### Б) Транспортні задачі

Припустимо, що маємо  $n$  - пунктів виробництва і  $m$  - пунктів споживання. Вартість перевезення одиниці продукції з  $i$ -го пункту виробництва в  $j$ -й пункт споживання  $c_{ij}$ . Необхідно скласти такий план перевезень  $x = \|x_{ij}\|_n^m$ , щоб загальна вартість всіх перевезень була мінімальною, і вся вироблена продукція була доставлена в пункти споживання.

Нехай  $a_i$  - об'єм виробництва в  $i$ -му пункті, і  $b_j$  - об'єм споживання в  $j$  пункті.

**Варіант 1.** Припустимо, що вся вироблена продукція споживається повністю.

Модель оптимізації буде мати вигляд

$$f(x) = \sum_{i=1}^n \sum_{j=1}^m C_{ij} X_{ij} \rightarrow \min, \quad (1.20)$$

$$\text{якщо } \sum_{j=1}^m X_{ij} = a_i, \text{ і } \sum_{i=1}^n X_{ij} = b_j.$$

Якщо  $\sum_{i=1}^n a_i = \sum_{j=1}^m b_j$ , то така транспортна задача називається закритою, якщо  $\sum_{i=1}^n a_i > \sum_{j=1}^m b_j$ , то — відкритою. Будь-яка відкрита транспортна модель легко зводиться до закритої за допомогою введення

фіктивного пункту виробництва або фіктивного пункту споживання, куди планується вивести різницю між  $\sum_{i=1}^n a_i$  і  $\sum_{j=1}^m b_j$ .

**Варіант 2.** Якщо маємо обмежену пропускну спроможність, то транспортна задача формулюється таким чином.

З пункту  $i$  в пункт  $j$  не можна перевозити більше продукції, ніж  $d_{ij}$ .

Модель оптимізації буде

$$f(x) = \sum_{i=1}^n \sum_{j=1}^m C_{ij} X_{ij} \rightarrow \min, \quad (1.21)$$

якщо

$$\begin{cases} 0 \leq X_{ij} \leq d_{ij}, \\ \sum_{j=1}^m X_{ij} = a_i, \\ \sum_{i=1}^n X_{ij} = b_j, \end{cases} \quad (1.22)$$

при цьому повинно бути  $d_{ij} > 0$ , і

$$\begin{cases} \sum_{j=1}^m d_{ij} \geq a_i \\ \sum_{i=1}^n d_{ij} \geq b_j, \end{cases} \quad (1.23)$$

## 1.7 Задачі оптимізації в технічному проектуванні

Для простоти розглянемо розрахунок стабілізатора на кремнієвому стабілітроні.

Необхідно забезпечити підтримку напруги  $U_n$  на навантаженні  $R_n$  в схемі, зображеній на рис. 1.5, в межах  $U_1 \leq U_n \leq U_2$  за умов: вхідна напруга  $E$  може змінюватися в межах  $E_1 \leq E \leq E_2$ , де  $E_1 = k E_2$ , опір навантаження  $R_n$  змінюється в межах  $R_n = R_n \pm \Delta R_n$ .

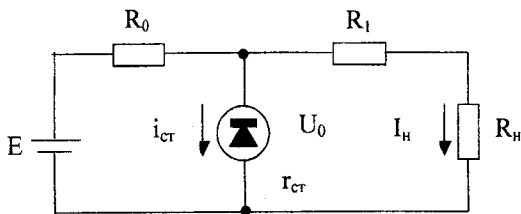


Рисунок 1.5 – Схема стабілізатора на кремнієвому стабілітроні

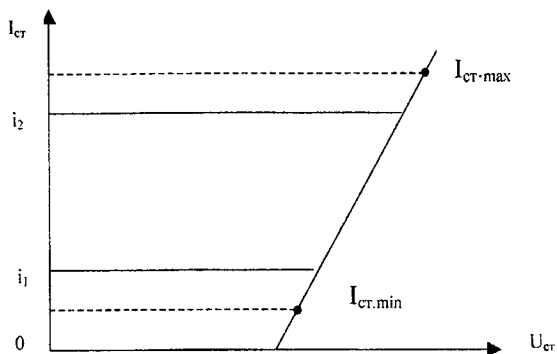


Рисунок 1.6 – Ідеалізована вольт-амперна характеристика стабілітрона

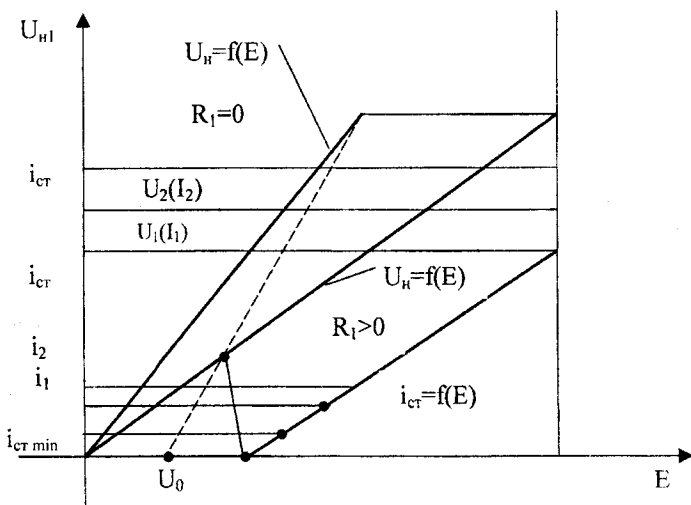


Рисунок 1.7 – Ідеалізована характеристика вхід-вихід схеми

На рис.1.6 зображена ідеалізована вольт-амперна характеристика стабілітрона, на рис.1.7 – ідеалізована характеристика вхід-вихід схеми. З урахуванням цих характеристик і закону Кірхгофа, робота схеми може бути описана рівнянням для двох режимів:

$$\begin{aligned} U_0 + r_{cm} i_2 &= U_2 + R_1 \cdot I_2, & i_1 &> i_2, \\ U_0 + r_{cm} i_1 &= U_1 + R_2 \cdot I_1, & i_1 &\geq I_{c \min}, \\ U_0 \cdot i_2 + r_{cm} i_2^2 &\leq P_{cm} & i_2 &< I_{c \max}, \end{aligned} \quad (1.24)$$

де:  $U_0$  - напруга стабілізації;

$r_{ct}$  - диференційований опір стабілітрона;

$i_1$  - мінімальний робочий струм стабілітрона;

$i_2$  - максимальний робочий струм стабілітрона;

$P_{ct}$  - допустима потужність розсіювання на стабілітроні;

$k$  - відношення мінімальної вхідної напруги до максимальної;

$I_1$  - мінімально допустимий струм навантаження;

$I_2$  - максимально допустимий струм навантаження;

За цільову функцію можна вибрати потужність, розсіювану в схемі, яку необхідно мінімізувати

$$P = U_0 \cdot i_2 + r_{cm} \cdot i_2^2 + R_1 \cdot I_2^2 + R_0 \cdot (I_2 + i_2)^2. \quad (1.25)$$

В моделі оптимізації незалежними змінними є параметри схеми  $R_0$ ,  $R_1$ ,  $U_0$ ,  $r_{ct}$ . Модель лінійна відносно цих змінних. Змінні  $R_0$  і  $R_1$  в моделі можуть приймати дискретні значення.

Модель дуже проста за рахунок використання ідеалізованих характеристик. При реальних характеристиках з'являються нелінійні залежності, які ускладнюють визначення оптимальних параметрів.

### 1.8 Військові моделі оптимізації

Необхідно визначити, в якій кількості треба виробити зброю різних класів, щоб забезпечити максимальну ефективність зброї для заданих ресурсів і обмежених витрат.

Припустимо, що кожен комплекс діє по своїй групі цілей. Позначимо:

$x_i$  – число комплексів зброї  $i$ -го типу;

$P_i$  – середнє число цілей, уражених одним засобом  $i$ -го типу (вірогідність ураження цілі  $i$ -м комплексом);

$R_i$  – вартість одного комплексу  $i$ -го типу;

$b_j$  – обмеження по  $j$ -му фактору;  
 $a_{ji}$  – кількість одиниць  $j$ -го фактору, сировина, обладнання, робоча сила і т.д., необхідні для виробництва одного комплексу зброї  $i$ -го типу;  
 $b$  – витрати на всю систему озброєння.  
 Спрощену модель оптимізації можна зобразити у вигляді

$$\max \left\{ f(\bar{x}) = \sum_{i=1}^n P_i x_i, \quad \bar{x} \in D \subset E^n \right\}, \quad (1.26)$$

де

$$D = \left\{ \begin{array}{l} \sum_{i=1}^n R_i x_i \leq b, \\ \sum_{i=1}^n a_{ji} x_i \leq b_j. \end{array} \right. \quad (1.27)$$

Тут максимізації підлягає математичне очікування числа уражених цілей. Цю модель можна суттєво розширити, якщо врахувати умови зберігання, експлуатацію комплексів, обмеження по номенклатурі у вигляді:

$$x_i^* \leq x_i \leq x_i^{**}.$$

Модель оптимізації буде більш реальною, якщо врахувати точні статистичні характеристики  $P_i$ , що визначаються на полігонах в процесі випробування зброї.

### 1.9 Ціна математичної моделі

Під ціною математичної моделі в широкому змісті розуміється основна сукупність кількісних даних, що характеризують оптимізаційну модель

$$ЦM = f(\bar{W}), \quad \bar{W} \in W, \quad (1.28)$$

де  $\bar{W}$  — вектор узагальнених витрат.

Мінімальним набором вектора затрат  $\bar{W}$  потрібно вважати трійку чисел:

$T_1$  — час обчислення функції моделі оптимізації в одній точці;

$V_0$  — об'єм пам'яті під вихідні дані моделі;

$T_2$  — час введення і перевірки на достовірність початкових даних моделі.

Тоді

$$ЦМ = f(T_1, V_0, T_2). \quad (1.29)$$

В якості інших компонент (координат) вектора можуть бути:

$T_3$  — час підготовки початкових даних;

$T_4$  — час вибірки і перекомпонування початкових даних із банку даних тощо.

При оцінці координат  $T_1, V_0, T_2$  всі операції арифметичні і логічні можна привести до однієї операції, наприклад до операції додавання з допомогою перехідних коефіцієнтів. Тобто

$$t_{ум} = K_{ум} \cdot t_{с л}, \quad \text{звідки} \quad K_{ум} = \frac{t_{ум}}{t_{с л}} \quad (1.30)$$

$$t_{\partial} = K_{\partial} \cdot t_{с л}, \quad K_{\partial} = \frac{t_{\partial}}{t_{с л}} \quad (1.31)$$

Довжини операцій  $t_{ум}, t_{\partial}, t_{л}, t_{с л}$  приведені в технічних паспортах конкретних ЕОМ.

## Контрольні запитання і завдання

1. Покажіть математично і графічно у чому різниця локального і глобального оптимумів.
2. Оптимізувати функцію  $f(x) = -x^3 + 3x^2 + 9x + 10$  на інтервалі  $-2 \leq x \leq 4$ . Знайти глобальний максимум.
3. Які моделі оптимізації називаються статичними?
4. Які моделі оптимізації називаються динамічними?
5. Які моделі оптимізації називаються лінійними, нелінійними та квадратичними?
6. Які моделі оптимізації називаються сепарабельними, дискретними та стохастичними?
7. Які моделі оптимізації називаються векторними?
8. Дослідити функцію на опуклість та вгнутість  $f(x) = (2x + 1)^2(x - 4)$ .
9. В чому сутність транспортної задачі. Зробіть постановку задачі.
10. В чому сутність задачі управління запасами. Зробіть постановку задачі.

11. Мінімізувати функцію  $f(x) = 2x^2 + \frac{16}{x}$  на відрізку  $1 \leq x \leq 5$ .
12. Як перевірити чи є функція опуклою або вгнутою?
13. Що таке точка перегину і як вона визначається математично?
14. Які моделі оптимізації називаються дробово-лінійними? Що таке гіперболічне програмування?
15. Що таке критерій оптимізації? Дайте математичне обґрунтування.
16. Дослідити та максимізувати функцію

$$f(x) = 5x^6 - 36x^5 + \left(\frac{165}{2}\right) \cdot x^4 - 60x^3 + 36.$$

17. Які економіко-математичні моделі оптимізації ви знаєте? Наведіть приклади.
18. В результаті експерименту встановлено, що траєкторія руху космічного тіла описується таким рівнянням

$$f(x) = 4x^3 - 3x^2 + 2x + e^x.$$

Дослідити дану функцію на опуклість і вгнутість. Визначити глобальний оптимум.

19. Встановити області, в яких функція  $f(x) = e^{-x^2}$  опукла або вгнута. Знайти глобальний максимум та глобальний мінімум в цій функції.
20. Що таке градієнт і антиградієнт?
21. Що таке сідлова точка? Яку роль відіграє рішення задачі про сідлову точку в умовах оптимізації.
22. Встановити які з наступних функцій є опуклими або вгнутими:

$$f(x) = e^x, f(x) = e^{-x}, f(x) = \frac{1}{x^2}, f(x) = x + \log x,$$

якщо  $x > 0$ ,  $f(x) = x^{2k}$ , де  $k$  – ціле число.

## 2 АНАЛІЗ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

### 2.1 Мінімум і максимум функцій за відсутності обмежень

Основною задачею математичного аналізу оптимізаційних моделей є задача визначення, до якого типу відноситься модель оптимізації: однокстремальної або багатокстремальної. Існування локального або глобального максимуму або мінімуму повністю визначається властивостями цільової функції і множиною допустимих рішень.

Для функції цілі можна встановити її вид, розмірність, затрати пам'яті ЕОМ, диференційованість, опуклість і т. ін. Для множини допустимих рішень – опукла множина або не опукла, однозв'язна або багатозв'язна, наявність неіснуючих обмежень. Все це дозволить вибрати найкращий метод оптимізації і визначити можливість його реалізації на ЕОМ.

#### А) Абсолютний max і min функції за відсутності обмежень

Функція  $f(\bar{x})$ , визначена на замкнутій множині  $X \subset R^n$ , досягає на ній абсолютного max в точці  $x_1$ , якщо  $f(\bar{x}) \leq f(x_1)$ .

Функція  $f(\bar{x})$ , визначена на замкнутій множині  $X \subset E^n$ , досягає на ній абсолютного min в точці  $\bar{x}^* \in X$ , якщо  $f(\bar{x}) \leq f(\bar{x}^*)$ .

Точка  $x_1$  на рис.2.1 є точкою абсолютного max, глобального max(min).

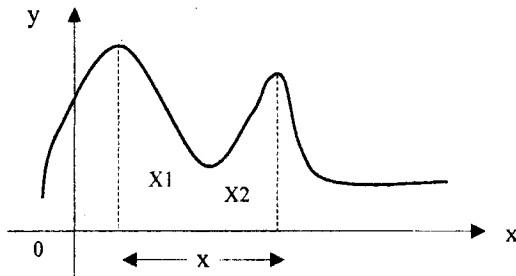


Рисунок 2.1 – Абсолютний (глобальний) максимум функції  $f(x)$

Якщо замкнута множина  $X$  обмежена, а  $f(\bar{x})$  неперервна в цій множині, то абсолютний max(min) на  $X$  досягається в одній з кількох точок.

Якщо множина  $X$  не є обмеженою, то абсолютного max функції  $f(\bar{x})$  можна не досягнути в жодній із точок  $X$  рис.2.2.



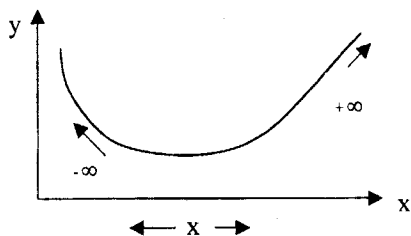


Рисунок 2.2 – Локальний та абсолютний максимуми функції  $f(x)$  за відсутності обмежень

Якщо множина  $X$  не є обмеженою, то для деяких функцій  $f(\bar{x})$  абсолютний  $\max$  може виявитися рівним граничному значенню функції  $f(\bar{x})$ , коли точка  $\bar{x}$  прямує до свого нескінченного значення будь-яким чином. Такий  $\max(\min)$  називається асимптотичним рис.2.3.

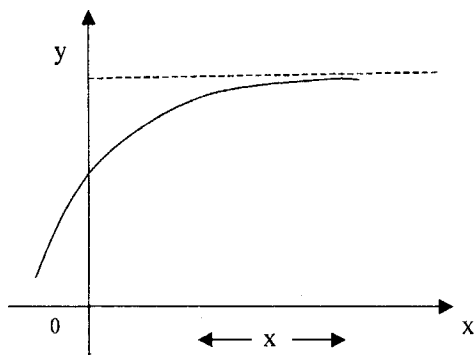


Рисунок 2.3 – Асимптотичний максимум функції  $f(x)$  за відсутності обмежень

### **Б) Сильний відносний $\max$ і $\min$**

Якщо  $f(\bar{x})$  визначена в усіх точках деякого околу  $\delta$  точки  $\bar{x}^* \in X \subset E^n$ , то в точці  $\bar{x}^*$  має сильний відносний  $\max$ , якщо знайдеться таке число  $\varepsilon$ ,  $0 < \varepsilon < \delta$ , що  $\forall \bar{x}$ , таких, що  $0 < |\bar{x} - \bar{x}^*| < \varepsilon$  буде  $f(\bar{x}) < f(\bar{x}^*)$  (рис.2.4).

Аналогічно можна визначити сильний відносний  $\min$ .

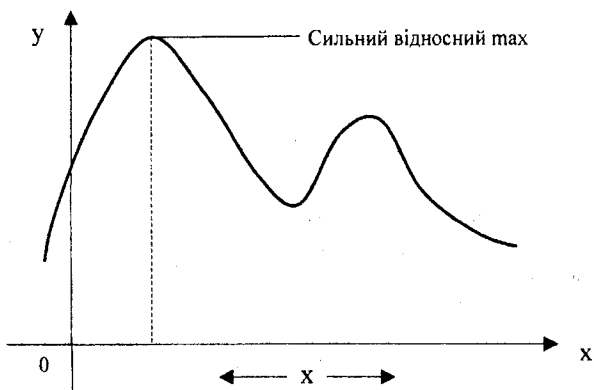


Рисунок 2.4 – Сильний відносний максимум функції  $f(x)$  за відсутності обмежень

### В) Слабкий відносний max і min

Якщо  $f(\bar{x})$  визначена в усіх точках деякого околу  $\delta$  точки  $\bar{x}^* \in X \subset E^n$ , то  $f(\bar{x})$  в точці  $\bar{x}^*$  має слабкий відносний max, якщо  $\exists \varepsilon$ ,  $0 < \varepsilon < \delta$ , що  $\forall \bar{x}$ , таких, що  $0 < |\bar{x} - \bar{x}^*| < \varepsilon$  буде  $f(\bar{x}) \leq f(\bar{x}^*)$ . Відповідним чином визначається і слабкий відносний min. І слабкий і сильний max(min) називається локальним (рис.2.5).

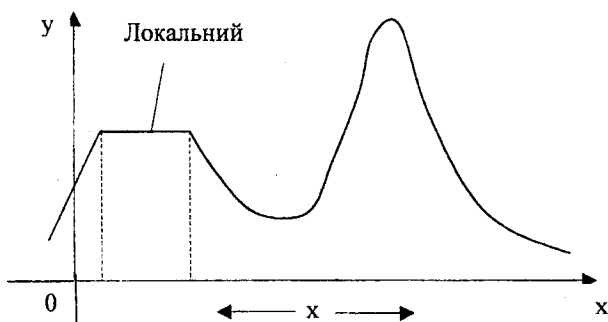


Рисунок 2.5 – Локальний максимум функції  $f(x)$  за відсутності обмежень

## 2.2 Стаціонарні точки функцій цілі і необхідні й достатні умови екстремуму

Точка  $\bar{x}^0 \in X \subset E^n$ , що є розв'язком системи рівнянь  $\frac{df(x)}{dx_i} = 0$ , називається стаціонарною точкою функції  $f(x)$ .

Точка  $\bar{x}^0$  є стаціонарна для  $f(\bar{x})$ , якщо  $\Delta f(\bar{x}^0) = 0$ . Система рівнянь, розв'язком якої є стаціонарна точка, в загальному випадку є нелінійною системою рівнянь, а тому стаціонарних точок може бути як завгодно багато. В сукупність стаціонарних точок входять точки перегинів, максимумів, мінімумів, сідлові точки та інші.

Нехай  $f(\bar{x}) \in C$ . Тоді для того, щоб в точці  $x^* \in X \subset R^n$  був тах або мінім необхідно, щоб  $\frac{df(x^*)}{dx_i} = 0$ , тобто  $\bar{x}^*$  повинна бути стаціонарною точкою.

Якщо  $f(x) \in C$  і  $f(x)$  має відносний тах або мінім в точці  $\bar{x}^*$ , то ця точка є розв'язком системи рівнянь  $\frac{df(x^*)}{dx_i} = 0$ .

### Висновки:

1. В точці  $\bar{x}^*$ , де  $f(\bar{x})$  має будь-який відносний тах або мінім,  $\frac{df(x^*)}{dx_i} = 0$ .
2. Градієнтом функції  $f(\bar{x})$  в точці  $\bar{x}^*$  є нульовий вектор  $\Delta f(\bar{x}^*) = 0$ .
3. Дотична площина до поверхні  $f(\bar{x})$  в точці  $\bar{x}^*$  має рівняння  $f(\bar{x}^*)$ , тобто дотична площина горизонтальна.

## 2.3 Дослідження цільових функцій $W_k, k=1, 2, \dots$ й множини

1. Функція множини  $W_1$ .

Функція  $f(\bar{x})$  на опуклій множині  $X \subset E^n$  називається опуклою (вниз), якщо  $\bar{x}_2 \in X$  і  $\lambda \in [0,1]$ ,

$$f[\lambda x_2 + (1-\lambda)x_1] \leq \lambda f(x_2) + (1-\lambda)f(x_1), \quad (2.1)$$

або

$$f[\lambda x_1 + (1 - \lambda)x_2] \leq \lambda f(x_1) + (1 - \lambda)f(x_2). \quad (2.2)$$

Функція  $f(\bar{x})$  називається вгнутою, якщо

$$f[\lambda x_2 + (1 - \lambda)x_1] \geq \lambda f(x_2) + (1 - \lambda)f(x_1). \quad (2.3)$$

Функція  $f(\bar{x})$  на опуклій множині  $X \subset R^n$  називається строго опуклою або строго вгнутою, якщо  $\bar{x}_1, \bar{x}_2 \in X$ , і  $\lambda \in [0, 1]$ , виконується відповідно

$$f[\lambda x_2 + (1 - \lambda)x_1] < \lambda f(x_2) + (1 - \lambda)f(x_1), \quad (2.4)$$

або

$$f[\lambda \bar{x}_2 + (1 - \lambda)\bar{x}_1] > \lambda f(\bar{x}_2) + (1 - \lambda)f(\bar{x}_1). \quad (2.5)$$

На рис.2.6 і 2.7 приведені геометричні моделі опуклої і вгнутої функції однієї змінної.

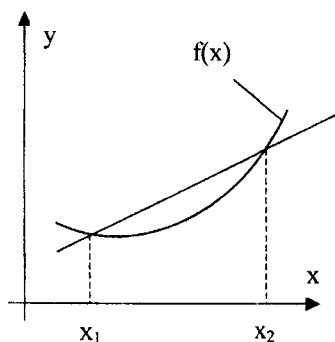


Рисунок 2.6 – Приклад вгнутої функції

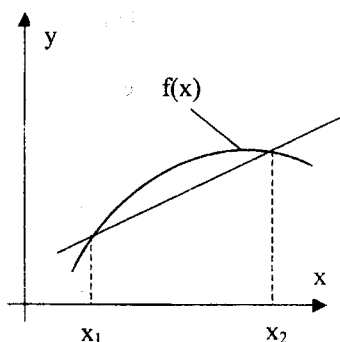


Рисунок 2.7 – Приклад опуклої функції

Введемо такі позначення:

$W_1^{Op}$  — множина опуклих функцій;

$W_1^{СОП}$  — множина строго опуклих функцій;

$W_1^{ВГ}$  — множина вгнутих функцій;

$W_1^{СВГ}$  — множина строго вгнутих функцій.

Очевидні наступні співвідношення:

Якщо функція  $f(\bar{x})$  опукла, то  $-f(\bar{x})$  вгнута функція і навпаки;

$$\begin{aligned}W_1^{СОП} &\subset W_1^{ОП} \\W_1^{СВГ} &\subset W_1^{ВГ} \\W_1 &= W_1^{ВГ} \cup W_1^{ОП}\end{aligned}\tag{2.6}$$

Лінійна функція  $f(\bar{x}) = C^T \bar{x}$  є опуклою і вгнутою.

Невід'ємна квадратична функція  $f(x) = x^T$  є опуклою функцією.

Від'ємна квадратична функція  $f(\bar{x}) = x^T$  є вгнутою функцією  $W_1^{ВГ}$ .

Додатно визначена квадратична функція  $f(\bar{x}) = x^T > 0$  належить  $W_1^{СОП}$ .

Від'ємно визначена квадратична функція  $f(\bar{x}) = x^T < 0$  належить  $W_1^{СВГ}$ .

### Теорема Зонтейдайка та інші теореми

Нехай  $f(x) \in W_1^{ОП}$  і існують всі часткові похідні другого порядку функції  $f(\bar{x})$ , що утворюють матрицю  $C(\bar{x})$ , тоді:

1.  $f(\bar{x}_2) - f(\bar{x}_1) \geq \Delta f(\bar{x}_1)^T (\bar{x}_2 - \bar{x}_1)$ ,
2.  $\Delta f(\bar{x} + \lambda \bar{S})^T \bar{S}$  — не спадна функція
3.  $C(\bar{x})$  невід'ємно визначена  $\bar{x} \in X^n$

#### Наслідки:

Якщо  $f(\bar{x}) \in W_1^{ОП}$ , і  $f(\bar{x}_2) < f(\bar{x}_1)$ , то  $\nabla f(\bar{x}_1)^T - (\bar{x}_2 - \bar{x}_1) < 0$ ;

Якщо  $f(x) \in W_1^{ОП}$ , і  $\nabla f(\bar{x}_1)^T (\bar{x}_2 - \bar{x}_1) \geq 0$ , то  $f(\bar{x})$  досягає мінімуму в точці  $\bar{x}_1$ .

Ця і наступна теореми наводяться з виведеннями в [3].

Функціональна композиція функцій класу  $W_k$  наведена в таблиці 2.1.

Логічна композиція функцій класу  $W_k$  наведена в таблицях 2.2 і 2.3.

## 2.4 Методи пошуку мінімуму і максимуму функцій за наявності обмежень

### 2.4.1 Метод множників Лагранжа

Математичні моделі задач оптимізації в більшості випадків зводяться до задач мінімізації або максимізації цільової функції на деякій множині допустимих розв'язків, що задається системою обмежень.

Класичний підхід при розв'язанні задач оптимізації можна зобразити в такому вигляді:

$$f(\bar{x}) \rightarrow \min(\max), \quad (2.7)$$

$$\text{якщо} \quad \varphi_j(x) = \varphi_j^0, \quad j = \overline{1, m}, \quad m < n. \quad (2.8)$$

Необхідно відшукати точку  $\bar{x}^*$ , яка забезпечить абсолютний мінімум або максимум функції  $f(\bar{x})$  на множині  $X \subset E^n$ , що задовольняє умови (2.8). Припускається, що  $f \in C$  і  $\varphi_j \in C$ ,  $j = \overline{1, m}$  на  $E^n$ . Множину точок  $\bar{x}$ , які задовольняють умову (2.8), позначимо через  $D$ .

Суть методу множників Лагранжа полягає в наступному [1].

Для отримання необхідних умов мінімуму або максимуму функції  $f(\bar{x})$  складемо функцію Лагранжа

$$f(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \sum_{j=1}^m \lambda_j \cdot [\varphi_j^0 - \varphi_j(\bar{x})] \quad (2.9)$$

і порівняємо до нуля її частинні похідні за кожною із  $m+n$  змінних  $x_i$ ,  $i = \overline{1, n}$ ,  $n\lambda_j$ ,  $j = \overline{1, m}$ .

**Таблиця 2.1** – Функціональна композиція функцій класу  $W_k$

$\varphi(y)$		$\psi_j(\bar{x})$	$f(x) = \varphi(\psi(\bar{x}))$	$\varphi(y)$		$\psi_j(x)$	$f(x) = \varphi(\psi(x))$
Зростаюча	ВГ	ВГ	ВГ	Спадаюча	ВГ	ВГ	ВГ
	ОП	ОП	ОП		ОП	ОП	ОП
	(Я)КВГ	ВГ	(Я)КВГ		(Я)КВГ	ВГ	(Я)КВГ
	(Я)КОП	ОП	(Я)КОП		(Я)КОП	ОП	(Я)КОП
Строго зростаюч	ВГ	СВГ	СВГ	Строго спадуюч	ВГ	СВГ	СВГ
	ОП	СОП	СОП		ОП	СОП	СОП
	КВГ	СВГ	СКВГ		КВГ	СВГ	СКВГ
	КОП	СОП	СКОП		КОП	СОП	СКОП

**Таблиця 2.2** – Логічна композиція функцій класу  $W_k$

$\psi_1(x)$	$\psi_2(x)$	$\psi_3(x)$	$f(x) = \frac{1}{\psi_1(x)}$	$f(x) = \psi_1(x) \cdot \psi_2(x)$	$f(x) = \frac{\psi_1(x)}{\psi_3(x)}$
ВГ $\geq 0$	ВГ $\geq 0$	ОП $> 0$	ОП	ЯКВГ	ЯКВГ
ВГ $\geq 0$	ОП $\leq 0$	ВГ $< 0$	ВГ	ЯКОП	ЯКОП
ВГ $\leq 0$	—	ОП $< 0$	ЯКВГ	—	ЯКОП
ВГ $\leq 0$	—	ВГ $> 0$	ЯКОП	—	ЯКВГ
СВГ $> 0$	ВГ $> 0$	ОП $> 0$	СОП	СКВГ	СКВГ
СВГ $> 0$	ОП $< 0$	ВГ $< 0$	СОП	СКОП	СКОП
СВГ $< 0$	—	ОП $< 0$	СКОП	—	СКОП
СВГ $< 0$	—	ВГ $> 0$	СКОП	—	СКВГ
ОП $\geq 0$	—	ВГ $> 0$	ЯКВГ	—	ЯКОП
ОП $\geq 0$	—	ОП $< 0$	ЯКВГ	—	ЯКВГ
ОП $\leq 0$	ОП $\leq 0$	ВГ $< 0$	ВГ	ЯКВГ	ЯКВГ
ОП $\leq 0$	ВГ $\geq 0$	ОП $> 0$	ВГ	ЯКОП	ЯКОП
СОП $> 0$	—	ВГ $> 0$	СКВГ	—	СКВГ
СОП $> 0$	—	ОП $< 0$	СКВГ	—	СКОП
СОП $< 0$	ОП $< 0$	ВГ $< 0$	СВГ	СКВГ	СКВГ
СОП $< 0$	ВГ $> 0$	ОП $> 0$	СВГ	СКОП	СКОП

**Таблиця 2.3** – Логічна композиція функцій класу  $W_k$

$\varphi(\bar{\zeta})$	$f(x) = \varphi(\psi(x))$	$\varphi(\bar{\zeta})$	$\psi(x)$	$f(x) = \varphi(\psi(x))$
Зростаюча	КВГ	Спадаюча	КВГ	КОП
	КОП		КОП	КВГ
Строго зростаюча	СКВГ	Строго спадуюча	СКВГ	СКВГ
	ЯКВГ		ЯКВГ	ЯКВГ
	СКОП		СКОП	СКОП
	ЯКОП		ЯКОП	ЯКОП

Отримаємо

$$\frac{dF}{dx_i} = \frac{df}{dx_i} - \sum_{j=1}^m \lambda_j \frac{d\varphi_j}{dx_i} = 0, \quad (2.10)$$

$$\frac{dF}{d\lambda_j} = \varphi_j^0 - \varphi_j(\bar{x}). \quad (2.11)$$

Отримання таким чином необхідних умов називається методом множників Лагранжа.

**Приклад.**

$$f(\bar{x}) = 4(x_1 - x_2 - 1)^2 + (x_1 + x_2 - 11)^2 \rightarrow \min,$$

$$\text{якщо } (x_1 - 3)^2 + (x_2 - 8)^2 = 9.$$

Складемо функцію Лагранжа

$$F(\bar{x}, \bar{\lambda}) = 4(x_1 - x_2 - 1)^2 + (x_1 + x_2 - 11)^2 + \lambda [9 - (x_1 - 3)^2 - (x_2 - 8)^2]$$

і взявши часткові похідні за всіма змінними, отримаємо систему необхідних умов

$$\frac{dF}{dx_1} = 8(x_1 - x_2 - 1) + 2(x_1 + x_2 - 11) - 2\lambda(x_1 - 3) = 0;$$

$$\frac{dF}{dx_2} = 8(x_1 - x_2 - 1) + 2(x_1 + x_2 - 11) - 2\lambda(x_2 - 8) = 0;$$

$$\frac{dF}{d\lambda} = 9 - (x_1 - 3)^2 - (x_2 - 8)^2 = 0.$$

Вирішивши цю систему, знайдемо, що  $f(\bar{x})$  досягає мінімуму в точці

$$\bar{x}_{10} = (5.121, 5.879), \text{ де } f(\bar{x}_{10}) = 12.35,$$

а максимуму в точці  $\bar{x}_{20} = (0.879, 10.12)$ , де  $f(\bar{x}_{20}) = 422$ .

#### 2.4.2 Теорема Куна-Таккера. Подвійність задачі

Теорема Куна-Таккера є узагальненням класичного методу множників Лагранжа для визначення екстремуму при наявності обмежень типу рівностей на випадок, коли з'являються обмеження типу нерівностей [1, 3].

Припустимо



$$\min \left\{ \frac{f(x)}{q_j(x)} \leq 0, x_i \geq 0 \right\}, \quad (2.12)$$

де  $f(\bar{x})$  і  $q_j(\bar{x})$  — опуклі функції. Тоді розв'язання задачі зводиться до мінімізації опуклої функції на опуклій допустимій множині

$$D = \left\{ \frac{f(x)}{q_j(x)} \leq 0, x_i \geq 0, i = 1, n \right\}. \quad (2.13)$$

Практично така задача є найбільш зрозумілим випадком в лінійному програмуванні, оскільки модель буде мати єдиний екстремум.

Функція Лагранжа для цієї задачі

$$F(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j q_j(x). \quad (2.14)$$

$$F(x, \lambda) = f(x) = \sum_{j=1}^m \lambda_j q_j(x). \quad (2.15)$$

Функція  $F(\bar{x}, \bar{\lambda})$  має сідлову точку  $(x^0, \lambda^0)$ , якщо

$$F(x^0, \lambda) \leq F(x^0, \lambda^0) \leq F(x, \lambda^0) \quad (2.16)$$

$\bar{x}$  з  $\varepsilon$ -околу  $\bar{x}^0$  і  $\bar{\lambda}$  із  $\varepsilon$ -околу  $\bar{\lambda}^0$ . Якщо нерівність (2.16) виконується для всіх  $\bar{x}$  і  $\bar{\lambda}$ , то  $F(\bar{x}, \bar{\lambda})$  має в  $(x^0, \lambda^0)$  глобальну сідлову точку. Теорему Куна-Таккера називають теоремою про сідлову точку, оскільки задача мінімізації функції  $f(\bar{x})$  в області  $D$  відповідає задачі про знаходження сідлової точки функції Лагранжа.

Теорема. Вектор  $\bar{x}^0$  тоді і тільки тоді є розв'язком задачі (2.12), коли існує вектор  $\lambda^0$  такий що

$$x^0 \geq 0, \quad \lambda^0 \geq 0; \quad (2.17)$$

$$F(\bar{x}^0, \lambda) \leq F(\bar{x}^0, \lambda^0) \leq F(\bar{x}, \lambda^0). \quad (2.18)$$

Ці положення є умовами достатності. Можна показати, що умови (2.17, 2.18) є також необхідними.

Якщо функції  $F$  і  $q_j$  є диференційовними, то умови (2.17, 2.18) можна записати у вигляді:

$$\frac{F(\bar{x}^0, \bar{\lambda})}{dx} \geq 0; \quad \frac{F(\bar{x}^0, \bar{\lambda})}{d\lambda_j} \leq 0; \quad (2.19)$$

$$x_i \frac{F(\bar{x}^0, \bar{\lambda})}{dx_i} = 0; \quad \lambda_j \frac{F(\bar{x}^0, \bar{\lambda})}{d\lambda_j} = 0. \quad (2.20)$$

Взагалі кажучи, якщо точка  $\bar{x}^0$  задовольняє умови Куна-Таккера, то для того щоб  $\bar{x}^0$  була розв'язком задачі  $\min_x \left\{ \frac{f(x)}{q_j(x)} \leq 0, \quad x \geq 0 \right\}$  достатньо, щоб цільова функція  $f$  була псевдоопукла, а функція  $q_j$  квазіопукла.

Для лінійних моделей в розглянутому випадку можна ввести поняття подвійних задач.

Якщо пряма задача записується у вигляді

$$\min_x \left\{ \frac{f(x)}{q_j(x)} \leq 0, \quad x \geq 0 \right\}, \quad (2.21)$$

то подвійна задача має вигляд

$$\max \left\{ F(\bar{x}, \bar{\lambda}) / \frac{f(\bar{x})}{q_j(\bar{x})} \leq 0, \quad \bar{\lambda} \geq 0 \right\}. \quad (2.22)$$

## 2.4.3 Метод штрафних функцій

### 2.4.3.1 Загальні поняття і положення

Метод штрафних функцій дозволяє звести початкову модель, складену із цільової функції і системи обмежень, до задачі відшукування абсолютного екстремуму деякої допоміжної (штрафної) функції. Цей

метод не оперує обмеженнями в явному вигляді і дуже ефективний в обчислювальному відношенні, особливо для моделей з нелінійною цільовою функцією і нелінійними обмеженнями [3]. Розв'язки при цьому виходять приблизними. Шляхом апроксимації нелінійностей можна добитися хорошого збігу розв'язків початкової задачі і допоміжної. Удалий вибір функцій штрафу залежить від обмежувальної умови: рівності чи нерівності.

Нехай є деяка модель

$$\min \left\{ \begin{array}{l} f(x) \\ q_j(x) = 0, \quad q_j(x) \leq 0, \quad j = m_1 + 1, \dots, m \end{array} \right\}. \quad (2.23)$$

Всі обмеження можна привести до вигляду, описаних в моделі. Використовуючи функції штрафу, можна побудувати допоміжну функцію штрафу

$$Y(x) = f(x) + k_0 \left[ \sum_{j=1}^{m_1} \Phi_j[\varphi_j(x)] + \sum_{j=m_1+1}^m \Phi_j[q_j(x)] \right] \rightarrow \min, \quad (2.24)$$

де:  $\Phi_j$  — функції штрафу, які накладаються за порушення обмежень

$$\Phi_j = \begin{cases} 0, & \text{якщо обмеження виконуються,} \\ \Phi_j & \text{якщо порушуються,} \end{cases}$$

$k_0$  — коефіцієнт штрафу, від вибору якого залежить ступінь збігу розв'язків допоміжної задачі з початковою.

#### 2.4.3.2 Функції штрафу для обмежень у вигляді нерівностей

За найпростішу функцію штрафу можна розглянути таку

$$\Phi^1[q_j(x)] = S_q[q_j(x)] \quad (2.25)$$

де

$$S_q[q_j(\bar{x})] = \begin{cases} 0, & \text{якщо } q_j(\bar{x}) \leq 0, \\ 1, & \text{якщо } q_j(\bar{x}) > 0, \end{cases}$$

В цьому випадку модель має вигляд як на рис.2.8 і

$$\min_x \left\{ f(\bar{x}) + K_0 \sum_{j=1}^m \Phi_j[q_j(x)] \right\}. \quad (2.26)$$

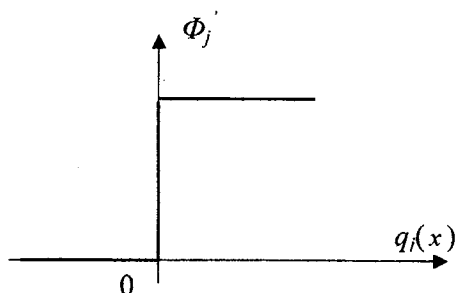


Рисунок 2.8 – Графічна модель функції штрафу з найпростішими обмеженнями

За  $K_0$  може бути вибрано  $|f(\bar{x})|$ .

Перевагою даної функції штрафу є її простота і обмеженість штрафу. Однак через її не диференційовність неможливо застосувати добре розроблені градієнтні методи.

Функцію  $\Phi_j$  можна апроксимувати яким-небудь аналітичним виразом, як показано, наприклад, на рис. 2.9.

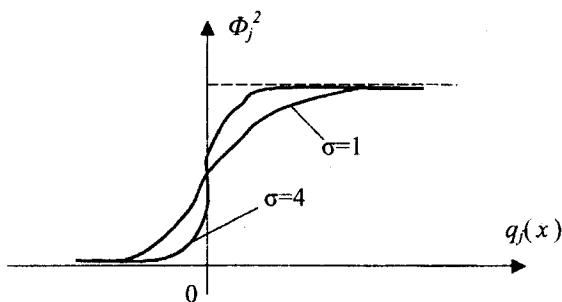


Рисунок 2.9 – Графічна модель функції штрафу, отримана за допомогою аналітичних методів апроксимації

З метою зменшення обчислень іноді застосовують лінійну  $q_j(\bar{x})$  функцію штрафу (рис.2.10).

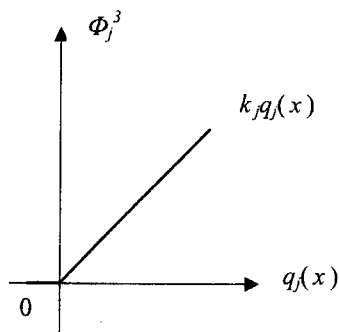


Рисунок 2.10 – Графічна модель лінійної функції штрафу

Математичні моделі зображаються відповідно формулами:

$$\Phi_j^2[q_j(x)] = \frac{1}{1 + e^{-\delta q_j(\bar{x})}}, \quad (2.27)$$

$$\Phi_j^3[q_j(x)] = K_j q_{jj}(x) S_q[q_j(x)] = \max\{0, q_j(x)\}, \quad (2.28)$$

$$\Phi_j^3[q_j(x)] = K_j \frac{1}{2} \{q_i(x) + q_j(x)\} \quad (2.29)$$

Застосування лінійної моделі обмежено через можливість розриву похідної в точці  $q_j(\bar{x}) = 0$ . На практиці частіше застосовується квадратична функція штрафу (рис.2.11)

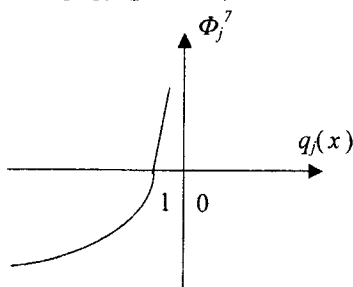


Рисунок 2.11 – Графічна модель квадратичної функції штрафу

$$\Phi_j^4[q_j(x)] = K_j q_j^2(x) S_q[q_j(x)], \text{ або } \Phi_j^4[q_j(x)] = K_j \left\{ \frac{1}{2} [q_i(x) + q_j(x)] \right\}^2 \quad (2.30)$$

Ця функція швидко зростає з порушенням обмеження, неперервна і має неперервні парні похідні (якщо ці властивості має  $q_j(\bar{x})$ ), але менш чутлива до малих порушень обмежень. Цю чутливість можна регулювати вибором коефіцієнта  $K_j$ .

Інколи замість другої беруть більш високу парну степінь

$$\Phi_j^5[q_j(x)] = K_j [q_i(x) + q_j(x)]^\alpha, \quad \alpha = 2, 4, 6, \dots \quad (2.31)$$

Недоліком при цьому є мала чутливість до малих порушень. Вільним від цього недоліку є функція штрафу, зображена на рис. 2.12.

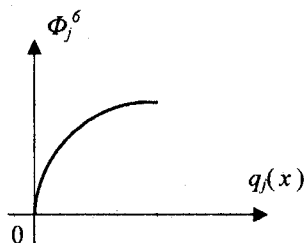


Рисунок 2.12 – Графічна модель функції штрафу вищих степенів

$$\Phi_j^6[q_j(x)] = K_j [q_i(x) + q_j(x)]^{2.3}. \quad (2.32)$$

В цьому випадку від початкової задачі можна перейти до наступної, рис. 2.13.

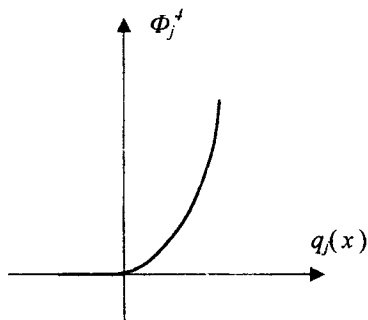


Рисунок 2.13 – Графічна модель функції штрафу з  $0 < \alpha < 1$

$$\min \left\{ f(\bar{x}) + \sum_{j=1}^m \Phi_j [q_j(\bar{x})] \right\} = \min \left\{ f(\bar{x}) - \sum_{j=1}^m K_j \ln[-q_j(\bar{x})] \right\}. \quad (2.33)$$

Задача при цьому зводиться до знаходження

$$\min \left[ f(\bar{x}) = \sum_{j=1}^m K_j \frac{1}{-q_j(\bar{x})} \right]. \quad (2.34)$$

При використанні функцій штрафу  $\Phi_j^7$  і  $\Phi_j^8$  необхідно, щоб початкова точка в алгоритмах пошуку знаходилась в допустимій області. Коефіцієнти  $K_j$  є при цьому додатковими змінними.

### 2.4.3.3 Штрафні функції для обмеження рівностей

Вплив обмежень рівностей виду  $\varphi_j(\bar{x}) = 0$  необхідно враховувати функціями штрафу, відмінними від нуля за наявності обмежень як в одну, так і в іншу сторону.

$$\Phi_j[\varphi_j(\bar{x})] = \begin{cases} 0, & \text{якщо } \varphi_j(\bar{x}) = 0, \\ \Phi_j, & \text{якщо } \varphi_j(\bar{x}) \neq 0. \end{cases} \quad (2.35)$$

$\Phi_j$  може бути const або var. Самою простою функцією штрафу є лінійна (рис. 2.14).

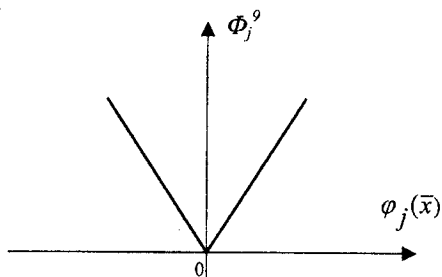


Рисунок 2.14 – Найпростіша лінійна функція штрафу

Така функція штрафу дозволяє перейти від задачі

$$\min \begin{cases} f(x) \\ \varphi_j(x) = 0 \end{cases} \quad (2.36)$$

до задачі на безумовний екстремум вигляду

$$\min_{x \in E^n} \left\{ f(x) + K_0 \sum_{j=1}^m k_j [\varphi_j(x)] \right\}. \quad (2.37)$$

Іншим прикладом може бути квадратична функція штрафу (рис. 2.15)

$$\Phi_j^{10}[\varphi_j(x)] = K_j [\varphi_j(x)]. \quad (2.38)$$

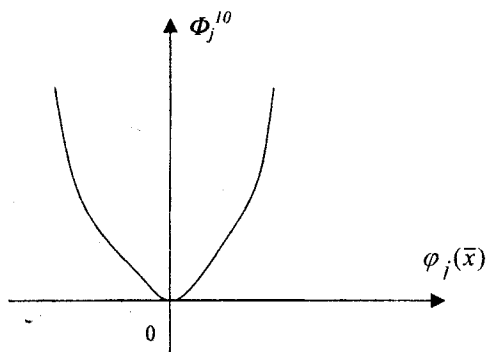


Рисунок 2.15 – Графічна модель квадратичної функції штрафу

Ця функція ефективніша лінійної і більш зручна для розрахунків в алгоритмах оптимізації, однак вона має меншу чутливість для малих порушень обмежень.

Можуть використовуватися функції штрафу (рис. 2.16)

$$\Phi_j^{11}[\varphi_j(x)] = K_j [\varphi_j(x)]^\alpha, \quad \alpha = 2, 4, 6, \dots \quad (2.39)$$

В цьому випадку величина штрафу росте значно швидше при порушенні обмежень, але чутливість до малих обмежень ще нижча.

Добру чутливість до таких порушень має функція штрафу (рис. 2.17)

$$\Phi_j^{12}[\varphi_j(\bar{x})] = K_j [\varphi_j(\bar{x})]^{2/3}. \quad (2.40)$$



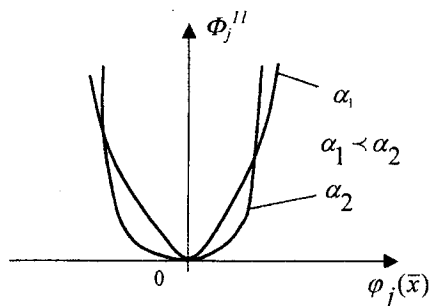


Рисунок 2.16 – Графічна модель функції штрафу вищих степенів з  $\alpha_2 > 2$

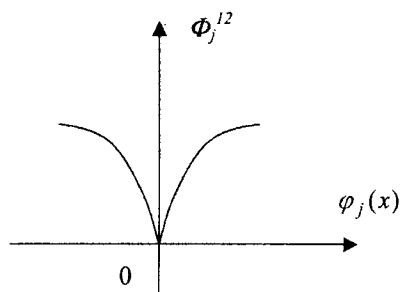


Рисунок 2.17 – Графічна модель функції штрафу з  $\alpha < 1$

Застосування функцій штрафу для обмежень рівностей зв'язане з значно більшими обчислювальними і алгоритмічними складностями, ніж застосування функцій штрафу для обмежень нерівностей.

Одним з найосновніших недоліків методу штрафних функцій є відсутність строгого методу вибору коефіцієнтів функцій штрафу.

Під час пошуку розв'язку вони можуть корегуватися. При застосуванні бар'єрних функцій оптимальний розв'язок початкової задачі не може бути знайдено без послідовних коефіцієнтів в функціях штрафу.

Другим недоліком можна вважати неможливість порівняння значень функції, що мінімізується, між ітераціями, внаслідок порушення обмежень.

## 2.5 Засоби пошуку екстремуму

Необхідною умовою екстремуму диференційовної функції кількох незалежних змінних [2]

$$f(x) = f(x_1, x_2, \dots, x_n) \quad (2.41)$$

є рівність нулю в точці екстремуму частинних похідних цієї функції:

$$\frac{df}{dx_1} = 0; \quad \frac{df}{dx_2} = 0; \dots \quad \frac{df}{dx_n} = 0. \quad (2.42)$$

Градiєнтом функції називається векторна величина

$$\text{grad } f = k_1 \frac{df}{dx_1} + k_2 \frac{df}{dx_2} + \dots + k_n \frac{df}{dx_n}, \quad (2.43)$$

де  $k_1, k_2, \dots, k_n$  — одиничні вектори осей, по яких відраховуються величини  $x_1, x_2, \dots, x_n$ .

В точці екстремуму

$$\text{grad } f = 0. \quad (2.44)$$

Задача пошуку екстремуму розбивається на дві частини:

1. Визначення градієнта або його відхилень від точки екстремуму;
2. Організація руху до точки екстремуму.

Це можна зробити безліччю способів.

### 2.5.1 Способи визначення градієнтів

#### *Спосіб синхронного детектування*

Спосіб може здійснюватися як при регулярних, так і при випадкових сигналах. Для регулярних сигналів основні величини  $x_1^*, x_2^*, \dots, x_m^*$ , що повільно змінюються, доповнюються невеликими гармонічними складовими, які мають різні частоти:

$$\begin{aligned} x_1 &= x_1^* + x_{10} \sin x_1 t; \\ x_2 &= x_2^* + x_{20} \sin x_2 t; \\ &\dots \dots \dots \\ x_m &= x_m^* + x_{m0} \sin x_m t; \end{aligned} \quad (2.45)$$

Вихідна величина  $f$  надходить на синхронні детектори (рис. 2.18), до яких підводяться також гармонічні складові

$$x_{10} \sin \omega_1 t, x_{20} \sin \omega_2 t, \dots, x_{m0} \sin \omega_m t.$$

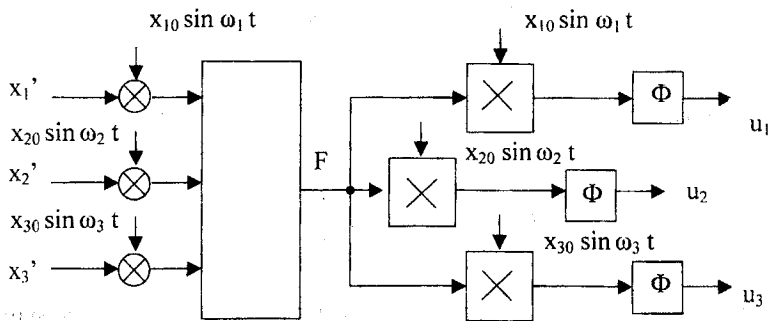


Рисунок 2.18 – Графічна модель синхронного детектування

Синхронні детектори виконують множення величини  $f$  на гармонічні складові і усереднення за часом отриманих добутків. Згідно з цим вихідні величини синхронних датчиків дорівнюють:

$$u_1 = x_{10} \overline{f \sin \omega_1 t}, \quad u_2 = x_{20} \overline{f \sin \omega_2 t}, \quad \dots, \quad u_m = x_{m0} \overline{f \sin \omega_m t}. \quad (2.46)$$

Неважко показати, що в стаціонарному режимі величини з точністю до малих вищого порядку пропорційні частинним похідним  $\frac{df}{dx_1}, \frac{df}{dx_2}, \dots, \frac{df}{dx_n}$  в точці  $x_1 = x_1^*, x_2 = x_2^*, \dots, x_n = x_n^*$  і визначають градієнт функції  $f$  в даній точці.

### Спосіб похідної за часом

Сутність способу заключається в диференціюванні функції  $f(\bar{x})$  за часом:

$$\frac{df}{dt} = \frac{df}{dx_1} \cdot \frac{dx_1}{dt} + \frac{df}{dx_2} \cdot \frac{dx_2}{dt} + \dots + \frac{df}{dx_n} \cdot \frac{dx_n}{dt}, \quad (2.47)$$

де частинні похідні відповідають поточним значенням координат  $x_1, x_2, \dots, x_n$ . Задаючи тим чи іншим способом величини  $\frac{dx_1}{dt}, \frac{dx_2}{dt}, \dots, \frac{dx_n}{dt}$  і,

змінюючи похідну  $\frac{df}{dt}$ , можна визначити компоненти градієнта  $\frac{df}{dx_1}, \frac{df}{dx_2}, \dots, \frac{df}{dx_n}$ .

Один з можливих варіантів полягає в почерговому послідовному задаванні за часом постійних швидкостей зміни регульованих величин (рис.2.19, 2.20). Робота схеми циклічна, час циклу  $\tau$  розбивається на  $m$  інтервалів  $\tau$ .

З виразу (2.47) видно, що в інтервалі, який розглядається:

$$\frac{df}{dt} = \pm \frac{df}{dx_i}, \quad i = \overline{1, n}. \quad (2.48)$$

### *Спосіб запам'ятовування екстремуму*

Вимушеним або автоколивальним рухом зображувана точка  $x_1, x_2, \dots, x_n$  переміщується в околі екстремуму  $x_1 = x_{1e}, x_2 = x_{2e}, \dots, x_n = x_{ne}$ . Кожного разу, коли в процесі такого руху функція  $f$  досягає екстремального значення  $f = f_e$ , воно фіксується спеціальним пристроєм.

Градієнт функції визначається з різниці між його поточним і екстремальним значеннями  $f - f_e$  з допомогою пошукових рухів.

Оскільки в точці екстремуму

$$\left( \frac{df}{dx_i} \right) = 0, \quad (2.49)$$

то з точністю до малих вищого порядку, можна показати, що

$$\frac{d(f - f_e)}{dx_i} = \frac{df}{dx_i}. \quad (2.50)$$

Вимірюючи різницю  $f - f_e$ , можна отримати  $\frac{df}{dx_i}$  (рис. 2.21).

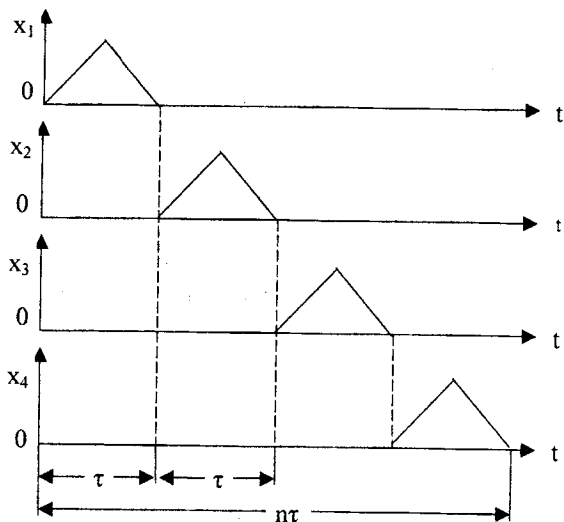


Рисунок 2.19 – Діаграма почергового послідовного задавання за часом постійних швидкостей зміни регульованих величин

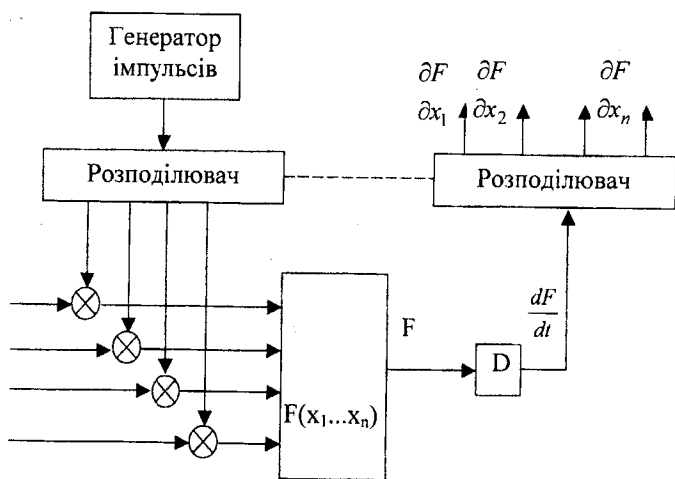


Рисунок 2.20 – Схема почергового послідовного задавання за часом постійних швидкостей зміни регульованих величин

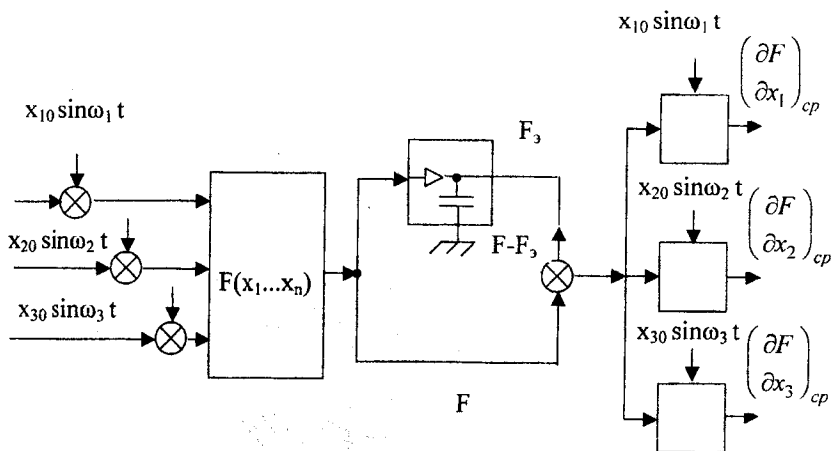


Рисунок 2.21 – Схема визначення  $\frac{df}{dx_i}$  по різниці між поточним і екстремальним значеннями  $f - f_e$ .

### 2.5.2 Методи організації процесу пошуку

Суть організації процесу пошуку полягає в забезпеченні руху системи до точки екстремуму. Для функцій кількох змінних можлива велика кількість методів, найбільш цікавими з яких є: метод Гауса-Зайделя, метод градієнта і метод найшвидшого спуску.

Сутність методу Гауса-Зайделя полягає в почерговій зміні координат  $x_1, x_2, \dots, x_n$  і визначенні частинних екстремумів

$$\frac{df}{dx_i} = 0,$$

при  $x_{i-1} = const$ .

Спочатку змінюється координата  $x_1$  в напрямку зменшення абсолютної величини компонента градієнта до  $\frac{df}{dx_1} = 0$  за постійних значень інших координат. Потім змінюється координата  $x_2$  за постійності інших і т.д. (лінія 1 на рис. 2.22).

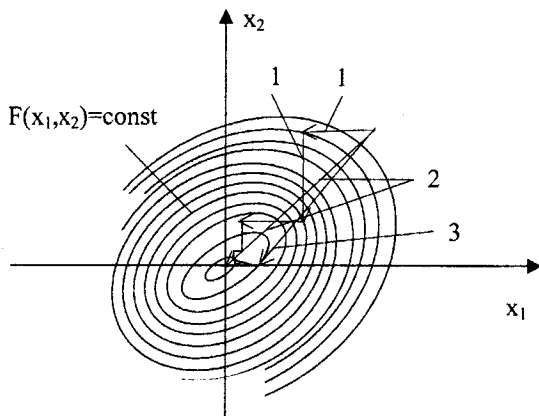


Рисунок 2.22 – Графічна модель почергової зміни координат  $x_1$ ,  $x_2$  для визначення частинних екстремумів

Сутність методу градієнта полягає в тому, що визначаються всі компоненти градієнта і забезпечується рух, зображуючи точки в напрямку, близькому до миттєвого напрямку градієнта, причому рух може бути неперервним або кроковим.

У випадку неперервного руху швидкості зміни координат встановлюються пропорційними відповідним компонентам миттєвого вектора градієнта:

$$\dot{x}_1 = a \frac{df}{dx_1}; \dot{x}_2 = a \frac{df}{dx_2}; \dots; \dot{x}_m = a \frac{df}{dx_m}, \quad (2.51)$$

де  $a = const > 0$  для випадку екстремуму-максимуму і  $a = const < 0$  для екстремуму-мінімуму.

При кроковому пошуку після обчислення градієнта виконується крок по осях координат, який відповідає компонентам градієнта:

$$\Delta x_i = a \frac{df}{dx_i}. \quad (2.52)$$

Далі знову визначається градієнт, виконується новий крок в напрямку вектора градієнта (лінія 2 на рис. 2.22).

Необхідно відзначити, що неперервний рух забезпечує більшу стійкість, ніж кроковий.

Сутність методу найшвидшого спуску полягає в визначенні вектора градієнта в початковій точці і організації руху в цьому напрямку і до тих пір, доки частинна похідна  $\frac{df}{dl}$  вздовж вказаного напрямку не перетвориться в нуль. Потім знову визначається напрям вектора градієнта і відбувається рух вздовж цього вектора до обернення похідної в нуль і т.д. (крива 3 рис. 2.22).

Для методу найшвидшого спуску характерний порівняно малий час досягнення екстремуму.

Можлива комбінація різноманітних методів руху до екстремуму, наприклад, при великих відхиленнях — метод найшвидшого спуску, при малих, для забезпечення більшої точності, — метод градієнта.

В наш час для розв'язання вище розглянутих задач оптимізації на ЕОМ використовують пакети програм MatLab, MathCad або табличний процесор Excel [17,18, 19].

## Контрольні запитання і завдання

1. Що таке абсолютний максимум (мінімум)?
2. Порівняйте поняття локальний, глобальний і асимптотичний мінімум (максимум) при відсутності обмежень. Наведіть приклади.
3. Що таке точка перегину і як її ідентифікувати? Наведіть приклади.
4. В чому сутність теореми Зонтейдака?
5. Як перевірити, чи є функція опуклою, чи вгнутою?
6. Що таке стаціонарна точка цільової функції?
7. Вкажіть необхідні і достатні умови екстремуму.
8. В чому сутність методу множників Лагранжа?
9. Мінімізувати функцію  $f(x) = x_1^2 + x_2^2$  при обмеженні  $h_1(x) = 2x_1 + x_2 - 2$ , перетворивши цю задачу на задачу безумовної оптимізації за допомогою функції Лагранжа  $L(x; v)$ .
10. Мінімізувати функцію  $f(x) = x_1 + x_2$  при обмеженні  $x_1^2 + x_2^2 = 1$ , перетворивши цю задачу на задачу безумовної оптимізації за допомогою функції Лагранжа  $L(x; v)$ .
11. В чому полягає сутність умови Куна-Таккера?
12. Що таке активні і неактивні обмеження?
13. Мінімізувати функцію  $f(x) = x_1^2 - x_2$  при обмеженнях  $x_1 + x_2 = 6$ ,  $x_1 - 1 \geq 0$ ,  $x_1^2 + x_2^2 \leq 26$ , перетворивши задачу до задачі Куна-Таккера.



## 3 МЕТОДИ ОПТИМІЗАЦІЇ ОДНОВИМІРНИХ ЗАДАЧ

Процес оптимізації лежить в основі всієї інженерної діяльності, оскільки класичні функції інженера полягають у тому, щоб, з одного боку, проектувати нові, ефективніші та менш дорогі технічні системи, а з другого - розробляти методи підвищення якості функціонування існуючих систем. Ефективність методів оптимізації, які дозволяють вибирати найкращий варіант без перевірки всіх можливих варіантів, тісно пов'язана із широким використанням досягнень у галузі математики шляхом реалізації ітераційних обчислювальних схем, що опираються на досить обґрунтовані методи та алгоритми із застосуванням обчислювальної техніки.

В даний час, в зв'язку з доступністю персональних комп'ютерів, велика увага приділяється використанню чисельних методів оптимізації в інженерній практиці, які можливо розділити на дві великі групи: *методи безумовної і умовної оптимізації*. Цей розподіл пов'язаний з різним описом простору проектування. Область дослідження, тобто область, в якій інженер намагається визначити оптимум певної задачі, називають [5] *простором проектування*.

Простір проектування, який визначається проектними параметрами, зазвичай обмежений низкою умов, пов'язаних з фізичною сутністю задачі і розглядається в [5] у вигляді двох варіантів:

- 1) якщо проектний параметр один то, зазвичай, обмеження пов'язані з його значеннями, тобто область проектування звужується до відрізка дослідження  $[a, b]$ ;
- 2) якщо проектних параметрів декілька, то обмеження можуть накладатись або на їх значення, обмежуючи область дослідження, або в вигляді взаємозалежності між проектними параметрами, які повинні враховуватись при пошуку рішення (ці залежності в реальних задачах можуть відображати закони природи, економіки, права, наявність необхідних матеріалів і т. ін.).

В даному розділі розглядаються методи безумовного одновимірного пошуку оптимуму цільової функції, які базуються на використанні першого варіанта зображення простору проектування, де *цільова функція* – це вираз (функція), значення якого інженер намагається мінімізувати або максимізувати. При цьому передбачається [5, 6], що цільові функції, які досліджуються є *унімодальними*, тобто мають на інтервалі дослідження, який розглядається, тільки один оптимум (рис. 3.1). Таке обмеження на характер цільової функції не є таким жорстким, як може здатися, оскільки багато задач, з якими інженер стикається в своїй практиці, виявляються унімодальними.

Чисельні методи, які орієнтуються на розв'язання задач безумовної оптимізації, можна розділити на три класи [7, 8]:

- *методи прямого пошуку*, що базуються на обчисленні тільки значень цільової функції;

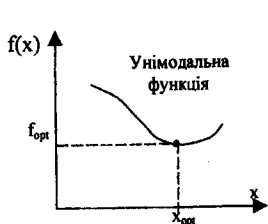


Рисунок 3.1 – Унімодальна цільова функція

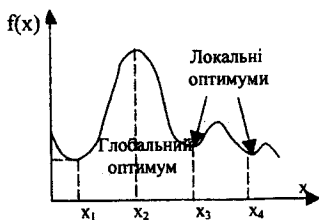


Рисунок 3.2 – Цільова функція з локальним та глобальним оптимумом

- *градієнтні методи*, в яких використовуються точні значення перших похідних  $f'(x)$ ;

- *методи другого порядку*, в яких поряд з першими похідними використовуються також другі похідні функції  $f(x)$ .

**Задача одновимірної оптимізації ставиться таким чином:** значення параметра  $X$  цільової функції  $f(x)$ , який називають *проектним параметром*, знаходяться на інтервалі дослідження  $[a, b]$ . В процесі пошуку оптимуму цільової функції цей інтервал, який називається *інтервалом невизначеності*, постійно зменшується (звужується), тому методи одновимірної оптимізації іноді називають методами звуження інтервалу невизначеності.

Вибір чисельного методу в першу чергу залежить від виду цільової функції, яка може бути [5] однопараметричною і багатопараметричною (рис. 3.3, 3.4).

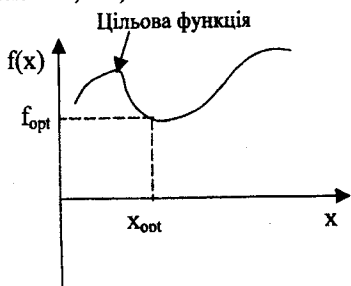


Рисунок 3.3 – Однопараметрична цільова функція

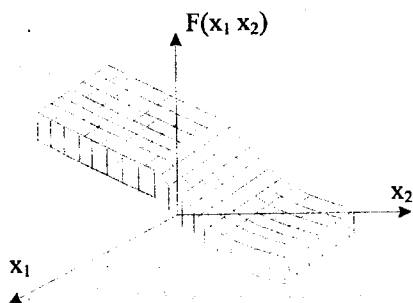


Рисунок 3.4 – Двопараметрична цільова функція

Деякі алгоритми оптимізації пристосовані до пошуку максимуму, а інші – для пошуку мінімуму.

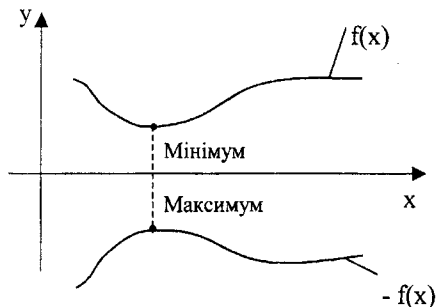


Рисунок 3.5 – Зміною знаку цільової функції на протилежний задача на мінімум перетворюється в задачу на максимум

Однак, незалежно від типу задачі, яка розв'язується на екстремум (оптимум) можливо користуватись одним і тим же алгоритмом, оскільки задачу максимізації можливо легко переробити в задачу на пошук мінімуму, змінивши знак цільової функції на протилежний (рис. 3.5).

**Загальна постановка задачі для методів одновимірної оптимізації** ставиться таким чином: нехай значення параметра  $X$  знаходиться на відрізку  $[a, b]$ , цільова функція унімодальна в області, яку досліджуємо, знайти оптимум цільової функції на відрізку  $[a, b]$ .

Більшість чисельних методів одновимірної оптимізації - це методи звуження відрізку  $[a, b]$ , а саме [5, 8, 9, 10]: метод розділення відрізку навпіл, метод дихотомії, метод золотого перерізу, метод Фібоначчі.

В процесі одновимірної оптимізації цільової функції на ЕОМ можна виділити два етапи:

- 1) встановлення меж відрізку, на якому реалізується процедура пошуку оптимуму;
- 2) зменшення відрізку до заданої похибки обчислення  $\varepsilon$ .

Перший етап реалізується за допомогою евристичних методів пошуку і є дуже складним [5, 11]. Другий - називають правилом виключення відрізків, реалізують алгоритм пошуку, що дозволяє знайти точку оптимуму шляхом послідовного виключення частин початкового обмеженого відрізку  $[a, b]$ , тобто за допомогою ітераційних алгоритмів. В якості умови закінчення ітераційного процесу використовується момент, коли алгоритми методів відрізняються один від одного способами пошуку оптимуму за як можливо меншу кількість кроків (ітерацій), коли підінтервал, що залишився, зменшиться до достатньо малих розмірів (зазвичай для цього задають значення заданої похибки обчислення  $\varepsilon$ ) [5, 6, 11, 12].

### 3. 1 Метод загального пошуку



Рисунок 3.6 – Метод загального пошуку

говорять про дроблення інтервалу невизначеності, яке характеризується коефіцієнтом  $f$ . Розділивши інтервал невизначеності на  $N$  частин, отримаємо  $N+1$  вузол, і тоді

$$f = \frac{2}{N+1}.$$

Щоб отримати значення  $f = 0,01$ , необхідно обчислити цільову функцію в 199 точках, а при  $f = 0,001$   $N = 1999$ . Звідси видно, що ефективність цього методу при зменшенні інтервалу невизначеності швидко падає. Напрощується інший шлях: щоб отримати  $f = 0,01$ , необхідно обчислити спочатку функцію в 19 точках і отримати  $f = 0,1$ , а потім, обчислити ще 19 значень функції на скороченому інтервалі невизначеності, отримати  $f = 0,01$ , зробивши при цьому всього 38, а не 199 обчислень. Таким чином, за деякої винахідливості ефективність пошуку можна різко збільшити.

### 3. 2 Метод половинного ділення (розділення відрізка навпіл)

Суть метода полягає [4, 5, 7] в постійному діленні відрізка дослідження цільової функції  $[a, b]$  навпіл і визначенні внутрі нього координат трьох точок  $x_1, x_2, x_m$ . При чому значення їх визначаються як:

$$x_m = \frac{a+b}{2}, \quad L = b-a, \quad x_1 = a + \frac{L}{4}, \quad x_2 = b - \frac{L}{4}.$$

Точки  $x_1, x_m, x_2$ , поділяють відрізок  $[a, b]$  на чотири рівні частини (рис. 3.6), обчислюємо значення цільової функції в точках  $x_1, x_2$ . Потім порівнюємо значення  $f(x_1)$  і  $f(x_m)$ , якщо  $f(x_1) < f(x_m)$ , то виключаємо з

Очевидно, найбільш природним способом звуження інтервалу невизначеності для одновимірної унімодалної функції є ділення його на декілька рівних частин з наступним обчисленням значень цільової функції в вузлах отриманої сітки (рис. 3.6). В результаті інтервал невизначеності звужується до двох кроків сітки. Звичайно

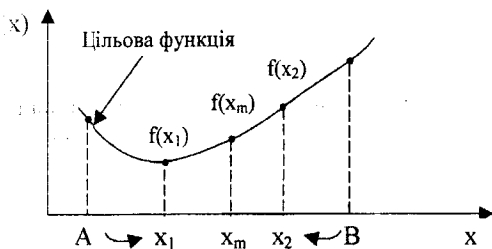


Рисунок 3.7 - Геометрична інтерпретація методу половинного ділення

дослідження відрізок  $[x_m, b]$  та покладемо  $b = x_m$ . Тоді середньою точкою нового відрізка  $[a, b]$  стає  $x_1$  ( $x_m = x_1$ ). Але якщо  $f(x_1) \geq f(x_m)$ , то порівнюємо значення цільової функції  $f(x_2)$  і  $f(x_m)$ ; якщо  $f(x_2) < f(x_m)$ , то виключаємо

відрізок  $[a, x_m]$ , покладемо  $a = x_m$ ,  $x_m = x_2$ ; якщо

$f(x_2) \geq f(x_m)$ , то виключаємо відрізок  $[a, x_1]$  та  $[x_2, b]$ , покладемо  $a = x_1$ ,  $b = x_2$ , тобто формуємо новий відрізок дослідження. Обчислюємо  $L = b - a$ , якщо умова  $|L| < \varepsilon$ , не виконується, то знову повертаємося до початку.

Даний алгоритм ітераційний, тому зазвичай в якості умови закінчення ітераційного процесу обирають умову  $a - b < \varepsilon$ , тобто звуження відрізка виконується до тих пір, поки його величина не зменшиться до заданої обчислювальної похибки  $\varepsilon$ . Алгоритм методу зображений на рис. 3.8.

Цей метод відрізняється великою ефективністю.

### 3.3 Метод дихотомії

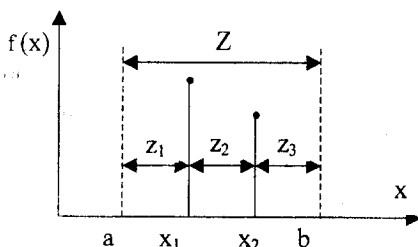


Рисунок 3.9 – Позначення, які використовуються в методі дихотомії

Обчислення цільової функції в двох точках інтервалу невизначеності дозволяє його звужити. Можна таким чином обрати ці точки, що інтервал невизначеності буде мінімальним. На рис. 3.9 показані позначення, які використовуються в цій схемі. Якщо значення цільової функції для  $x_1$  більше, ніж для  $x_2$ , то новий інтервал невизначеності дорівнює  $Z_1 = z_1 + z_2$ . Інакше він

визначається виразом  $Z_2 = z_2 + z_3$ . Задача полягає в тому, щоб одночасно мінімізувати  $Z_1$  і  $Z_2$ , задовольнивши умови [3, 5, 8]

$$\begin{aligned} z_1 + z_2 + z_3 &= Z, \\ z_1 > 0, z_2 > 0, z_3 > 0. \end{aligned}$$

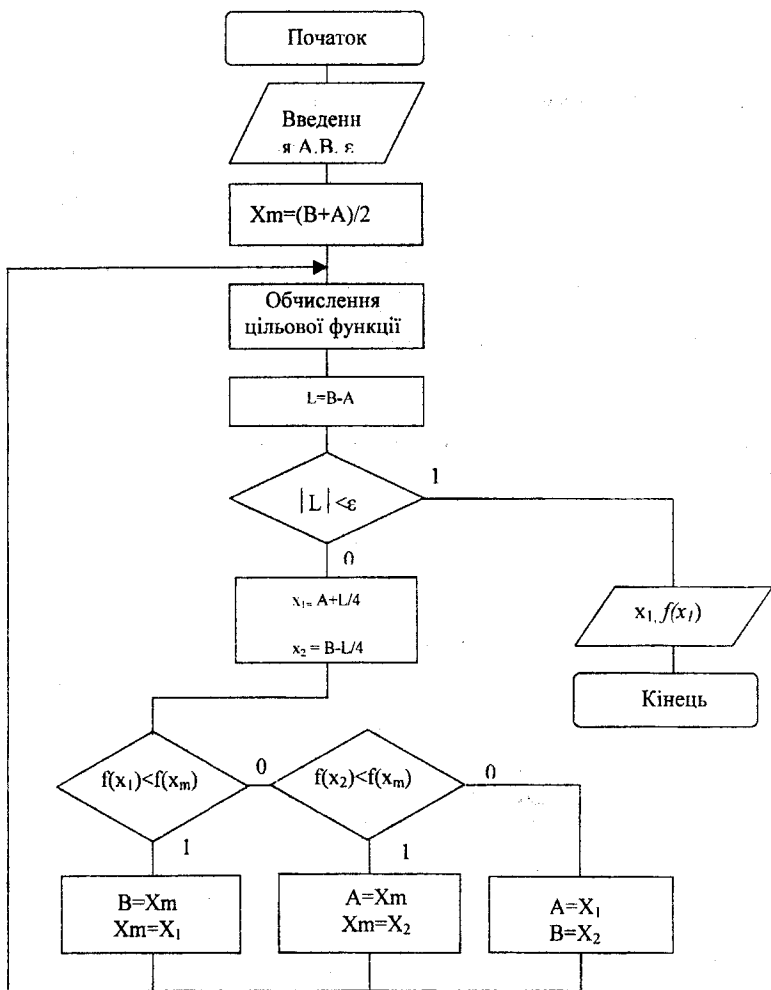


Рисунок 3.8 – Схема алгоритму метода половинного ділення

Із рівності можна виключити  $z_2$ . Тоді

$$Z - z_3 = \min, \quad Z - z_1 = \min.$$

Оскільки величина  $Z$  задана, то праві частини цих рівнянь будуть тим менші, чим більші  $z_1$  і  $z_3$ . Отже, оптимум відповідає умові

$$z_1 = z_3 = 0,5Z.$$

Але тоді  $z_2 = 0$ , що суперечить умові  $z_2 > 0$ .

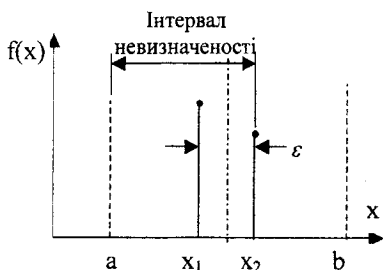


Рисунок 3.10 – Метод дихотомії

Нехай  $z_2$  має деяке дуже маленьке значення  $\varepsilon$ . Тоді із  $z_1$  і  $z_3$  віднімемо по  $\frac{\varepsilon}{2}$ . В результаті після обчислення першої пари значень цільової функції для близьких значень  $x$  інтервал невизначеності звужиться, як показано на рис. 3.10, і коефіцієнт дроблення буде

$$f = \frac{l}{2} + \frac{\varepsilon}{2}.$$

В межах коли  $\varepsilon \rightarrow 0$ ,  $f \rightarrow \frac{l}{2}$ . Надалі при використанні методу дихотомії виконуються ті операції, що і при використанні методу ділення відрізка навпіл. Відзначимо [9], що для досягнення однакових звужень інтервалу невизначеності метод дихотомії потребує обчислення цільової функції в точках на одну менше.

### 3.4 Метод “золотого перерізу”

З кожних трьох значень цільової функції, які були обчислені в інтервалі невизначеності в подальшому використовуються лише два, а третє не дає додаткової інформації і в подальшому не використовується. В методі золотого перетину [4, 5] цільова функція обчислюється в точках інтервалу невизначеності, які розташовані таким чином, щоб кожне

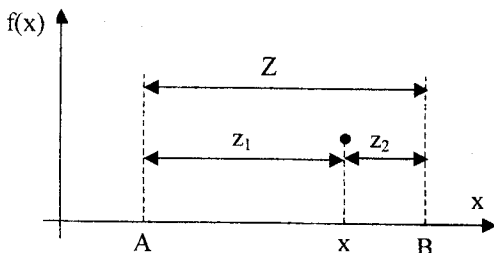


Рисунок 3.11 – Позначення, які використовуються в методі золотого перерізу

обчислене значення цільової функції давало нову корисну інформацію. Сутність цього методу полягає [1] в наступному. Інтервал невизначеності ділиться на дві нерівні частини, що відношення довжини більшого відрізка до довжини всього інтервалу дорівнює відношенню довжини

меншого відрізка до довжини більшого відрізка. На рис. 3.11 показаний інтервал невизначеності  $Z$ , який складається з відрізків  $z_1$  і  $z_2$ , відношення довжин яких визначається правилом золотого перетину.

$$\frac{z_1}{Z} = \frac{z_2}{z_1}$$

Крім того,  $z_1 + z_2 = Z$ . Із першого рівняння випливає  $z_1^2 = Z \cdot z_2$ . Підставивши значення  $Z$  з другого рівняння і поділивши обидві частини на  $z_1^2$ , отримаємо

$$\left(\frac{z_2}{z_1}\right)^2 + \frac{z_2}{z_1} - 1 = 0.$$

Розв'язуючи це квадратне рівняння, знаходимо для додатного кореня значення

$$\frac{z_2}{z_1} = \frac{-1 + \sqrt{5}}{2} = 0,618.$$

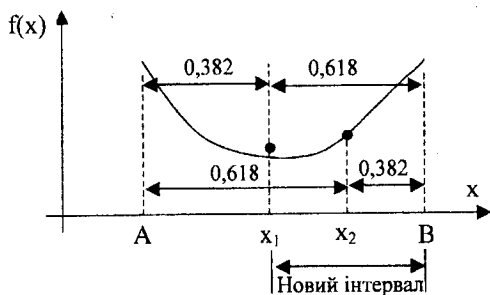


Рисунок 3.12 – Метод золотого перерізу

порівнянні з методом дихотомії, однак їх добре видно при подальшому діленні інтервалу, оскільки виявляється, що одне із значень цільової функції, яке необхідно обчислити на наступному кроці, вже відоме. Тому, щоб зменшити невизначеність ще в  $1/0,618$  раз, потрібно додатково обчислити тільки одне значення цільової функції в точці, яка визначається правилом золотого перерізу.

Для  $n > 2$  ефективність методу золотого перерізу вища, ніж у методу дихотомії, оскільки для кожного наступного обчислення цільової функції інтервал невизначеності скорочується в  $1/0,618$  раз. Після обчислення  $N$  значень цільової функції коефіцієнт дроблення інтервалу невизначеності складає

$$f = 0,618^{N-1}.$$

Метод золотого перерізу [3] дозволяє відмітити цікаву закономірність: найбільше скорочення наступних інтервалів невизначеності досягається при обчисленні цільової функції в точках, рівновіддалених від його центру. Якщо продовжувати таким чином і кожного разу, обчислюючи цільову



функцію, скорочувати інтервал невизначеності, то будуть справедливими такі відношення:

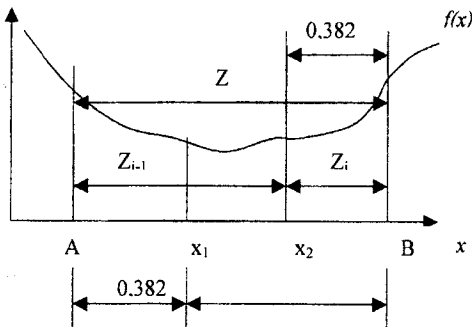
$$Z_{j-2} = Z_{j-1} + Z_j, \quad 1 < j < N,$$

де  $Z_j$  довжина інтервалу невизначеності після обчислення  $J$ -го значення цільової функції.

На рис. 3.13 представлений алгоритм вибору наступної точки пошуку. Задана точність може, звичайно, змінюватися вибором значення. Для функції  $f(x) = -e^{-x} \ln(x)$ , пошук відбувався в інтервалі  $(0, 2)$ .

Істинний мінімум знаходиться в точці 1,76322211, де значення функції дорівнює -0,0972601313.

Розробляючи програми для розв'язання задач однопараметричної оптимізації використовують такі співвідношення:



$$z_1 = 0.382$$

$$z_2 = 1 - z_1 = 0.618$$

$$x_1 = A + z_1(b - a)$$

$$x_2 = A + z_2(b - a)$$

Рисунок 3.13 - Послідовність етапів вибору наступної точки пошуку

### 3.5 Метод Фібоначчі

Припустимо, що потрібно визначити мінімум цільової функції як можна точніше, тобто з найменшим можливим інтервалом невизначеності, але при цьому можна виконати тільки  $n$  обчислень функції. Як слід вибрати  $n$  точок, в яких обчислюється функція? З першого погляду здається ясным, що не слід шукати розв'язку для всіх точок, які одержані в результаті експерименту. Навпаки, треба спробувати зробити так, щоб значення функції, отримані в попередніх експериментах, визначали положення наступних точок [1]. Справді, знаючи значення функції, ми тим самим маємо інформацію про саму функцію і положення її мінімуму і використаємо цю інформацію в подальшому пошуку.

Припустимо, що с інтервал невизначеності  $(x_1, x_3)$  і відомо значення функції  $f(x_2)$  всередині цього інтервалу (див. рис.3.15). Якщо можна

обчислити функцію всього один раз в точці  $x_4$ , то де слід помістити точку  $x_4$ , для того щоб отримати найменший можливий інтервал невизначеності?

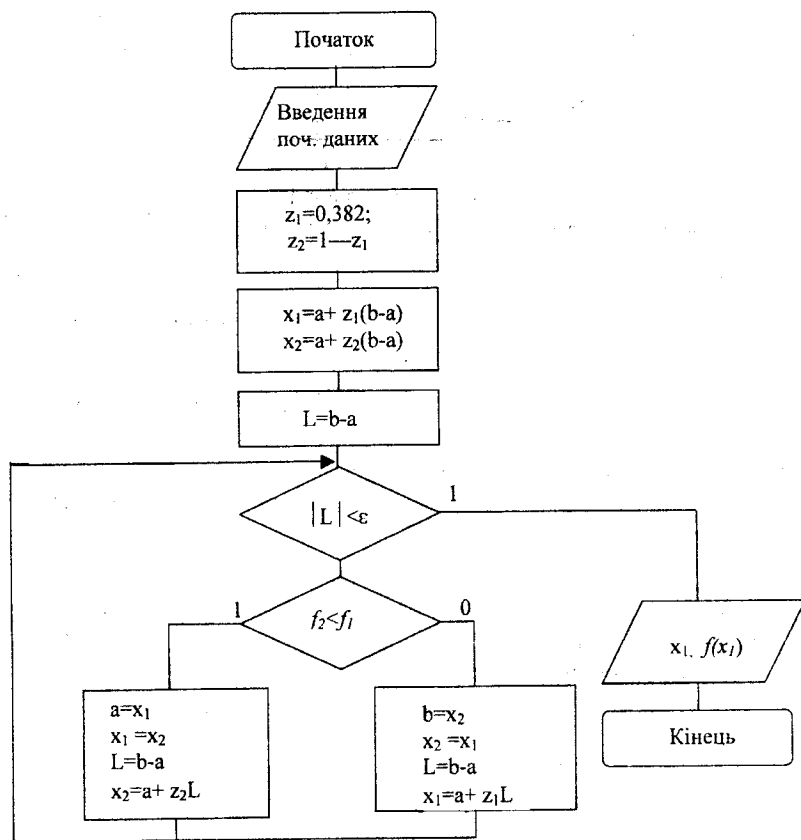


Рисунок 3.14 - Схема алгоритму метода „золотого перерізу”

Припустимо  $x_2 - x_1 = L$  і  $x_3 - x_2 = R$ , причому  $L > R$ , як показано на рис.3.15, і ці значення будуть фіксовані, якщо відомі  $x_1, x_2$  і  $x_3$ . Якщо  $x_4$  знаходиться в інтервалі  $(x_1, x_2)$ , то:

1. Якщо  $f(x_4) < f(x_2)$ , то новим інтервалом невизначеності буде  $(x_1, x_2)$  довжиною  $x_2 - x_1 = L$ .

2. Якщо  $f(x_4) > f(x_2)$ , то новим інтервалом невизначеності буде  $(x_4, x_3)$  довжиною  $x_3 - x_4$ .

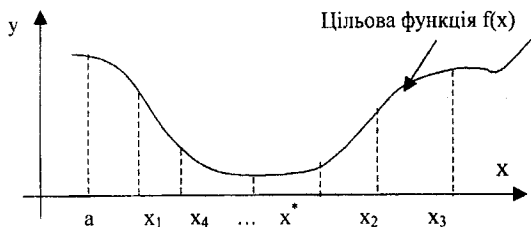


Рисунок 3.15 – Унімодална цільова функція. Геометрична інтерпретація методу Фібоначчі

Оскільки невідомо, яка з цих ситуацій буде мати місце, виберемо  $x_4$  таким чином, щоб зробити мінімально найбільшу з довжин  $x_3 - x_4$  і  $x_2 - x_1$ . Досягнути цього можна, зробивши

довжини  $x_3 - x_4$  і  $x_2 - x_1$  рівними, тобто помістивши  $x_4$  всередину інтервалу симетрично відносно точки  $x_2$ , що вже лежить всередині інтервалу. Будь-яке інше положення точки  $x_4$  може призвести до того, що отриманий інтервал буде більший  $L$ . Помістивши  $x_4$  симетрично відносно  $x_2$ , ми нічим не ризикуємо в будь-якому випадку.

Якщо виявиться, що можна виконати ще одне обчислення функції, то слід застосувати описану процедуру до інтервалу  $(x_1, x_2)$ , в якому вже є значення функції, обчислене в точці  $x_2$ . Отже, стратегія зрозуміла з самого початку. Потрібно помістити наступну точку всередину інтервалу невизначеності симетрично відносно точки, яка вже знаходиться там. Парадоксально, але, щоб зрозуміти, як потрібно починати обчислення, необхідно розібратися в тому, як його потрібно закінчувати.

На  $n$ -ому обчисленні  $n$ -у точку потрібно помістити симетрично по відношенню до  $(n-1)$ -ї точки. Положення цієї останньої точки в принципі залежить від нас. Для того щоб отримати найбільше зменшення інтервалу на даному етапі, слід поділити навпіл попередній інтервал. Тоді точка  $x_n$  буде збігатися з точкою  $x_{n-1}$ . Однак при цьому ми не одержуємо жодної нової інформації. Звичайно точки  $x_{n-1}$  і  $x_n$  знаходяться одна від одної на достатній відстані, щоб визначити, в якій половині, лівій чи правій, знаходиться інтервал невизначеності. Вони розміщуються на відстані  $\varepsilon/2$  по обидві сторони від середини відрізка  $L_{n-1}$ ; можна самостійно задати величину  $\varepsilon$  або вибрати цю величину рівною мінімально можливій відстані між двома точками. (Припустимо, що в нашому прикладі інженер може регулювати температуру з інтервалом в  $1\text{C}^\circ$ , тому  $\varepsilon = 1$ .)

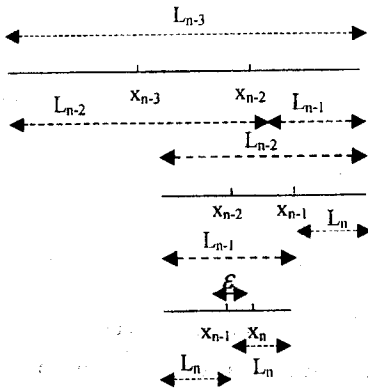


Рисунок 3.16 – Геометрична інтерпретація ітеративного процесу Фібоначчі

Інтервал невизначеності буде мати довжину  $L_n$ , отже,  $L_{n-1} = 2L_n - \varepsilon$  (рис. 3.16, нижня частина).

На попередньому етапі точки  $x_{n-1}$  і  $x_{n-2}$  повинні бути поміщені симетрично всередині інтервалу з  $L_{n-2}$  на відстані  $L_{n-1}$  від кінців цього інтервалу. Отже,  $L_{n-2} = L_{n-1} + L_n$  (рис. 3.16, середня частина).

**Зауваження.** З рисунку зрозуміло, що на передостанньому етапі  $x_{n-2}$  залишається в якості внутрішньої точки [1, 5].

Аналогічно

$$L_{n-3} = L_{n-2} + L_{n-1}. \quad (\text{рис. 3.16, верхня частина})$$

В загальному випадку

$$L_{j-1} = L_j + L_{j+1} \quad \text{якщо } 1 < j < n. \quad (3.1)$$

Таким чином,

$$L_{n-1} = 2L_n - \varepsilon, \quad (3.2)$$

$$L_{n-2} = L_{n-1} + L_n = 3L_n - \varepsilon,$$

$$L_{n-3} = L_{n-2} + L_{n-1} = 5L_n - 2\varepsilon,$$

$$L_{n-4} = L_{n-3} + L_{n-2} = 8L_n - 3\varepsilon \text{ і т.д.}$$

Якщо визначити послідовність чисел Фібоначчі таким чином:  $F_0 = 1, F_1 = 1$ , та  $F_k = F_{k-1} + F_{k-2}$  для  $k = 2, 3, \dots$ , тоді

$$L_{n-j} = F_{j+1}L_n - F_{j-1}\varepsilon, \quad j = 1, 2, \dots, n-1 \quad (3.3)$$

Якщо початковий інтервал  $(a, b)$  має довжину  $L_1 = (b - a)$ , то

$$L_1 = F_n L_n - \varepsilon \cdot F_{n-2}.$$

Тобто

$$L_n = \frac{L_1}{F_n} + \varepsilon \frac{F_{n-1}}{F_n}. \quad (3.4)$$

Отже, зробивши п обчислень функції, ми зменшимо початковий інтервал невизначеності в  $1/F_n$  раз у порівнянні з його початковою довжиною (нехтуючи  $\varepsilon$ ), і це буде найкращий результат.

Якщо ми вже почали пошук, то його нескладно продовжити, використовуючи описане вище правило симетрії. Отже, необхідно знайти положення першої точки, що розміщується на відстані  $L_2$  від одного з кінців початкового інтервалу, причому не важливо від якого кінця, оскільки друга точка розміщується згідно з правилом симетрії на відстані  $L_2$  від другого кінця інтервалу:

$$L_2 = F_{n-1}L_n - \varepsilon F_{n-3} = F_{n-1} \frac{L_1}{F_n} + \varepsilon \frac{(F_{n-1}F_{n-2} - F_n F_{n-3})}{F_n} = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n \varepsilon}{F_n}. \quad (3.5)$$

Після того як знайдене положення першої точки, числа Фібоначчі більше не потрібні. Використане значення  $\varepsilon$  може бути визначене з практичних міркувань. Воно повинне бути менше  $L_1/F_{n+1}$ , інакше ми будемо марно витратити час на обчислення функції.

Таким чином, пошук методом Фібоначчі, названий так через появу під час пошуку чисел Фібоначчі, є ітераційною процедурою [1]. В процесі пошуку інтервалу  $(x_1, x_2)$  з точкою  $x_2$ , що вже лежить в цьому інтервалі, наступна точка  $x_4$  завжди вибирається такою, що  $x_3 - x_4 = x_2 - x_1$  або  $x_4 - x_1 = x_3 - x_2$ , тобто

$$x_4 = x_1 - x_2 + x_3. \quad (3.6)$$

Якщо  $f(x_2) = f_2$  та  $f(x_4) = f_4$ , тоді можна розглянути чотири випадки (рис. 3.17).

### Приклад 1

Використати метод Фібоначчі для пошуку мінімуму функції  $f(x) = 2x^2 - e^x$  в інтервалі  $(0,1)$  для 10-кратного обчислення функції.

Остаточний інтервал невизначеності має довжину

$$0,359550563 - 0,348314606 = 0,01123957 \approx \frac{1}{89} = \frac{1}{F_{10}}.$$

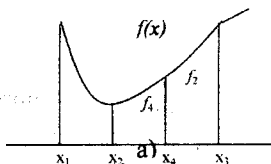
З точністю до шостого знаку після коми мінімум досягається в точці  $x=0,357403$ , і в цій точці  $f(x^*) = -1,17413$ .

а)  $x_4 < x_2$

$f_4 < f_2$

Новий інтервал  $(x_1, x_2)$ ,

що містить точку  $x_4$

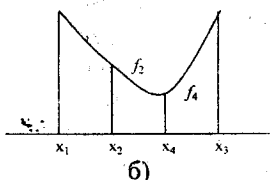


б)  $x_4 > x_2$

$f_4 < f_2$

Новий інтервал  $(x_2, x_3)$ ,

що містить точку  $x_4$

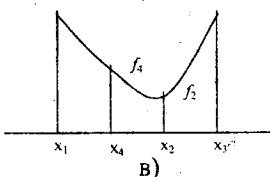


в)  $x_4 < x_2$

$f_4 > f_2$

Новий інтервал  $(x_4, x_3)$

що містить точку  $x_2$



г)  $x_4 > x_2$

$f_4 > f_2$

Новий інтервал  $(x_1, x_4)$

що містить точку  $x_2$

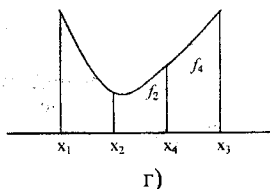


Рисунок 3.17 – Геометрична інтерпретація алгоритму визначення цільової функції на  $i$ -му кроці ітерації

### 3.6 Порівняння методів одновимірного пошуку

Найкращими критеріями порівняння методів пошуку, які були описані вище, є їх ефективність і універсальність. Під ефективністю алгоритму розуміють число обчислень функції, необхідне для досягнення необхідного звуження інтервалу невизначеності. Із табл. 3.1 видно, що найкращим в цьому відношенні є метод Фібоначчі, а найгіршим – метод загального пошуку. Конструктор не з великим задоволенням використовує метод Фібоначчі, оскільки, застосовуючи його, необхідно заздалегідь задавати число обчислень значень функції. Однак він може скористатися методом золотого перерізу. Як правило, методи Фібоначчі і золотого перерізу, мають високу ефективність, найбільш підходять для розв'язання одновимірних унімодальних задач оптимізації.

Універсальність алгоритму означає, що його можна легко застосувати для розв'язання найрізноманітніших задач. В цьому відношенні метод Фібоначчі, поступається іншим, оскільки потребує окремого обчислення положення точок, в яких будуть визначатися значення цільової функції на кожному новому кроці. Цим приходиться розплачуватися за підвищення ефективності методу. З точки зору універсальності, малоефективний метод загального пошуку має по крайній мірі одну перевагу – його можна з успіхом застосовувати і для неунімодалних функцій, якщо всі вони достатньо гладкі. Нерідко заздалегідь не відомо, чи є розглянута цільова функція унімодалною. В таких випадках слід використати декілька різних алгоритмів і подивитись, чи дають вони усі один і той самий оптимум. Звідси витікає важливий висновок, який слід мати на увазі, розв'язуючи задачі оптимізації: не існує універсального алгоритму, який дозволяв би розв'язувати будь-які задачі. Вирішуючи складні задачі оптимізації, слід користуватися різними методами, оскільки це дозволяє збільшити частку вигідних розв'язків.

Таблиця 3.1 – Порівняння методів одновимірного пошуку за значеннями коефіцієнта дроблення інтервалу невизначеності  $f$

Кількість обчислень цільової функції $N$	Загальний пошук	Ділення відрізка навпіл	Метод дихотомії	Метод золотого перерізу	Метод Фібоначчі
1	1,0	1,0	1,0	1,0	1,0
2	0,667	-	0,500	0,618	0,500
3	0,500	0,500	-	0,382	0,333
4	0,400	-	0,250	0,236	0,200
5	0,333	0,250	-	0,146	0,125
6	0,286	-	0,125	0,090	0,077
7	0,250	0,125	-	0,056	0,048
8	0,222	-	0,0625	0,345	0,0294
9	0,200	0,0625	-	0,0213	0,0182
10	0,182	-	0,0312	0,0132	0,0112
11	0,167	0,0312	-	0,00813	0,00694
12	0,154	-	0,0156	0,00502	0,00429
13	0,143	0,0156	-	0,00311	0,00265
14	0,133	-	0,00781	0,00192	0,00164
15	0,125	0,00781	-	0,00119	0,00101
16	0,118	-	0,00391	0,000733	0,000626
17	0,111	0,00391	-	0,000453	0,000387
18	0,105	-	0,00195	0,000280	0,000239
19	0,100	0,00195	-	0,000173	0,000148
20	0,095	-	0,000976	0,000107	0,0000913

## Контрольні запитання і завдання

1. В чому полягає властивість унімодальності функцій і в чому полягає важливість цієї властивості при розв'язуванні задач оптимізації з однією змінною?
2. Якщо точка цільової функції задовольняє достатні умови існування локального мінімуму, то як встановити чи є цей мінімум глобальним?
3. Чи є методи виключення інтервалів в цілому більш ефективними, ніж методи точкового оцінювання? Чому?
4. При реалізації пошукових алгоритмів на ЕОМ в якості умови закінчення ітераційного процесу використовуються як аналіз абсолютної величини різниці поточних значень кінців відрізка дослідження, так і абсолютної величини різниці поточних значень цільової функції. Чи можлива ситуація, коли результат однієї з перевірок вказує на збіжність до точки мінімуму, тоді як отримана точка в дійсності мінімуму не відповідає. Поясніть відповідь рисунком.
5. Задані такі функції однієї змінної:

$$\text{а) } f(x) = x^5 + x^4 - \left(\frac{x^3}{3}\right) + 2; \quad \text{б) } f(x) = (2x+1)^2(x-4).$$

Для кожної з заданих функцій знайдіть:

- інтервал(и) зростання, спадання;
  - точки перегину (якщо такі існують);
  - інтервал(и), в якому функція вгнута, опукла;
  - локальний і глобальний максимуми (якщо такі існують);
  - локальний і глобальний мінімуми (якщо такі існують).
6. Встановіть області, в яких наступна функція опукла чи вгнута:  $f(x) = e^{-x^2}$ . Знайдіть глобальний максимум і глобальний мінімум цієї функції.
  7. В чому суть алгоритму методу дихотомії? Складіть схему алгоритму методу. Виберіть і обґрунтуйте умову виходу з ітераційного циклу.
  8. Дослідіть функцію  $f(x) = x^3 - 12x + 32$  в інтервалі  $-4 \leq x \leq 4$ . Знайдіть локальні мінімуми, локальні максимуми, глобальний мінімум і глобальний максимум  $f$  в заданому інтервалі.
  9. В чому сутність алгоритму методу Фібоначчі? Складіть схему алгоритму методу, наведіть основну математичну модель методу.
  10. Розробіть програму для ЕОМ, яка реалізує пошук оптимуму за методом золотого перерізу і методом Фібоначчі для функції  $f(x) = 3x^2 + \frac{12}{x^3} - 5$  в інтервалі  $\frac{1}{2} \leq x \leq \frac{5}{2}$ . Порівняйте результуючі інтервали пошуку, які отримані за допомогою перерахованих методів.



## 4 МЕТОДИ ОПТИМІЗАЦІЇ БАГАТОВИМІРНИХ ЗАДАЧ

У цьому розділі розглядаються фундаментальні поняття і конкретні методи, які використовують під час пошуку безумовних мінімумів функцій декількох змінних. Викладене ґрунтується на матеріалі розділу 3, оскільки одновимірні методи відіграють дуже важливу роль при дослідженні функції декількох змінних.

На перший погляд може здатися, що відмінність між методами багатовимірного і одновимірного пошуку полягає лише в тому, що перші вимагають більшого об'єму обчислень, і, що в принципі методи, які придатні для функцій однієї змінної, можна застосовувати і для функцій багатьох змінних. Однак це не так, оскільки багатовимірний простір якісно відрізняється від одновимірного. Передусім зі збільшенням числа вимірів зменшується ймовірність унімодальності цільової функції. Крім того, безліч елементів, які утворюють багатовимірний простір, значно потужніші множини елементів одновимірного простору. Об'єм обчислень, які необхідні для звуження інтервалу невизначеності в багатовимірному просторі, є степеневою функцією, показник якої рівний розмірності простору. Так, якщо в випадку одновимірного простору для досягнення  $f=0,1$  вимагається обчислити 19 значень цільової функції, то у випадку двовимірного простору [5] це число складає 361, тривимірного – 6859, чотиривимірного – 130321, а п'ятивимірного – 2476099! Оскільки при виборі оптимальної конструкції нерідко потрібно мати діло з п'ятьма і більше змінними, серйозність труднощів, зумовлених багатовимірністю, стає очевидною.

Спочатку розглянемо питання аналізу «в статистиці» з використанням положення лінійної алгебри і диференціального обчислення [2, 3, 6, 8], а також умови, які (в достатньо загальних можливих випадках) дозволяють ідентифікувати точки оптимуму. Такі умови використовують для перевірки обраних точок і дають можливість з'ясувати, чи є ці точки точками мінімуму чи сідловими точками. При цьому задача вибору вказаних точок залишається за межами цього аналізу; основна увага віддається розв'язанню питання про те, чи відповідають досліджувані точки розв'язку багатовимірної задачі безумовної оптимізації, у якій вимагається мінімізувати  $f(x)$ ,  $x \in R^N$ , (4.1) коли обмеження відсутні на  $x$ , де  $x$  — вектор керованих змінних розмірності  $N$ ,  $f$  — скалярна цільова функція. Звичайно допускається, що  $x_i$  (для всіх значень  $i=1, 2, 3, \dots, N$ ) можуть приймати будь-які значення, хоча інколи в практичних застосуваннях область значень  $x$  обирається у вигляді дискретної множини. Крім того, часто зручно допускати, що функція  $f$  і її похідні існують і неперервні скрізь, хоча ми знаємо, що оптимуми можуть досягати в точках розриву  $f$  чи її градієнта

$$\nabla f = \left[ \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \dots, \frac{\partial f}{\partial x_N} \right]^T. \quad (4.1)$$

Потрібно мати на увазі, що функція  $f$  може приймати мінімальні значення в точці  $\bar{X}$ , в якій  $f$  чи  $\nabla f$  розриваються. Крім того, цієї точки  $\nabla f$  може не існувати. Для того, щоб побудувати систему конструктивних критеріїв оптимальності, необхідно (якнайменше на першій стадії дослідження) виключити із розгляду подібні ситуації, які дуже ускладнюють аналіз. У ряді випадків приходиться обмежуватись лише ідентифікацією локальних оптимумів, оскільки нелінійна цільова функція  $f$  не завжди має опуклий характер, може бути мультимодальною. На рис. 4.1 відображені лінії рівняння функції Хіммельблау

$$f(x) = [x_1^2 + x_2 - 11]^2 + [x_1 + x_2^2 - 7]^2, \quad (4.2)$$

неважко побачити, що ця функція має чотири різних мінімуми.

Далі підійдемо до питання аналізу «в динаміці» [2, 6, 8], яке формулюється таким чином: якщо точка  $x^{(0)}$  не задовольняє

### Функція Хіммельблау:

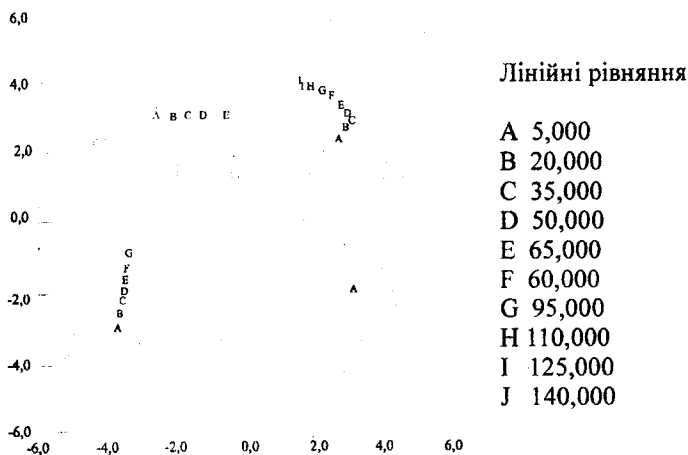


Рисунок 4.1 – Лінії рівняння мультимодальної функції

вимоги, які визначаються критеріями оптимальності, то як отримати «добре» нове наближення  $x^{(1)}$  до розв'язку  $x$ ? Спроба дати відповідь на це питання приводить до необхідності розгляду ряду методів. Методи, які

розглядаються, класифікуються у відповідності з тим, чи використовується інформація про похідні досліджуваної функції.

#### 4.1 Критерії оптимальності

Розглянемо умови, які дозволяють характеризувати (тобто класифікувати) точки простору керованих змінних. Критерії оптимальності необхідні для розпізнавання розв'язків і, крім того, складають основу більшості методів пошуку розв'язків, що використовуються.

Для цього розглянемо розкладання Тейлора для функції декількох змінних :

$$f(x) = f(\bar{X}) + \nabla f(\bar{X})^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(\bar{X}) \Delta x + O_3(\Delta x), \quad (4.3)$$

де  $\bar{X}$  — точка розкладення із простору  $R^N$ ;  $\Delta x = x - \bar{X}$  — величина зміни  $x$ ;  $\nabla f(x)$  —  $N$ -вимірний вектор-стовпець перших похідних  $f(x)$ , обчислених в точці  $\bar{X}$ ;  $\nabla^2 f(\bar{X}) = H_f(\bar{X})$  — симетрична матриця порядку  $N \times N$  других частинних похідних  $f(x)$ , вирахованих у точці  $\bar{X}$ . (Цю матрицю часто називають *матрицею Гессе*. Її елемент, розташований на перетині  $i$ -го рядка і  $j$ -го стовпця, дорівнює  $\partial^2 f / \partial x_i \partial x_j$ ).  $O_3(\Delta x)$  — сума всіх членів розкладання, які мають порядок по  $\Delta x$  вище другого. Нехтуючи членами найвищих порядків (тобто виключаючи  $O_3(\Delta x)$ ), визначимо величину зміни цільової функції  $f(x)$ , відповідно довільній зміні  $x$ :

$$\Delta f(x) = f(x) - f(\bar{X}) = \nabla f(\bar{X})^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(\bar{X}) \Delta x. \quad (4.4)$$

Нагадаємо, що за означенням у всіх точках із околу точки мінімуму цільова функція приймає значення, які вищі мінімального, тобто має місце нерівність

$$\Delta f = f(x) - f(\bar{X}) \geq 0. \quad (4.5)$$

Точка  $\bar{X}$  є точкою *глобального* мінімуму, якщо нерівність (4.5) виконується для всіх  $x \in R^N$ ; такі точки будемо позначати через  $x^{**}$ . Коли формула (4.5) справедлива лише в деякому  $\delta$ -околі точки  $\bar{X}$ , тобто для всіх  $x$ , таких, що  $\|x - \bar{X}\| < \delta$  для заданого  $\delta > 0$ , то  $\bar{X}$  є точкою *локального* мінімуму, чи  $x^*$ . Якщо

$$\Delta f = f(x) - f(\bar{X}) \leq 0, \quad (4.6)$$

то  $\bar{X}$  є точкою максимуму (локального чи глобального у відповідності з даним вище означенням). Виключення знака рівності із формул (4.5) і (4.6) дозволяє визначити точку *строого* мінімуму чи максимуму. В тому випадку, коли  $\Delta f$  приймає як додатне, так і від'ємне, так і нульове значення в залежності від вибору точок із  $\delta$ -околу, точка  $\bar{X}$  являє собою *сідлову точку*.

Повернемось до рівняння (4.4) і згадаємо про висунуте раніше припущення про те, що  $f(x)$ ,  $\Delta f(x)$  і  $\Delta^2 f(x)$  існують і неперервні для всіх  $x \in R^N$ . Як випливає з формули (4.4), для того щоб знак  $\Delta f$  не змінювався при довільній зміні  $\Delta x$ , градієнт  $\nabla f(\bar{X})$  повинен дорівнювати нулю, тобто  $\bar{X}$  повина бути *стаціонарною точкою*. Інакше різниця  $\Delta f$  може приймати додатні чи від'ємні значення в залежності від знаків  $\nabla f(\bar{X})$  і  $\Delta x$ . Таким чином, точка  $\bar{X}$  повинна задовольняти умову *стаціонарності*

$$\nabla f(\bar{X})=0, \quad (4.7)$$

і формула (4.4) приймає такий вигляд :

$$\Delta f(x) = \frac{1}{2} \Delta x^T \nabla^2 f(\bar{X}) \Delta x. \quad (4.8)$$

Звідси видно, що знак  $\Delta f(x)$  визначається квадратичною формою

$$Q(x) = \Delta x^T \nabla^2 f(\bar{X}) \Delta x, \quad (4.9)$$

чи

$$Q(x) = z^T A z. \quad (4.10)$$

Із лінійної алгебри ми знаємо, що:

- A – додатно визначена матриця, якщо  $Q(z) > 0$  для будь-яких  $z$ ;
- A – додатно напіввизначена матриця, якщо  $Q(z) \geq 0$  для будь-яких  $z$ ;
- A – від'ємно визначена матриця, якщо  $Q(z) < 0$  для будь-яких  $z$  ;
- A – від'ємно напіввизначена матриця, якщо  $Q(z) \leq 0$  для будь-яких  $z$  ;
- A – невизначена матриця, якщо  $Q(z) > 0$  для деяких  $z$  і  $Q(z) < 0$  для інших  $z$ .

Із (4.10) видно, що стаціонарна точка  $\bar{X}$  є

- точкою мінімуму, якщо  $\nabla^2 f(\bar{X})$  — додатно напіввизначена матриця;
- точкою максимуму, якщо  $\nabla^2 f(\bar{X})$  — від'ємно напіввизначена матриця;
- сідлова точка, якщо  $\nabla^2 f(\bar{X}) > (\text{або } <) 0$  — невизначена матриця.

Крім того, іноді корисно провести аналіз стаціонарної точки  $\bar{X}$  в деякому іншому аспекті. Розглянемо стаціонарну точку  $\bar{X}$  разом з навколишнім  $\delta$ -околом і векторами, що виходять із точки  $\bar{X}$  (рис. 4.2). При цьому

$$\tilde{x} = x + \alpha s(\bar{X}). \quad (4.11)$$

Шляхом відповідного вибору  $\alpha$  і  $s$  можливо побудувати всі точки  $\tilde{x}$  із кола точки  $\bar{X}$ . Підстановка (4.11) в (4.8) дає формулу

$$\Delta f(x) = (\alpha^2/2)s^T \nabla^2 f(x)s. \quad (4.12)$$

Тепер за допомогою (4.10) і (4.11) можливо класифікувати  $s(x)$  як напрямок спуску, напрямок підйому чи напрямок загального вигляду. Якщо напрямок спуску знайти *не вдається*, то  $\bar{X}$  є точкою локального мінімуму  $x^*$ , що відповідає випадку, коли  $\nabla^2 f(x)$ - додатно напіввизначена матриця.

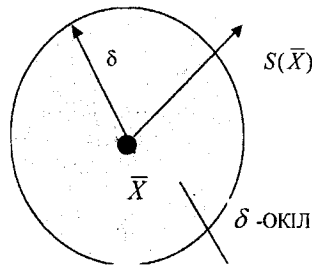


Рисунок 4.2 – Окіл стаціонарної точки

### Теорема 1. Необхідні умови

Для наявності в точці  $x^*$  локального мінімуму необхідно, щоб виконувалась рівність

$$\nabla f(x^*) = 0 \quad (4.13)$$

і матриця  $\nabla^2 f(x^*)$  була додатно напіввизначеною:

$$\nabla^2 f(x^*) \geq 0. \quad (4.14)$$

### Теорема 2. Достатні умови

Якщо

$$\nabla f(x^*) = 0 \quad (4.15a)$$

і матриця  $\nabla^2 f(x^*)$  додатно визначена, то

$$\nabla^2 f(x^*) > 0. \quad (4.156)$$

Доведення цих теорем випливає із наведеного вище судження. Звичайно доводиться обмежуватись знаходженням локального мінімуму; разом з тим, якщо можна показати, що  $x^T \nabla^2 f(x) x \geq 0$  для всіх  $x$ , то  $f(x)$  називають *опуклою* функцією, а локальний мінімум називають глобальним.

#### Приклад 4.1. Критерії оптимальності

Розглянемо функцію

$$f(x) = 1x_1^2 + 4x_1x_2^3 - 10x_1x_2 + x_2^2,$$

лінії рівня якої зображені на рис. 4.3. Треба класифікувати точку  $\bar{X} = [0, 0]^T$

$$f(x) = 2(x_1)^2 + 4x_1x_2^3 - 10x_1x_2 + x_2^2$$

Розв'язування:

$$\frac{\partial f}{\partial x_1} = 4x_1 + 4x_2^3 - 10x_2,$$

$$\frac{\partial f}{\partial x_2} = 12x_1x_2^2 - 10x_1 + 2x_2,$$

$$\nabla f(x) = [0, 0]^T.$$

Звідси, точка  $\bar{X}$  — стаціонарна.

$$\frac{\partial^2 f}{\partial x_1^2} = 4,$$

$$\frac{\partial^2 f}{\partial x_2^2} = 24x_1x_2 + 2 = +2,$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 12x_2^2 - 10 = -10.$$

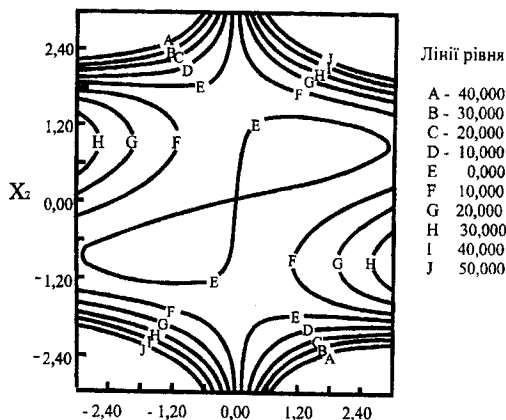


Рисунок 4.3 – Лінії рівня нелінійної функції двох змінних з прикладу 4.1.

Звідси,

$$\nabla^2 f(x) = H_f(x) = \begin{bmatrix} +4 & -10 \\ -10 & +2 \end{bmatrix}$$

Матриця  $\nabla^2 f(\bar{x})$  є невизначеною, оскільки квадратична форма  $z^T H_f z$  приймає додатне значення якщо  $z=(0,1)$  і від'ємне значення якщо  $z=(1,1)$ . Тому  $\bar{x}$  представляє собою *сідлову точку*, що і відображено на рис. 4.3.

#### 4.2 Методи прямого пошуку

В цьому розділі розглядається *питання аналізу* «в динаміці» для функції декількох змінних, тобто досліджуються методи і алгоритми, що дозволяють на ітераційній основі отримувати оцінки  $x^*$  вектора керованих змінних, якому відповідає мінімальне значення функції  $f(x)$ . Указані методи застосовуються [2, 4, 8] також і до задач максимізації [2, 4, 8], в яких цільову функцію треба замінити на  $-f(x)$ . Методи, орієнтовані на розв'язання задач оптимізації, можна розділити на три широких класи:

1. *Методи прямого пошуку*, оснований на обчисленні тільки значень цільової функції.

2. *Градiєнтні методи*, в яких використовуються точні значення перших похідних  $f(x)$ .

3. *Методи другого порядку*, в яких поряд з першими похідними використовуються також похідні функції  $f(x)$  більших порядків.

Методи прямого пошуку можливо використовувати для розв'язання задач, в яких  $\nabla f$  існує, і вони часто використовуються в тих випадках, коли  $\nabla f$  являє собою складну векторну функцію керованих змінних. Отже, в цьому і в наступних розділах припускається, що функція  $f(x)$  унімодальна в області, яка розглядається. Якщо методи оптимізації використовуються для аналізу мультимодальних функцій, то доводиться обмежитися ідентифікацією локальних мінімумів.

Багатовимірні методи, які реалізують процедуру пошуку оптимуму на основі обчислення значень функції, можна з загальних позицій розділити на евристичні та теоретичні. Евристичні методи реалізують процедуру пошуку за допомогою інтуїтивних геометричних зображень і забезпечують отримання частинних емпіричних результатів. З другого боку, теоретичні методи ґрунтуються на фундаментальних математичних теоремах і мають такі операційні якості, як збіжність (як мінімум при виконанні деяких визначених умов). Нижче ми розглянемо три методи прямого пошуку:

- 1) пошук по симплексу, чи  $S^2$ -метод;
- 2) метод покоординатного підйому;
- 3) метод виключення областей.

Перші два із перерахованих методів відносяться до категорії евристичних і реалізують принципово різні стратегії пошуку. В процесі пошуку по  $S^2$ -методу послідовно оперують регулярними симплексами в просторі керованих змінних, тоді як при реалізації методу покоординатного підйому використовується фіксована множина (координатних) напрямків, які обираються рекурсивним способом. Метод виключення областей оснований на теоретичних результатах і орієнтований на розв'язання задач з квадратичними цільовими функціями; для таких задач метод збігається за кінцеве число ітерацій. До числа загальних особливостей всіх трьох методів варто віднести відносну простоту відповідних обчислювальних процедур, які легко реалізуються і швидко коректуються. З другого боку, реалізація вказаних методів може вимагати (і часто вимагає) більш значних затрат часу порівняно з методами використання похідних.



## 4.2.1 Метод пошуку по симплексу ( $S^2$ -метод)

Перші спроби розв'язання оптимізаційних задач без обмежень, на основі прямого пошуку, зв'язані з використанням одновимірних методів оптимізації. Як правило, при реалізації таких методів допустима область визначення показника якості функціонування системи (цільової функції) замінюється дискретною множиною (решіткою) точок простору керованих змінних, а потім використовуються різні стратегії зменшення області, яка має розв'язок задач. Часто ця процедура виявляється [4, 8] еквівалентною рівномірному пошуку в вузлах і є недостатньою для розв'язання задач з числом змінних, яке більше двох. Більш корисна ідея полягає у виборі базової точки і оцінюванні значень цільової функції в точках, що обмежують базову точку. Наприклад, при розв'язанні задач з двома змінними можливо використати квадратний зразок, відображений на рис. 4.4. Далі «найкраща» з п'яти досліджуваних точок вибирається за наступну базову точку, біля якої будується аналогічний зразок. Якщо жодна з кутових точок не має переваг над базовою, розміри зразка слід зменшити, після чого продовжити пошук.

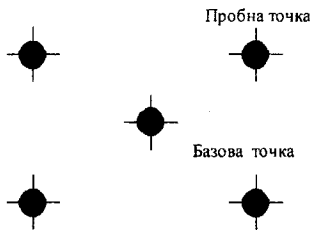


Рисунок 4.4 – Квадратичний зразок (частинний випадок кубічного зразка)

Цей тип *еволюційної оптимізації* був використаний Боксом [2] і іншими дослідниками для аналізу функціонування промислових підприємств, коли ефект варіацій значень змінних, що описують промислові процеси, вимірюється з похибкою. В задачах великої розмірності обчислення значень цільової функції проводиться в усіх вершинах, а також у центрі ваги гіперкуба, тобто в точках *кубічного зразка*. Якщо кількість змінних (розмірність простору, в якому ведеться пошук) дорівнює  $N$ , то пошук по кубічному зразку вимагає  $2^N + 1$  обчислень значень функції для одного зразка. При збільшенні розмірності задачі необхідна кількість обчислень значення цільової функції зростає дуже швидко. Таким чином, не зважаючи на логічну простоту пошуку по кубічному зразку, виникає необхідність використання більш ефективних

методів прямого пошуку для розв'язання задач оптимізації, що виникають на практиці.

Одна з тих стратегій пошуку, що викликає особливий інтерес, ґрунтується на *методі пошуку по симплексу*, запропонованому Спендлі, Хекстом і Хімсвордом [2, 3]. Слід відзначити, що вказаний метод і інші подібні методи не мають відношення до симплекс-методу лінійного програмування, а подібність назв носить випадковий характер. Процедура симплексного пошуку Спендлі, Хекста і Хімсворда базується на тому, що експериментальним зразком, який має найменшу кількість точок, є регулярний симплекс. Регулярний симплекс у  $N$ -вимірному просторі представляє собою многокутник, утворений  $N+1$  рівновіддаленими одна від одної точками-вершинами. Наприклад, у випадку двох змінних симплексом є трикутник з однаковими сторонами; в тривимірному просторі симплекс представляє собою тетраедр. У алгоритмі симплексного пошуку використовується [5, 8] важлива властивість симплексів, згідно з якою новий симплекс можливо побудувати на будь-якій грані початкового симплекса шляхом переносу вибраної вершини на належну відстань вздовж прямої, проведеної через центр ваги інших вершин початкового симплекса.

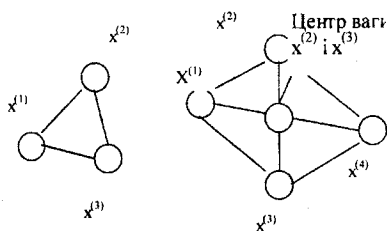


Рисунок 4.5 – Побудова нового симплекса: а – початковий симплекс,  $x^{(1)}$ ,  $x^{(2)}$ ,  $x^{(3)}$ ; б – новий симплекс,  $x^{(1)}$ ,  $x^{(2)}$ ,  $x^{(4)}$

Отримана таким чином точка є вершиною нового симплекса, а вибрана при побудові вершина початкового симплекса виключається. Неважко побачити, що при переході до нового симплекса необхідно одне обчислення значень цільової функції. Рис. 4.5 ілюструє процес побудови нового симплекса на площині.

Робота алгоритма симплексного пошуку починається з побудови регулярного симплекса в просторі незалежних змінних і оцінювання значень цільової функції в кожній із вершин симплекса. При цьому визначається вершина, якій відповідає найбільше значення цільової функції. Далі найдена вершина проектується через центр ваги інших вершин симплекса в нову точку, яка використовується за вершину нового симплекса. Якщо функція зменшується достатньо плавно, ітерації

продовжуються до тих пір, поки не буде накрыта точка мінімуму, або не почнеться циклічний рух по двох чи більше симплексах. В таких ситуація можливо використати такі три правила.

### **Правило 1. «Накрив» точки мінімуму**

Якщо вершина, якій відповідає найбільше значення цільової функції, побудована на попередній ітерації, то замість неї береться вершина, якій відповідає наступне за величиною значення цільової функції.

### **Правило 2. Циклічний рух**

Якщо деяка вершина симплекса не виключається протягом більш ніж  $M$  ітерацій, то необхідно зменшити розміри симплекса за допомогою коефіцієнта редукції і побудувати новий симплекс, вибравши за базову точку, якій відповідає мінімальне значення цільової функції. Спендлі, Хекст і Хімсворд запропонували обчислити  $M$  за формулою

$$M=1,65N+0,05N^2,$$

де  $N$  – розмірність задачі, а  $M$  округлюється до найближчого цілого числа. Для використання цього правила треба встановити величину коефіцієнта редукції.

### **Правило 3. Критерії закінчення пошуку**

Пошук завершується, коли або розміри симплекса, або різниці між значеннями функції в вершинах стають достатньо малими. Щоб можна було застосувати ці правила, необхідно задати величину параметра закінчення пошуку.

Реалізація розглянутого алгоритму, основана на обчисленні двох типів:

1) побудова регулярного симплекса при заданій базовій точці і масштабному множнику;

2) обчислення координат відображеної точки. Побудова симплекса є достатньо простою процедурою, оскільки із елементарної геометрії ми знаємо, що при заданих початковій (базовій) точці  $x^{(0)}$  і масштабному множнику  $\alpha$  координати інших  $N$  вершин симплекса в  $N$ -вимірному просторі обчислюються за формулою

$$x^{(i)} = \begin{cases} x_j^{(0)} + \delta_1, & j \neq i, \\ x_j^{(0)} + \delta_2, & j = i, \end{cases} \quad (4.16)$$

для  $i$  і  $j=1,2,3,\dots,N$ .

Приріст  $\delta_1$  і  $\delta_2$ , який залежить тільки від  $N$  і вибраного масштабного множника  $\alpha$ , який визначається за формулою

$$\delta_1 = \left[ \begin{array}{c} (N+1)^{1/2} + N - 1 \\ N \quad 2 \end{array} \right] \alpha, \quad (4.17)$$

$$\delta_2 = \left[ \begin{array}{c} (N+1)^{1/2} - 1 \\ N \quad 2 \end{array} \right] \alpha. \quad (4.18)$$

Зауважимо, що величина масштабного множника  $\alpha$  обирається дослідником, виходячи з характеристики задачі, яка вирішується. При  $\alpha=1$  ребра регулярного симплекса мають одиничну довжину.

Обчислення другого типу, пов'язані з відображенням відносного центра ваги, також являє собою нескладну процедуру. Нехай  $x^{(j)}$  — точка, яка має відобразитись. Центр ваги інших  $N$  точок розташовується в точці

$$x_c = \frac{1}{N} \sum_{\substack{i=0 \\ i \neq j}}^N x^{(i)}. \quad (4.19)$$

Всі точки прямої, які проходять через  $x^{(j)}$  і  $x_c$ , задаються формулою

$$x = x^{(j)} + \lambda(x_c - x^{(j)}). \quad (4.20)$$

Якщо  $\lambda = 0$  отримуємо початкову точку  $x^{(j)}$ , коли  $\lambda = 1$  відповідає центру ваги  $x_c$ . Для того щоб побудований симплекс мав властивості регулярності, відображення повинно бути симплексним. Звідси, нова вершина отримується якщо  $\lambda = 2$ . Таким чином,

$$x_{\text{нов.}}^{(j)} = 2x_c - x_{\text{поперед.}}^{(j)} \quad (4.21)$$

**Приклад 4.2. Обчислення в відповідності з методом пошуку по симплексу**

Мінімізувати  $f(x) = (1-x_1)^2 + (2-x_2)^2$

**Р о з в ' я з а н н я.** Для побудови початкового симплексу треба задати початкову точку і масштабний множник. Нехай  $x^{(0)} = [0,0]^T$  і  $\alpha = 2$ . Тоді

$$\delta_1 = \left[ \begin{array}{c} 3+1 \\ 2 \quad 2 \end{array} \right] \alpha = 1,9318,$$

$$\delta_2 = \left[ \begin{array}{c} 3-1 \\ 2 \quad 2 \end{array} \right] \alpha = 0,5176.$$

Використовуючи ці два параметри, обчислимо координати двох інших вершин симплексу:

$$x^{(1)} = [0 + 0,5176, 0 + 1,9318]^T = [0,5176, 1,9318]^T,$$

$$x^{(2)} = [0 + 1,9318, 0 + 0,5176]^T = [1,9318, 0,5176]^T,$$

яким відповідає значення цільової функції, рівне  $f(x^{(1)})=0,2374$  і  $f(x^{(2)})=3,0658$ . Оскільки  $f(x^{(0)})=5$ , необхідно відобразити точку  $x^{(0)}$  відносно центра ваги двох інших вершин симплекса

$$x_c = \frac{1}{2} \sum_{i=1}^2 x^{(i)} = \frac{1}{2} (x^{(1)} + x^{(2)}).$$

Використовуючи формулу (4.21), отримуємо

$$x^{(3)} = x^{(1)} + x^{(2)} - x^{(0)},$$

$$x^{(3)} = [2,4494, 2,4494]^T.$$

В отриманій точці  $f(x^{(3)})=2,3027$  ми спостерігаємо зменшення цільової функції. Новий симплекс сформований точками  $x^{(1)}$ ,  $x^{(2)}$  і  $x^{(3)}$ . У відповідності з алгоритмом слід відобразити точку  $x^{(2)}$ , якій відповідає найбільше значення цільової функції, відносно центра ваги точок  $x^{(1)}$  і  $x^{(3)}$ . Ітерації продовжуються до тих пір, поки правило 1, 2 і 3 стане непотрібним.

Викладений вище алгоритм  $S^2$ -методу має декілька переваг.

1. Розрахунки і логічна структура методу відрізняються простотою, і, звідси, відповідна програма для ЕОМ є відносно короткою.

2. Рівень вимог до обсягу пам'яті ЕОМ невисокий, масив має розмірність  $(N+1, N+2)$ .

3. Використовується відносно невелике число раніше визначених параметрів: масштабний множник  $\alpha$ , коефіцієнт зменшення множника  $\alpha$  (якщо застосовується правило 2) і параметри закінчення пошуку.

4. Алгоритм є ефективним навіть в тих випадках, коли похибка обчислення значень функції велика, оскільки при його реалізації оперують найбільшими значеннями функції у вершинах, а не найменшими.

Алгоритм має також ряд істотних недоліків.

1. Можливе виникнення ускладнень, зв'язаних з масштабуванням, оскільки всі координати вершин симплекса залежать від одного і того ж масштабного множника  $\alpha$ . Щоб уникнути ускладнень такого виду, в

практичних задачах слід промасштабувати всі змінні з тим, щоб їх значення були рівні за величиною.

2. Алгоритм працює дуже повільно, оскільки отримана на попередніх ітераціях інформація не використовується для прискорення пошуку.

3. Не існує простого способу розширення симплекса, що не вимагає переліку значень цільової функції в усіх точках зразка. Таким чином, якщо  $\alpha$  з будь-якої причини зменшується (приклад, якщо зустрічається область з вузьким «яром» чи «хребтом»), то пошук повинен продовжуватися зі зменшеною величиною кроку.

Модифікована процедура пошуку по симплексу, розроблена Нелдером і Мідом [3, 5, 6], частково вилучає деякі із перерахованих недоліків.

Неважко показати, що хоча формула для визначення вершин регулярного симплекса є дуже зручною при побудові початкового зразка, але основ для збереження властивостей регулярності симплекса в процесі пошуку немає.

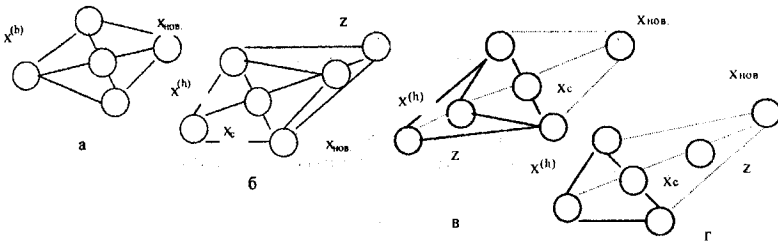


Рисунок 4.6 – Розтягування та стискування симплекса

а – нормальне відображення ( $\theta = a = 1$ ),  $f^{(l)} < f(x_{нов.})$ ;

б – розтягнення ( $\theta = \gamma > 1$ ),  $f(x_{нов.}) < f^{(l)}$ ;

в – стискування ( $\theta = \beta < 0$ ),  $f(x_{нов.}) > f^{(g)}$ ,  $f(x_{нов.}) \geq f^{(h)}$ ;

г – стискування ( $\theta = \beta > 0$ ),  $f^{(g)} < f(x_{нов.}) < f^{(h)}$ .

Звідси, при відображенні симплекса існує можливість як його розтягнення, так і стискування. Для розрахунків за методом Нелдера і Міда використовується вершина симплекса  $x^{(h)}$  (якій відповідає найбільше значення цільової функції  $f^{(h)}$ ),  $x^{(g)}$  (якій відповідає наступне за величиною значення цільової функції  $f^{(g)}$ ) і  $x^{(l)}$  (якій відповідає найменше значення цільової функції  $f^{(l)}$ ). Нагадаємо, що відображення вершини симплекса здійснюється по прямій.

$$\begin{aligned}x &= x^{(h)} + \lambda(x_c - x^{(h)}), \\x &= x^{(h)} + (1 + \theta)(x_c - x^{(h)}).\end{aligned}\tag{4.22}$$

Якщо  $\theta = 1$  має місце нормальне відображення симплекса, оскільки точка  $x_{\text{нов.}}$  лежить на відстані  $x_c - x^{(j)}$  від точки  $x_c$ . Якщо  $-1 \leq \theta < 1$ , то спостерігається стиснуте відображення, чи *стискування* симплекса, тоді як вибір  $\theta > 1$  забезпечує розтягнуте відображення, чи *розтягування* симплекса. На рис. 4.6 наведені можливі варіанти відображення. Три значення параметра  $\theta$ , які використовуються при нормальному відображенні, які позначаються через  $\alpha, \beta, \gamma$  відповідно. Реалізація методу починається з побудови початкового симплексу і визначення точок  $x^{(h)}, x^{(g)}, x^{(l)}$  і  $x_c$ . Після нормального відображення здійснюється перевірка значень цільової функції за критерієм закінчення пошуку в точці відображеного симплекса. Якщо пошук не завершений, то за допомогою тестів, наведених на рис. 4.6, обирається одна із операцій — нормальне відображення, розтягування чи стискування. Ітерації продовжуються, поки зміни значень цільової функції в вершинах симплекса не стануть незначними. В якості задовільних значень параметрів  $\alpha, \beta, \gamma$  Нелдер і Мід рекомендують [2] використовувати  $\alpha = 1, \beta = 0,5, \gamma = 2$ .

Цей алгоритм більш загальний і забезпечує локальне перетворення простору проектування, при якому досягається мінімум унімодальної функції вигляду

$$M = F(x_1, x_2, \dots, x_N).\tag{4.23}$$

Схема алгоритму методу Нелдера-Міда представлена на рис.4.7. Виконується він таким чином. Спочатку в просторі проектування вводять вхідний симплекс і обчислюють значення цільової функції в його вершинах. З отриманих значень вибирають гірше  $P_{\text{гір}}$  і краще  $P_{\text{кращ}}$ . Після цього визначають положення центра симплекса  $P$ , виключаючи найгіршу точку. Відображену точку  $P^*$  знаходять за формулою

$$P^* = (1 + \alpha)P - \alpha P_{\text{гір}},\tag{4.24}$$

де  $\alpha$  — додатна стала, яка називається коефіцієнтом відбивання. Якщо точці  $P^*$  відповідає краще з найдених значень цільової функції, то наступну точку визначають за формулою

$$P^{**} = \gamma P^* + (1 - \gamma)P, \quad (4.25)$$

де  $\gamma$  — заздалегідь вибрана стала, яка називається коефіцієнтом розширення (завжди більше одиниці). Якщо значення цільової функції в точці  $P^{**}$  краще її значень в усіх попередніх точках, то його використовують замість колишнього найкращого значення в точці  $P_{\text{крац}}$  і після перевірки на збіжність процес повторюється. Якщо ж  $P^{**}$  не краще  $P^*$ , то  $P_{\text{крац}}$  замінюють на  $P^*$  і продовжують пошук.

Якщо раніш знайдена точка  $P^*$  не краща за  $P_{\text{крац}}$ , то  $P_{\text{зир}}$  замінюють або колишньою точкою  $P_{\text{зир}}$ , або точкою  $P^*$ , причому вибирають ту з них, в якій цільова функція має краще значення. Після того як вибір зроблений, за формулою

$$P^{**} = \beta P_{\text{зир}} + (1 - \beta)P \quad (4.26)$$

визначають положення точки стискування, причому  $\beta$  — коефіцієнт стискування, значення якого знаходиться між нулем і одиницею. Якщо точка  $P^{**}$  виявляється більш вдалою, ніж  $P_{\text{зир}}$ , то в подальшому нею замінюють  $P_{\text{зир}}$ . Інакше перш ніж продовжити ітерації, всі  $P_i$  замінюють на нові:

$$P_{i,\text{нов}} = \frac{P_{i,\text{ст}} + P_{\text{крац}}}{2} \quad (4.27)$$

Описаний алгоритм достатньо гнучкий, що дозволяє враховувати локальні топологічні властивості поверхні, яка описана цільовою функцією. Симплекси витягуються в напрямку нахилу площин, їхні осі повертаються при зустрічі з западиною на поверхні цільової функції, а поблизу екстремуму вони стискаються. В цілому метод симплексів можна охарактеризувати як винятково ефективний і зручний для виконання на ЕОМ.

Результати окремих чисельних експериментів показують, що метод Нелдера-Міда має достатню ефективність і високу надійність в умовах наявності випадкових збурень чи похибок при визначенні значень цільової функції. В 1969 р. Бокс і Дрейпер [7] стверджували, що цей метод є «найбільш ефективним із всіх методів послідовної оптимізації, які



ми знаємо». У 1972 р. Паркінсон і Хатчинсон [8] досліджували вплив вибору параметрів  $\alpha, \beta, \gamma$  і способу побудови початкового симплекса на ефективність пошуку. Вони встановили, що орієнтація початкового

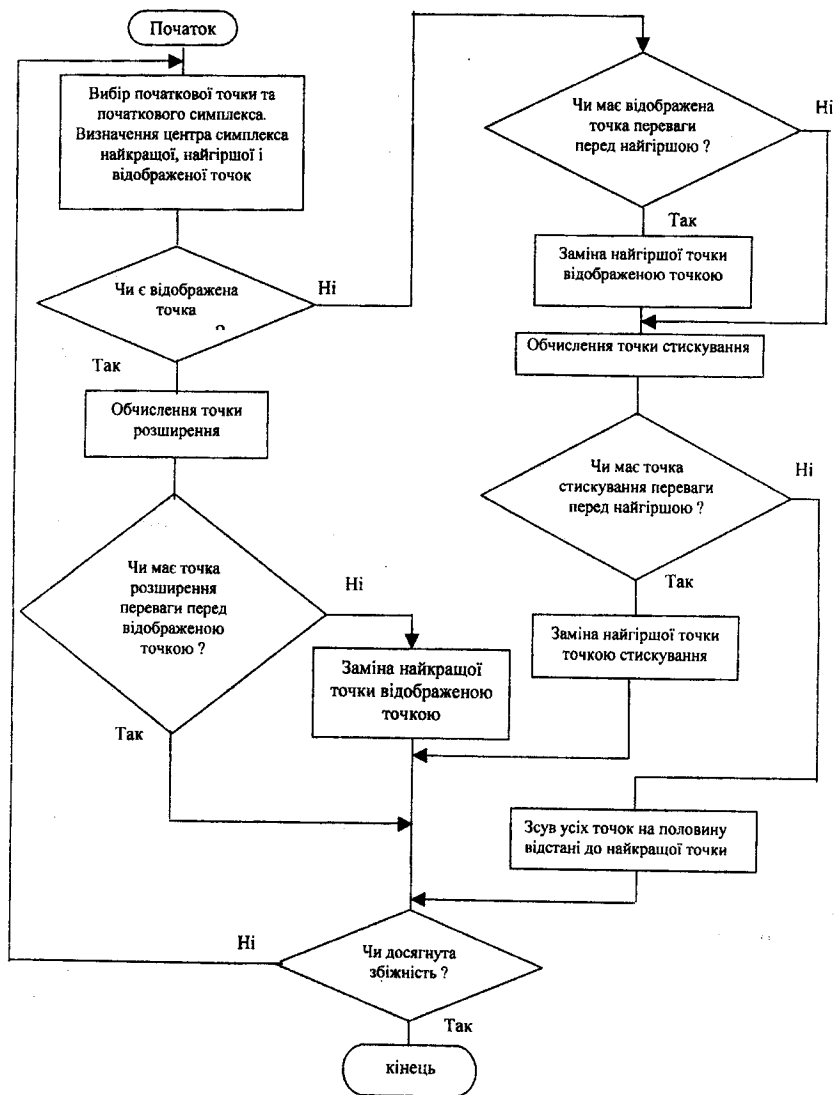


Рисунок 4.7 – Схема алгоритму методу Нелдера-Міда

початкового симплекса на відміну від його форми є важливим фактором, якій впливає на процедуру пошуку, і запропонували використовувати значення параметрів  $(\alpha, \beta, \gamma) = (2; 0,25; 2,5)$ . Такий вибір параметрів дозволить забезпечити добру роботу алгоритма під час повторення послідовних розтягнень симплекса.

#### 4.2.2 Метод покоординатного підйому

Логічним розвитком розглянутої вище методики одновимірного пошуку була б послідовна зміна кожного проектного параметра до тих пір, доки не буде досягнутий максимум цільової функції. По завершенні цієї процедури для всіх змінних можна повернутися до першої і подивитися, чи не можна ще більш удосконалити розв'язання. Цей засіб, який називається методом покоординатного підйому, не завжди дозволяє знайти оптимальний розв'язок. На рис.4.8, *а* показана двовимірна цільова функція, яка підходить для розв'язання задачі цим методом. Її особливість полягає в тому, що лінії рівня близькі за формою до кіл або еліпсів, осі яких паралельні осям координат. Якщо ж ці осі нахилені до осей координат (рис. 4.8, *б*), то ефективність алгоритма знижується, бо для знаходження оптимуму потрібно обчислювати значно більше значень цільової функції. Метод покоординатного підйому не використовується, якщо лінії рівня мають точки перегину (рис. 4.8, *в*). Оскільки лінії рівня такого типу надто часто зустрічаються в інженерній практиці, то перш ніж скористатися вказаним методом, слід переконатися, що задача, яка вирішується, не має подібного недоліку.

Незважаючи на це, метод покоординатного підйому часто використовують на першій стадії розв'язання задачі, застосовуючи після цього більш складні методи.

До переваг методу покоординатного підйому слід віднести можливість використання простих алгоритмів одновимірного пошуку, таких, як метод золотого перерізу.

#### 4.2.3 Метод виключення областей

Знаючи з попередньої глави, наскільки ефективно методи одновимірного пошуку дозволяють скорочувати інтервал невизначеності (одновимірний або двовимірний), можна спробувати застосувати ту ж методику і до багатовимірного простору. Один з найбільш очевидних методів виключення областей називається методом дотичної до лінії рівня, бо в ньому використовуються дотичні до ліній рівня цільової функції.

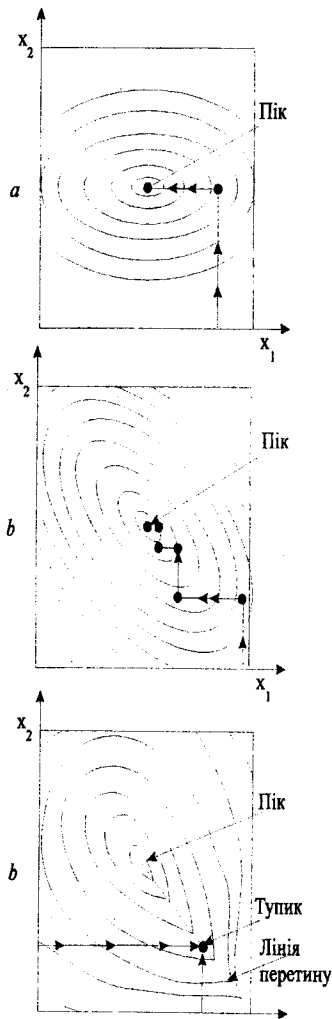


Рисунок 4.8 – Метод покоординатного підйому

Продемонструємо цей метод на прикладі двовимірної цільової функції, лінії рівня якої показані на рис. 4.9. Нехай довільно вибрана точка простору проектування лежить на лінії рівня, яка проходить дещо нижче піка, що відповідає оптимальному розв'язку. Проведемо через цю точку дотичну до лінії рівня. Зробити це неважко, бо дотична повинна лежати в локальному градієнті поверхні цільової функції. Якщо цільова функція



Рисунок 4.9 – Метод виключення областей (дотичною до лінії рівня) у випадку опуклих ліній рівня

достатньо гладка і унімодальна, то дотична до лінії рівня поділить простір проектування на дві частини, в одній з яких імовірність знаходження оптимуму велика, а в іншій мала. Користуючись цим прийомом в кількох вдало вибраних точках, для яких відомі значення цільової функції, можна суттєво звузити область пошуку. Однак виконання цього алгоритму зв'язано з деякими труднощами. Якщо лінії рівня вгнуті, а не опуклі, то може виявитись виключеною область, яка має екстремум (рис.4.10). Крім того, область невизначеності, яка залишилась після виключення, може мати конфігурацію, яка мало придатна для використання інших алгоритмів.

Одним з методів виключення є метод сіткового пошуку, який розробив Мішке і який дає непогані результати. В цьому випадку звужена область невизначеності являє собою гіперкуб – багатовимірний аналог квадрата або куба, – розміри якого можна визначити заздалегідь. Завдяки цьому метод Мішке [5] є одним з небагатьох методів багатовимірного пошуку, ефективність якого піддається вимірюванню. Щоб краще зрозуміти суть цього методу, розглянемо його для випадку простору проектування, який визначається двома змінними.

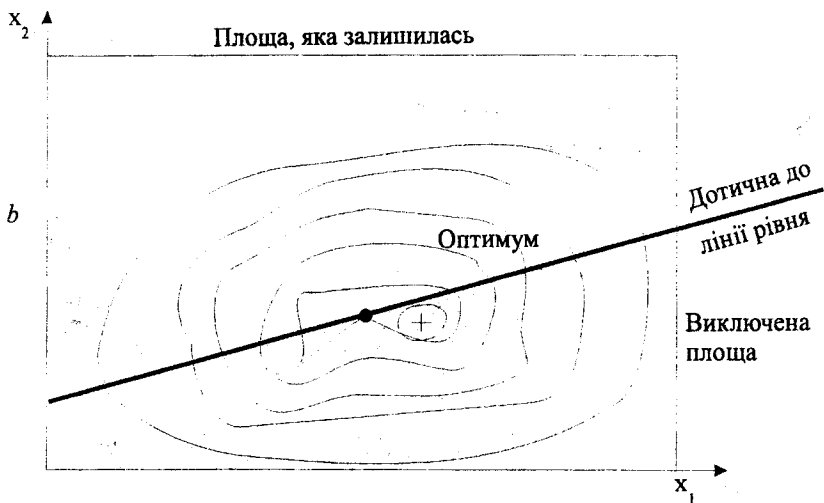


Рисунок 4.10 – Метод виключення областей (дотичною до лінії рівня) у випадку вгнутих ліній рівня

Вхідну область невизначеності в залежності від розмірності простору відобразимо на одиничний квадрат, куб або гіперкуб. Це дозволить вести пошук в нормованій області зі стороною, яка рівна одиниці. В гіперкубі побудуємо сітку, яку утворено попарно симетричними взаємно ортогональними площинами, паралельними координатним напрямкам, вздовж яких змінюються проектні параметри. Ці площини перетинаються по прямих, які в свою чергу перетинаються в точках, що називаються надалі вузлами (рис. 4.11). Обчислимо значення цільової функції в вузлах і в центрі куба. У випадку  $M$  проектних параметрів отримаємо  $2^{M+1}$  значень цільової функції, з яких виберемо найбільше. Приймемо відповідний вузол за центр гіперкуба менших розмірів і продовжимо дослідження. Процес триває до тих пір, доки не буде досягнутий необхідний ступінь звуження інтервалу невизначеності. Якщо в області допустимих значень позначити ступінь звуження вздовж будь-якої осі координат через  $r$ , то лінійне звуження для  $b$ -вимірного гіперкуба буде рівне  $f = r^2$ , а число обчислених значень цільової функції  $N = b(2^M) + 1$ .

Мішке рекомендує вибирати  $r$  в інтервалі значень  $2/3 < r < 1$ . Він відзначає також, що у випадку трьох і більше змінних більшу ефективність забезпечують не кубічні, а зіркоподібні області.

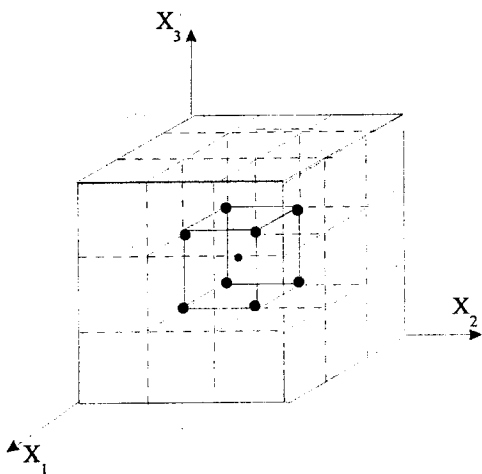


Рисунок 4.11 – Сітковий метод пошуку

#### 4.2.4 Метод випадкового пошуку

Вище в цьому розділі говорилося про громіздкі обчислення у випадку багатовимірного простору на прикладі числа значень цільової функції, які необхідно обчислити, для того, щоб користуючись методом сіток, отримати  $f = 0,1$ , і було показано, що це число зростає як степенева функція, показник степеня якої дорівнює розмірності простору. Оригінальний підхід, який дозволяє обійти цю складність, запропонований Бруксом і оснований на випадковому пошуці. Нехай простір проектування являє собою куб або гіперкуб зі стороною, яка рівна одиниці, і розділений на кубічні комірки шляхом ділення на 10 рівних частин кожної сторони куба, яка відповідає одному з проектних параметрів. Для  $N = 2$  число комірок дорівнює 100, для  $N = 3$  – 1000; в загальному випадку для  $N$  вимірювань число комірок дорівнює  $10^N$ . Ймовірність того, що обрана навмання комірка потрапить до числа 10% найбільш перспективних комірок, дорівнює 0,1, оскільки для  $N=1$  нас буде цікавити одна комірка з 10, для  $N = 2$  – одна з десяти кращих, якщо загальна кількість комірок 100 і т. д. Ймовірність того, що ми пропустимо одну з 10% найбільш перспективних комірок, складе 0,9. Якщо випадково обрати дві комірки, то вірогідність пропуску буде  $0,9^2$ , тобто 0,81. Взагалі ймовірність знаходження хоча б однієї комірки з найбільш перспективних, частка яких рівна  $f$ , після  $N$  спроб складе

$$P = 1 - (1 - f)^N.$$

В таблиці 4.1 вказано, скільки комірок необхідно обрати випадково, щоб забезпечити задану ймовірність при заданій частці найбільш перспективних комірок. З неї видно, що при випадковій вибірці 44 комірок ймовірність досягнення  $f = 0,1$  складе 99%. Це досить непогано, якщо згадати, що для 100%-го забезпечення цільову функцію у випадку п'яти змінних прийшлося би обчислити 2 476 099 разів.

Таблиця 4.1

$f$	Ймовірність			
	0,80	0,90	0,95	0,99
0,1	16	22	29	44
0,05	32	25	59	90
0,01	161	230	299	459
0,005	322	460	598	919

Метод випадкового пошуку має дві переваги:

1. Він підходить для будь-якої цільової функції незалежно від того, унімодальна вона чи ні.

2. Вірогідність успіху при спробах не залежить від розмірності простору, який розглядається.

Хоча цей метод не дозволяє безпосередньо знайти оптимальний розв'язок, він створює хороші умови для застосування в подальшому інших методів пошуку. Тому його часто застосовують в сукупності з одним або декількома методами інших типів.

#### 4.3 Градієнтні методи

В багатьох алгоритмах багатовимірної оптимізації так чи інакше використовується інформація про градієнти. Проілюструємо це положення таким простим прикладом. Уявимо собі, що альпіністу зав'язали очі і сказали, що він повинен добратися до вершини "унімодальної" гори. Навіть якщо він не буде нічого бачити, він може це зробити, якщо весь час буде рухатися вгору. Хоча будь-яка стежка, яка веде вгору, в кінці-кінців приведе його до вершини, найкоротшою з них буде найкрутіша, якщо, правда, альпініст не наштовхнеться на вертикальний обрив, який необхідно буде обійти. (Математичним еквівалентом обриву на поверхні, яку створює цільова функція, є ті її місця, де поставлені умовні обмеження). Уявимо поки що, що задача оптимізації не містить обмежень. Пізніше ми включимо їх в схему пошуку. Метод оптимізації, в основі якого лежить ідея руху по найкрутішій стежці, називається *методом*

найшвидшого підйому або найшвидшого спуску. Вектор градієнта перпендикулярний лінії рівня і вказує напрямок до нової точки в просторі проєктування. Відмітимо, що градієнтний метод на відміну від методу дотичної до лінії рівня можна використовувати до будь-якої унімодалної функції, а не тільки до тих, у яких ця властивість явно виражена.

Щоб краще зрозуміти ідею градієнтних методів, більш конкретно зупинимося на властивостях градієнтів. Розглянемо систему незалежних одиничних векторів  $e_1, e_2, e_3, \dots, e_N$ , які направлені вздовж осей координат  $x_1, x_2, x_3, \dots, x_N$ , які є в той же час проєктними параметрами. Вектор градієнта довільної цільової функції  $F = (x_1, x_2, x_3, \dots, x_N)$  має вигляд

$$\frac{\partial F}{\partial x_1} e_1 + \frac{\partial F}{\partial x_2} e_2 + \dots + \frac{\partial F}{\partial x_N} e_N,$$

де частинні похідні обчислюються в точці, яка розглядається. Цей вектор направлений вверх, в напрямку підйому; обернений йому вектор вказує напрямок спуску. Одиничний вектор градієнта часто зображають у вигляді

$$s_1 e_1 + s_2 e_2 + s_3 e_3 + \dots + s_N e_N,$$

де

$$s_i = \frac{\frac{\partial F}{\partial x_i}}{\left[ \sum_{j=1}^N \left( \frac{\partial F}{\partial x_j} \right)^2 \right]^{1/2}}. \quad (4.28)$$

Іноді характер цільової функції буває достатньо добре відомий, щоб можна було обчислити компоненти вектора градієнта шляхом безпосереднього диференціювання. Якщо таким способом частинні похідні отримати неможливо, то можна знайти їх наближені значення в безпосередньому околі точки, яка розглядається:

$$\frac{\partial F}{\partial x_i} = \frac{F(x_1, x_2, \dots, x_i + \Delta, \dots, x_N) - F(x_1, x_2, \dots, x_i, \dots, x_N)}{\Delta}.$$

Тут  $\Delta$  - невелике зміщення в напрямку  $x_i$ . Цю формулу часто називають "наближенням січної". Отриману інформацію про напрямок градієнта можна використовувати по-різному для побудови алгоритма пошуку.

*Постановка задачі оптимізації градієнтними методами:* мінімізація функції  $F(x_1, x_2, x_3, \dots, x_N)$  з  $N$  проєктними параметрами за допомогою ЕОМ розв'язується ітераційними методами. Розв'язок задачі



починається з вибору початкових значень  $x_i^{[0]}$  ( $i = 1, 2, \dots, N$ ), які за звичай визначаються із умов розв'язуваної задачі, і потім будують послідовні наближення, використовуючи ітераційну формулу, яку отримуємо з формули (4.12):

$$x_i^{[j+1]} = x_i^{[j]} + \lambda^{[j]} s_i^{[j]}, \quad (i = 1, 2, \dots, N; j = 0, 1, 2, \dots), \quad (4.29)$$

де  $\lambda^{[j]}$  - величина кроку ітерації по кожному з параметрів  $x_i$ ;

$s_i^{[j]}$  - параметр вибору "напряму", який звичайно визначається за формулою (4.12).

Дана формула забезпечує збіжність досліджуваної функції до деякого розв'язку  $x_k$  якщо  $j \rightarrow \infty$ . Величина кроку  $\lambda^{[j]}$  на кожній  $j$ -ій ітерації визначається одним з методів оптимізації однопараметричної оптимізації, наприклад методом ділення відрізка навпіл або методом "золотого перерізу" або Фібоначчі.

### 4.3.1 Найшвидший підйом з використанням одномірного пошуку

В деяких методах пошуку інформація про градієнт використовується для ведення одновимірного пошуку в напрямку найшвидшого підйому або спуску, причому використовується співвідношення

$$x_{i,нов} = x_{i,ст} + \lambda s_i, \quad (4.30)$$

де  $\lambda$  - величина кроку, значення якого визначаються в напрямку градієнта.

Отримавши одновимірний оптимум в напрямку даного градієнта, знаходять новий градієнт і повторюють процес до тих пір, поки наступні обчислення дозволяють покращувати отриманий результат. Головна перевага цього методу полягає в тому, що параметр  $\lambda$  можна використовувати в якості незалежної змінної для пошуку за методом Фібоначчі, і це забезпечує високу ефективність методу. Інша важлива перевага розглядуваних методів полягає в тому, що вони дозволяють відходити від сідлових точок поверхні, які описуються цільовою функцією (рис. 4.12).

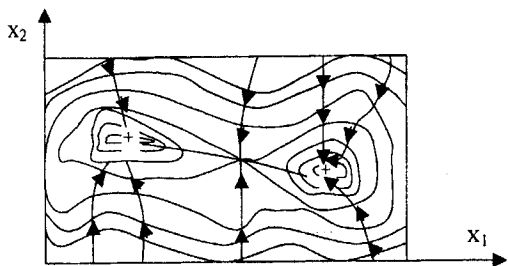


Рисунок 4.12 – Бімодальна цільова функція

Відмітимо, однак, що, як видно з рисунку, для мультимодальних функцій градієнтні методи дозволяють знайти лише локальний оптимум. Тому, якщо характер поверхні недостатньо добре відомий, то слід випробувати декілька

початкових точок і переконатися, що в усіх випадках отримується один і той ж оптимальний розв'язок. Іншою причиною, яка знижує ефективність градієнтних методів, являються злами ліній рівня цільової функції. Оскільки такі точки відповідають розриву в нахилі лінії контуру, то тут можливі помилки в визначенні напрямку подальшого пошуку. Тому пошук може уповільнитися і йти зигзагами поперек лінії зламу, а час необхідний для отримання розв'язку, буде настільки великим, що рахунок прийдеться припинити. В дійсності більшість досліджуваних поверхонь має одну або більше ліній зламу, які нерідко проходять через точку оптимуму. Тому, нашоствхнувшись на лінію зламу, слід далі рухатися вздовж неї.

### 4.3.2 Метод найшвидшого спуску

Даний метод оснований на використанні ітераційної формули

$$x_i^{[j+1]} = x_i^{[j]} + \lambda^{[j]} s_i^{[j]},$$

де  $s_i^{[j]} = -\frac{\partial F}{\partial x_i}$ , причому усі похідні обчислюються якщо  $\lambda_i = x_i^{[j]}$ ;

$\lambda^{[j]}$  - величина кроку, значення якого змінюється (зменшується або обчислюється) методом половинного ділення.

#### Алгоритм методу найшвидшого спуску:

1. Обираємо початкові значення координат вектора  $x^0 = (x_1^{(0)}, x_2^{(0)}, \dots, x_N^{(0)})$  і початкові значення кроку ітераційного процесу  $\lambda$ , які звичайно обираються з умов розв'язуваної конкретної задачі. Хоча загальних правил вибору  $x^0$  немає, однак якщо є додаткова інформація

ро область розташування мінімуму цільової функції, то  $x^0$  обираємо в цій області.

2. Задаємо номер ітерації  $k = 1$ .

3. Обчислюємо значення цільової функції в точці з координатами  $x^0$ .

4. Обчислюємо значення градієнта  $s_k$ .

5. Обчислюємо норму вектора градієнта NG за формулою (4.28).

6. Якщо  $|NG| < \text{заданої } \varepsilon$ , то ітераційний процес закінчується і оптимум знайдений.

7. Якщо умова  $|NG| < \varepsilon$  не виконується, то визначаються нові координати вектора  $x^1$ , які отримуються під час руху до мінімуму цільової функції з кроком  $\lambda$  (рис. 4.13).

8. Порівнюємо два значення цільової функції в двох точках з координатами векторів  $x^0$  і  $x^1$  за формулою

$$f(x^1) < f(x^0). \quad (4.31)$$

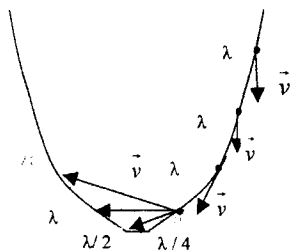


Рисунок 4.13 – Послідовність руху до мінімуму з заданим кроком  $\lambda$

9. Якщо умова не виконується, то крок був вибраний невірно, тобто з цим кроком перескочили через оптимум і крок потрібно зменшити, наприклад в два рази  $\lambda = \frac{\lambda}{2}$  і переходимо до пункту 7 (рис. 4.13).

10. Якщо умова (4.31) виконується, то запам'ятовуємо координати вектора  $x^1$  і переходимо до

пункту 4.

Схема алгоритму описаного методу зображена на рис. 4.14.

### 4.3.3 Метод Флетчера-Рівса

Цей метод дозволяє знайти мінімум нелінійної цільової функції багатьох змінних виду

$$M = F(x_1, x_2, x_3, \dots, x_N)$$

при відсутності обмежень. Метод заснований на застосуванні частинних похідних цільової функції по незалежних змінних і перевизначений для дослідження унімодальних функцій. За його допомогою можна досліджувати і мультимодальні функції, однак в цьому випадку слід брати

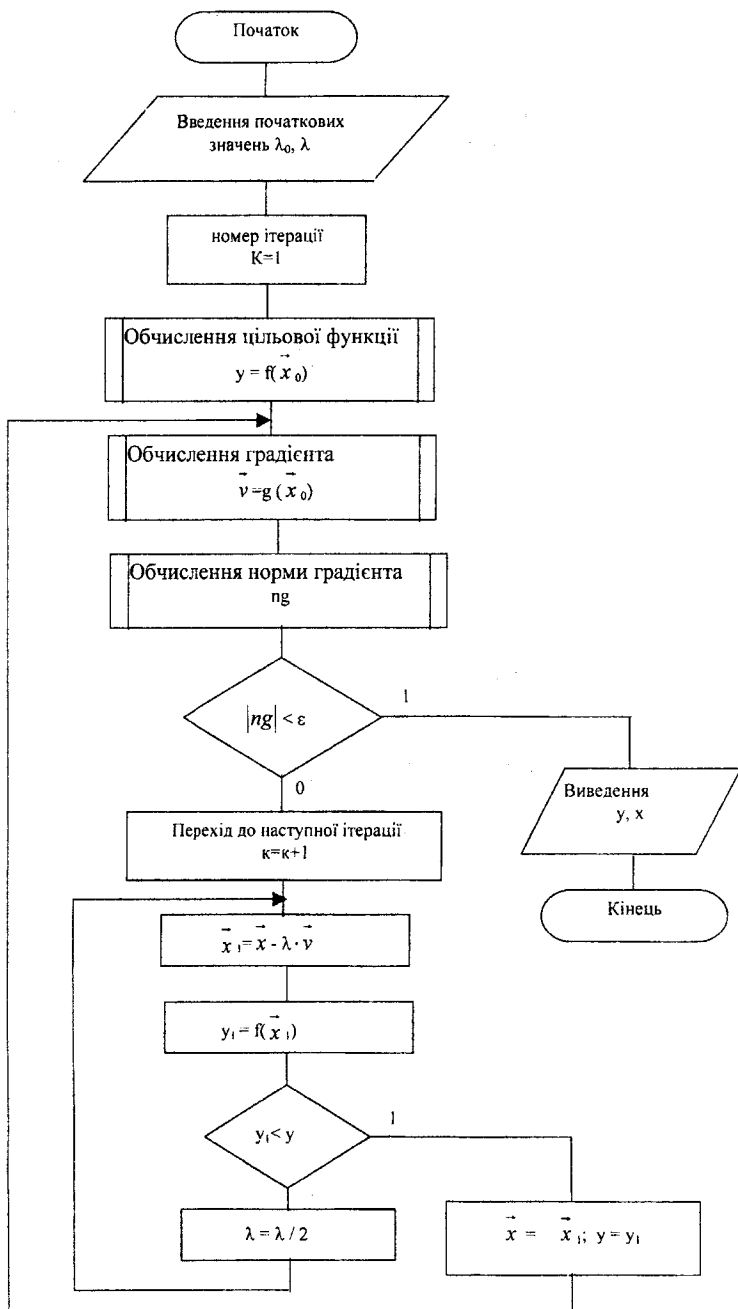


Рисунок 4.14 - Схема алгоритму методу найшвидшого спуску

декілька вхідних точок і перевіряти, однаково чи в усіх випадках зв'язання. Схема алгоритму методу Флетчера-Рівса зображена на рис.4.16. Виконується він таким чином. Спочатку вибирається підходяща початкова точка простору проектування й шляхом обчислення компонент вектора градієнта визначається напрямок найшвидшого спуску. Індекс  $k=1$  відповідає вхідній точці. Після цього в напрямку найшвидшого спуску ведеться одновимірний пошук за формулою

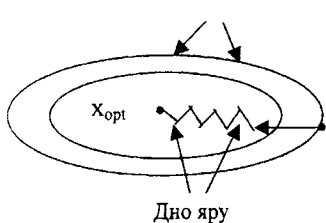


Рисунок 4.15 – Зміна напрямків руху  $s_i$  по дну яру

$$s_i^{(k)} = \frac{-\left(\frac{\partial F}{\partial x_i}\right)^{(k)}}{\left[\sum_{j=1}^N \left(\frac{\partial F}{\partial x_j}\right)^2\right]^{1/2}}, \quad i = 1, 2, \dots, N,$$

$$x_{i,нов} = x_{i,ст} + \lambda s_i, \quad i = 1, 2, \dots, N,$$

де  $\lambda$  – зміщення в напрямку вектора градієнта. Знайшовши мінімум в цьому напрямку, визначають напрямки нових одиничних векторів, які дещо відрізняються від напрямку нового вектора градієнта і являють собою лінійні комбінації вектора градієнта на даному кроці і вектора градієнта, отриманого на попередньому кроці. Нові компоненти одиничних векторів записуються у вигляді

$$x_i^{(k+1)} = \frac{-\frac{\partial F^{(k+1)}}{\partial x_i} + \beta^{(k)} s_i^{(k)}}{\left\{ \sum_{j=1}^N \left[ -\left(\frac{\partial F}{\partial x_i}\right)^{(k+1)} + \beta^{(k)} s_j^{(k)} \right]^2 \right\}^{1/2}}, \quad i = 1, 2, \dots, N, \quad (4.32)$$

де

$$\beta^{(k)} = \frac{\sum_{i=1}^N \left[ \left(\frac{\partial F}{\partial x_i}\right)^{(k+1)} \right]^2}{\sum_{i=1}^N \left[ \left(\frac{\partial F}{\partial x_i}\right)^{(k)} \right]^2}. \quad (4.33)$$

Індекс  $k$  вказує на послідовність обчислень в процесі ітерацій. Нові напрямки називаються «спряженими» і відповідають поточній локальній квадратичній апроксимації функції, а фактично представляє собою рух по дну яру (рис. 4.15). Після цього по новому напрямку (іншому схилу яру) проводять одновимірний пошук і, знайшовши мінімум, перевіряють, чи досягнутий необхідний ступінь збіжності. Якщо перевірка показує, що це так, то рахунок припиняється. Інакше визначають нові спряжені напрямки,  $k$  збільшують на одиницю і продовжують процес до тих пір, доки не буде забезпечена збіжність або доки пошук не буде проведений по всіх  $N+1$  напрямках. Закінчивши цикл пошуку по  $N+1$  напрямках, починають новий цикл, в якому знову використовується напрямок найшвидшого спуску. Достоїнство цього алгоритму полягає в тому, що він дозволяє використати переваги градієнтних методів, які проявляються під час дослідження цільової функції з перервними похідними. Оскільки  $N+1$  напрямків пошуку другої сукупності відрізняються від напрямків одиничних векторів градієнта, то пошук не «зависає на перегині», а йде вздовж лінії, яка з'єднує точки перегинів лінії рівня, яка, як правило, проходить через точку оптимуму. Взагалі можна стверджувати, що методи, основані на визначенні нових напрямків пошуку на основі накопичених даних про локальну поведінку функції, за своєю природою ефективніші, ніж методи, в яких напрямок пошуку задається заздалегідь. Саме тому метод Флетчера-Рівса має більші переваги у порівнянні з методами найшвидшого спуску або підйому. Його недолік полягає в тому, що, він є складнішим ніж вказані методи, він вимагає розробки більш складних програм.

#### 4.3.4 Метод Девідона-Флетчера-Пауела

Метод Девідона-Флетчера-Пауела [2, 5] являє собою алгоритм оптимізації, пристосований для відшукування безумовного мінімуму цільової функції, яка залежить від декількох змінних і яка має вигляд

$$M = F(x_1, x_2, \dots, x_N) \quad (4.34)$$

Необхідні частинні похідні цільової функції по незалежних змінних. Оскільки в основі методу лежить припущення про унімодальність цільової функції, в тих випадках, коли є підстави припускати, що вона не є такою, потрібно брати декілька вхідних точок. На рис. 4.17 зображена схема алгоритму методу Девідона-Флетчера-Пауела. Алгоритм виконується таким чином. Спочатку в просторі проектування вибирають підходящу початкову точку. Після цього, обчислюючи складові вектора градієнта визначають напрямок пошуку.

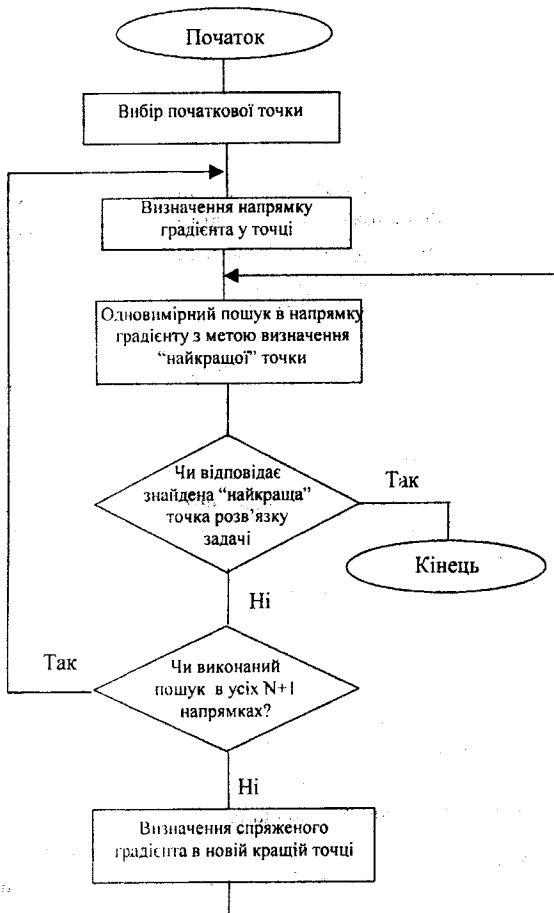


Рисунок 4.16 – Схема алгоритму методу Флетчера-Рівса

$$s_i^{(k)} = \frac{\sum_{j=1}^N H_{i,j} \left( \frac{\partial F}{\partial x_i} \right)^{(k)}}{\left\{ \sum_{l=1}^N \left[ \sum_{j=1}^N H_{l,j} \frac{\partial F}{\partial x_j} \right]^2 \right\}^{\frac{1}{2}}}, \quad i = 1, 2, \dots, N,$$

тут  $k$  – номер ітерації, а  $H_{i,j}$  – елементи симетричної позитивно визначеної матриці розмірності  $N \times N$ . В процесі ітерацій ця матриця

перетворюється в матрицю, зворотну матриці Гессе, елементами якої є другі часткові похідні цільової функції. Оскільки звичайно матриця заздалегідь невідома, то в якості початкової можна скористуватися будь-якою симетричною позитивно визначеною матрицею. Як правило, беруть найпростішу з них – одиничну матрицю. В цьому випадку пошук починається вздовж лінії найшвидшого спуску.

Одновимірний пошук ведеться вздовж вхідного напрямку у відповідності з співвідношенням

$$x_{i,нов} = x_{i,ст} + \lambda S_i, \quad i = 1, 2, \dots, N, \quad (4.35)$$

де  $\lambda$  – величина кроку в напрямку пошуку. Знайшовши одновимірний оптимум, перевіряють результат на збіжність і, якщо вона досягнута, пошук припиняють. Інакше для подальшого пошуку вибирають новий напрямок, причому використовують колишнє співвідношення і нову матрицю  $H$ , що визначається за формулою

$$H^{(k+1)} = H^{(k)} + A^{(k)} - B^{(k)}. \quad (4.36)$$

Елементи матриць  $A^{(k)}$  і  $B^{(k)}$ , які мають розмірність  $N \times N$  обчислюються за формулами

$$A^{(k)} = \frac{\Delta x^{(k)} (\Delta x^{(k)})'}{(\Delta x^{(k)})' \Delta G^{(k)}}, \quad (4.37)$$

$$B^{(k)} = \frac{H^{(k)} \Delta G^{(k)} (\Delta G^{(k)})' (H^{(k)})}{(\Delta G^{(k)})' H^{(k)} \Delta G^{(k)}}, \quad (4.38)$$

де верхнім індексом  $t$  позначені транспоновані матриці, а  $\Delta x^{(k)}$  і  $\Delta G^{(k)}$  – відповідно вектори-стовпці різниць значень  $x_i$  і градієнтів в двох точках.

Вектори-стовпці визначаються виразами

$$\Delta x^{(k)} = x^{(k+1)} - x^{(k)},$$



$$\Delta G^{(k)} = \frac{\partial F^{(k+1)}}{\partial x} - \frac{\partial F^{(k)}}{\partial x}$$

У відповідності з правилами матричного обчислення числівники виразів для  $A^{(k)}$  і  $B^{(k)}$  являють собою матриці розмірності  $N \times N$ , а знаменники є скалярами. Визначивши новий напрямок пошуку, проводять одновимірний пошук і продовжують ітераційний процес. При виконанні алгоритму, який описується, пошук після першої спроби ведеться в тих напрямках, в яких цільова функція в найближчому околі має значення, які наближаються до оптимального.

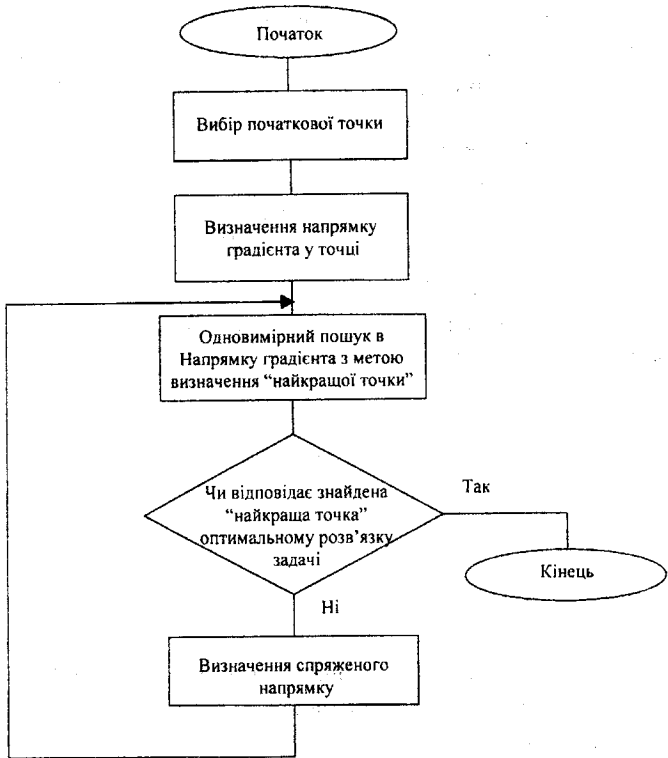


Рисунок 4.17 - Схема алгоритму методу Девідона-Флетчера-Пауела

Лише в окремих випадках ці напрямки збігаються з напрямком градієнта. Тому даний алгоритм часто називають методом «відхиленого» градієнта. Вказана властивість методу Девідона-Флетчера-Пауела дозволяє обходити труднощі, які зв'язані з розривами похідних в просторі

проектування. Широко розповсюджена думка, що цей метод є найефективнішим з усіх градієнтних методів. На відміну від методу Флетчера-Рівса він дає повну інформацію про кривизну поверхні цільової функції в точці мінімуму, однак при цьому вимагається більший об'єм пам'яті і більший час для розрахунків під час обробки матриці  $H$ .

### 4. 3. 5 Метод конфігурацій Хука-Дживса

Цей метод полегшує пошук і не вимагає обчислення похідних [2, 9]. Пошук ведеться вздовж ліній розриву похідних у припущенні, що зміщення в просторі проектування, які виявилися вдалими на ранній стадії пошуку, можуть призвести до успіху і на його більш пізніх стадіях. Метод Хука-Дживса перевизначений для відшукування мінімуму унімодальної функції багатьох змінних

$$M = F(x_1, x_2, \dots, x_N), \quad (4.39)$$

якщо відсутні обмеження. На рис. 4.18 зображена схема алгоритму цього методу. Виконується він таким чином. Спочатку вибирається вхідна базова точка простору проектування і величини кроків, що будуть використані при дослідженні функції. Після цього, у відповідності зі схемою рис. 4.19, проводиться дослідження із заданими приростом у напрямках, відповідних всім незалежним змінним. Там, де отримано уточнене значення функції, розміщують нову тимчасову базову точку. Закінчивши етап дослідження, вибирають нову базову точку і виконують «зрушення схеми». Ця операція полягає в екстраполяції вздовж лінії, що з'єднує нову і попередню базові точки. Відстань зсуву за нову базову точку дещо перевищує відстань між двома попередніми базовими точками. Математично екстраполяція визначається за формулою

$$x_{i,0}^{(k+1)} = x_i^{(k+1)} + \alpha \left( x_i^{(k+1)} - x_i^k \right), \quad (4.40)$$

де  $x_{i,0}^{(k+1)}$  – нова тимчасова базова точка, або «точка зростання»,  $i$  – змінний індекс,  $k$  – порядковий номер стадії пошуку, а  $\alpha$  – коефіцієнт підсилення, що більший або рівний одиниці. Після цього досліджують окіл нової тимчасової базової точки, щоб з'ясувати, чи не містить вона точку, прийнявши яку за наступну базову можна наблизитися до оптимального розв'язку. Цей пошук також ведеться за схемою, яка показана на рис. 4.19. Якщо знайдена тимчасова точка зростання або одна з сусідніх з нею точок має перевагу перед іншими, то вся процедура повторюється з використанням її в якості базової. Завдяки введенню коефіцієнта підсилення, кожне наступне дослідження околу точки

дійсноється на все більшому і більшому віддаленні від вхідної точки до тих пір, доки в процесі пошуку не виявиться пройденим пік або лінія розриву похідної. В цьому випадку вертаються до попередньої «кращої базової точки», звужують область дослідження і повторюють весь процес знову. Якщо крок, який зменшується, послідовно виявляється меншим за деяку задалегідь задану величину і при цьому відсутня помітна зміна значення цільової функції, пошук припиняється. Після декількох змін напрямку пошуку метод Хука-Дживса забезпечує збіг розподілу розрахункових точок з лінією розриву похідних. Звичайно крок, після завершення вибору схеми пошуку зсуву, на кожному наступному кроці збільшується, доки не перевищить величину вхідного кроку в 10 або навіть в 100 разів. Тому у випадку, коли зрушення виявляється невдалим, єдиний засіб продовжити пошук — повернутися до найвдалішої з базових точок і почати усе спочатку. Той факт, що даний алгоритм має властивість «прискорюватися», сприяє підвищенню його загальної ефективності. Друга перевага методу Хука-Дживса – можливість отримання за його допомогою наближеного розв'язку, якість якого безупинно підвищується на всіх стадіях чисельного розв'язання. Особливо явно переваги подібних засобів виявляються при відшуванні екстремумів на гіперповерхнях, які містять глибокі вузькі западини, тобто тоді, коли градієнтні методи неефективні.

#### 4. 3. 6 Метод конфігурацій Розенброка

Метод конфігурацій Розенброка [5, 8] оснований на пошуку мінімуму вздовж ліній розриву похідних і часто виявляється ефективним, коли інші методи не дозволяють отримати розв'язок. Його нерідко називають «методом обертання осей координат», оскільки дослідження в околі вибраної точки ведеться саме таким методом. На відміну від попередніх методів, в яких вхідним змінним надають незалежні прирости, в методі Розенброка система координат повертається так, щоб одна з осей була направлена вздовж лінії розриву похідних, положення якої визначається в результаті попереднього дослідження. Решта осей утворюють з нею ортогональну систему координат. Метод Розенброка оснований на припущенні про унімодальність цільової функції і перевизначений для відшукування мінімуму функції багатьох змінних виду

$$M = F(x_1, x_2, \dots, x_N) \quad (4.41)$$

при відсутності обмежень. На рис. 4.20 показана схема алгоритму, який використовується в цьому методі. Виконується він таким чином. Спочатку вибирають початкову точку, задають початкові величини кроків  $S_i$  ( $i = 1, 2, \dots, N$ ) і обчислюють цільову функцію. Після цього кожній

змінній  $x_i$  надають приріст  $S_i$  в напрямку, паралельному до відповідної осі координат в просторі проектування, і знову обчислюють цільову функцію  $F$ . Якщо її нове значення виявляється меншим за попереднє, то зсув вважається вдалим і наступний крок збільшується у відповідності з формулою

$$S_i = \alpha S_i, \quad (4.42)$$

де  $\alpha > 1$ . Якщо ж нове значення  $F$  виявляється більше за попереднє, то зсув вважається невдалим і наступний крок визначається за формулою

$$S_i = -\beta S_i, \quad (4.43)$$

де  $\beta < 1$ .

Здійснивши зсув по всіх змінних, перевіряють збіжність і, якщо вона досягнута, пошук припиняють. Інакше вводять додаткову перевірку, щоб з'ясувати, чи були зроблені хоча б одне успішне і одне безуспішне зрушення в кожному напрямку. Якщо такого поєднання зрушень не було, вся процедура повторюється, починаючи з першої змінної. Якщо ж подібне поєднання виявлене для кожного напрямку, осі повертають так, щоб вхідний напрямок пошуку збігся з найперспективнішим з раніше розглянутих напрямків. Після цього вибирають нові значення кроків і продовжують пошук по всіх змінних, користуючись новою системою координат.

На відміну від інших, даний алгоритм націлений на відшукання оптимальної точки в кожному напрямку, а не просто на фіксований зсув по всіх напрямках. Величина кроку в процесі пошуку безупинно змінюється в залежності від конфігурації рельєфу поверхні. Поєднання обертання вектора, направленого вздовж лінії розриву похідних, з регулюванням масштабу робить метод Розенброка виключно ефективним для розв'язання складних задач оптимізації.

#### 4.4 Аналіз сучасних пакетів програм для одновимірної та багатовимірної оптимізації

В попередніх параграфах обговорювалися проблеми розробки математичного апарату, який застосовується при розв'язуванні оптимізаційних задач. Але навіть, якщо б усі ці проблеми були успішно розв'язані в теоретичному плані, питання про розробку ефективного інструменту для розв'язування оптимізаційних задач залишалося б відкритим. Справа в тому, що програмна реалізація засобів розв'язання

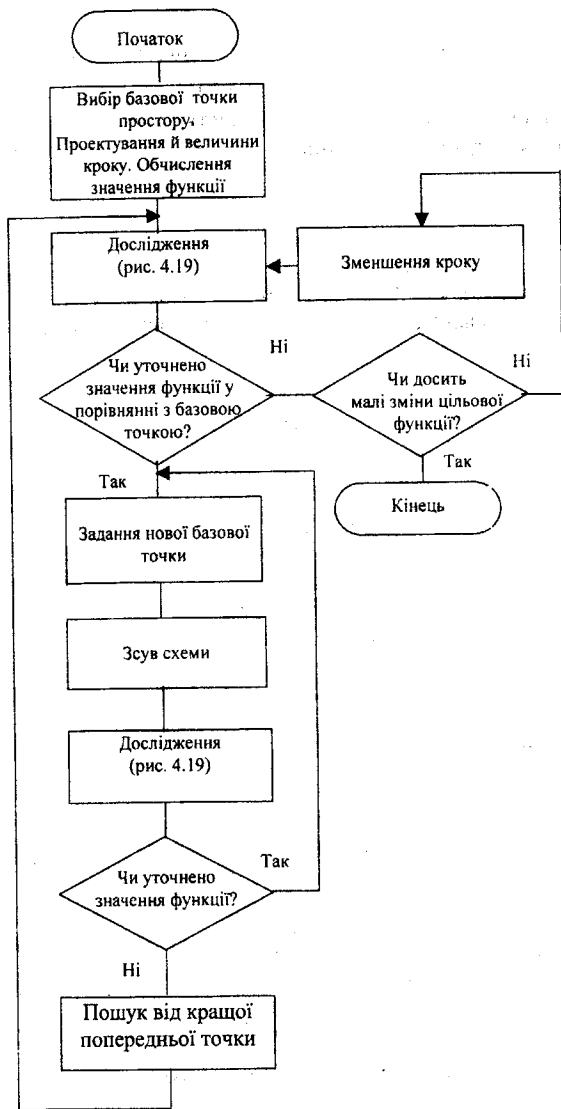


Рисунок 4.18 – Схема алгоритму методу конфігурацій Хука-Дживса

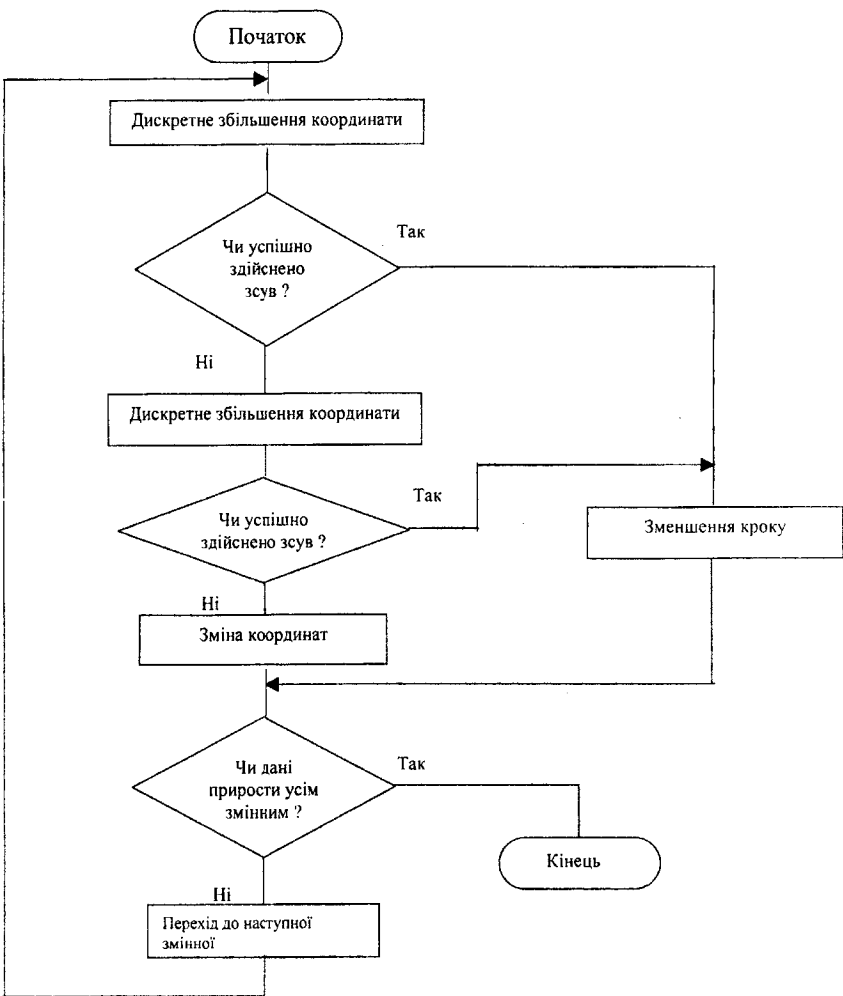


Рисунок 4.19 – Алгоритм дослідження цільової функції на основі методу Хука-Дживса

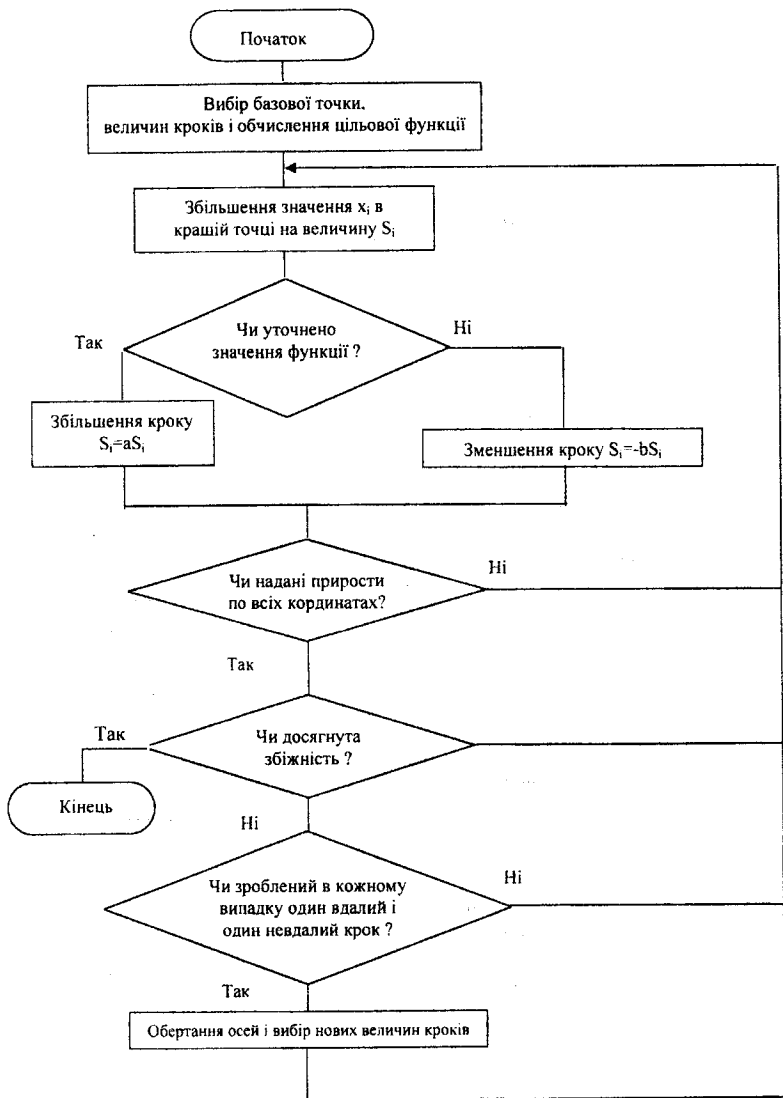


Рисунок 4.20 – Схема алгоритму методу конфігурацій Розенброка

оптимізаційних задач на ЕОМ призводить до виникнення нових не менш цікавих і складних проблем.

Математики розробляють методи оптимізації “на всі випадки життя”. Це природний підхід, оскільки враховування якоїсь специфіки властивостей задачі дозволяє істотно збільшити ефективність роботи методу. Але уся біда в тому, що в реальній ситуації ми, як правило, не знаємо властивостей задачі і тому не можемо зробити найкращий вибір методу розв’язання з набору, який маємо. Виникає природне питання: чом би не автоматизувати процес вибору властивостей задачі і відповідного аналізу методів і не доручити його комп’ютеру? На справді так і робиться в сучасних системах, але сьогодні це, скоріше тенденція, а не завершальний результат у вигляді універсального розв’язку оптимізаційних проблем. Більше того, існує і прямо протилежна тенденція, яка веде до підсилення ролі людини в управлінні процесом розв’язання задачі, при чому ці дві тенденції мирно уживаються одна з одною. Причина цього в великій різноманітності реальних оптимізаційних задач. Для одних кращий перший шлях, для інших – другий.

Для відносно простих задач оптимізації втручання людини в процес розв’язку не потрібне. Діалог в таких ситуаціях разом з тим існує, але його предметом є модифікація постановки задачі, вибір форми видачі результатів розв’язку, їхній аналіз і т. д. В наш час програмними продуктами такого класу являються два найбільш широко застосованих пакета програм MatLab і MathCad, які дозволяють мінімізувати і максимізувати однопараметричні і багатопараметричні функції.

Для мінімізації (максимізації) однопараметричних функцій в цих пакетах програм використовуються алгоритми методу половинного ділення, методу “золотого перерізу” і методу Фібоначчі [5,17,19]. В [17] наводиться порівняння ефективності алгоритмів і рекомендації з їх конкретного використання. Для знаходження локального мінімуму функції від декількох змінних в пакетах програм MatLab і MathCad використовується алгоритм Нелдера-Міда [17], і алгоритми градієнтних методів найшвидшого спуску і спряжених градієнтів.

**Приклад 4.3** Використаємо алгоритм Нелдера-Міда, щоб знайти мінімум функції  $f(x, y) = x^2 - 4x + y^2 - y - xy$  в середовищі MatLab.

Почнемо з трьох вершин:

$$V_1 = (0; 0), \quad V_2 = (1,2; 0,0), \quad V_3 = (0,0; 0,8).$$

Обчислюємо значення функції  $f(x, y)$  в вершинах

$$f(0; 0) = 0,0, \quad f(1,2; 0,0) = - 3,36, \quad f(0,0; 0,8) = - 0,16.$$



Слід порівняти значення функції, щоб визначити  $B$ ,  $G$  і  $W$ :

$$B = (1,2; 0,0), \quad G = (0,0; 0,8), \quad W = (0; 0).$$

Вершину  $W = (0; 0)$  слід замінити. Точки  $M$  і  $R$  мають координати

$$M = \frac{B+G}{2} = (0,6; 0,4) \quad \text{і} \quad R = 2M - W = (1,2; 0,8).$$

Значення функції  $f(R) = f(1,2; 0,8) = -4,48$  менше значення  $f(G)$ . Таким чином, приходимо до випадку (i). Оскільки  $f(R) \leq f(B)$ , рухаємося вправо, і необхідно побудувати вершину  $E$ :

$$E = 2R - M = 2(1,2; 0,8) - (0,6; 0,4) = (1,8; 1,2).$$

Значення функції  $f(E) = f(1,8; 1,2) = -5,88$  менше значення  $f(B)$ , і отримуємо новий трикутник з вершинами

$$V_1 = (1,8; 1,2), \quad V_2 = (1,2; 0,0), \quad V_3 = (0,0; 0,8).$$

Таблиця 4.2 – Логічне рішення для алгоритму Нелдера-Міда

Якщо  $f(R) < f(G)$ , ТО виконується випадок (i) {чи відображення, чи розтягання}

ІНАКШЕ виконуємо випадок (ii) {чи стискання, чи розтягання}

```

BEGIN {випадок (i)}
IF  $f(B) < f(R)$  THEN
    заміна  $W$  на  $R$ 
ELSE
    Обчислення  $E$  і  $f(E)$ 
    IF  $f(E) < f(B)$  THEN
        заміна  $W$  на  $E$ 
    ELSE
        заміна  $W$  на  $R$ 
    ENDIF
ENDIF
END {Випадок (i)}
    
```

```

BEGIN {Випадок (ii)}
IF  $f(R) < f(W)$  THEN
    заміна  $W$  на  $R$ 
    Обчислення  $C = (W + M)/2$ 
    або  $C = (M + R)/2$  і  $f(C)$ 
    IF  $f(C) < f(W)$  THEN
        заміна  $W$  на  $C$ 
    ELSE
        Обчислення  $S$  і  $f(S)$ 
        заміна  $W$  на  $S$ 
        заміна  $G$  на  $M$ 
    ENDIF
ENDIF
END {Випадок (ii)}
    
```

Продовжуємо процес і генеруємо послідовність трикутників, яка збігається до розв'язку в точці (3; 2) (рис. 4.21). В табл. 4.3 наведені значення функції в вершинах трикутника на різних кроках ітерації. Обчислення алгоритму на комп'ютері продовжується до 33-го кроку, коли найкращою вершиною буде  $B=(2,99996456; 1,99983839)$ , причому  $f(B) = -6,99999998$ . Ці значення є наближеними до  $f(3; 2) = -7$ . Основою для зупинки ітерації до обчислення точки (3; 2) є те, що функція плоска біля мінімуму. Значення функції  $f(B)$ ,  $f(G)$  і  $f(W)$  перевірені, і знайдено, що вони усі однакові (це приклад помилки округлення). Алгоритм завершено.

Таблиця 4.3 – Значення функції в різних трикутниках з прикладу 4.3

k	Найкраща точка	Хороша точка	Найгірша точка
1	$f(1,2; 0,0) = -3,36$	$f(0,0; 0,8) = -0,16$	$f(0,0; 0,0) = 0,00$
2	$f(1,8; 1,2) = -5,88$	$f(1,2; 0,0) = -3,36$	$f(0,0; 0,8) = -0,16$
3	$f(1,8; 1,2) = -5,88$	$f(3,0; 0,4) = -4,44$	$f(1,2; 0,0) = -3,36$
4	$f(3,6; 1,6) = -6,24$	$f(1,8; 1,2) = -5,88$	$f(3,0; 0,4) = -4,44$
5	$f(3,6; 1,6) = -6,24$	$f(2,4; 2,4) = -6,24$	$f(1,8; 1,2) = -5,88$
6	$f(2,4; 1,6) = -6,72$	$f(3,6; 1,6) = -6,24$	$f(2,4; 2,4) = -6,24$
7	$f(3,0; 1,8) = -6,96$	$f(2,4; 1,6) = -6,72$	$f(2,4; 2,4) = -6,24$
8	$f(3,0; 1,8) = -6,96$	$f(2,55; 2,05) = -6,7725$	$f(2,4; 1,6) = -6,72$
9	$f(3,0; 1,8) = -6,96$	$f(3,15; 2,25) = -6,9525$	$f(2,55; 2,05) = -6,7725$
10	$f(3,0; 1,8) = -6,96$	$f(2,8125; 2,0375) = -6,956406$	$f(3,15; 2,25) = -6,9525$

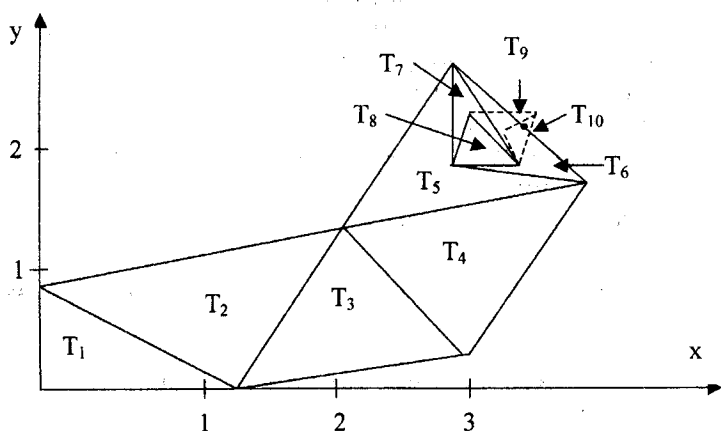


Рисунок 4.21 – Послідовність трикутників  $\{T_k\}$ , яка збігається до точки (3; 2) для методу Нелдера-Міда

В додатку А наводяться приклади програм на мові MatLab для пошуку мінімуму методом золотого перерізу, градієнтним методом найшвидшого спуску і методом мінімізації Нелдера-Міда.

В наш час для розв'язання реальних задач з прийняття рішення в бізнесі, економіці, управлінні підприємствами активно використовується табличний процесор Excel [18, 19].

### Контрольні запитання і завдання

1. Що таке градієнт і антиградієнт? Поясніть математично.
2. В чому полягає сутність градієнтних методів пошуку оптимуму цільової функції.
3. Поясніть узагальнену математичну модель градієнтних методів пошуку оптимуму функції.
4. Покажіть, що функція  
$$f(x) = 3x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 + 2x_2x_3 - 6x_1 - 4x_2 - 2x_3$$
 є опуклою.
5. Знайдіть і класифікуйте стаціонарні точки функції  
$$f(x) = 2x_1^3 + 4x_1x_2^2 - 10x_1x_2 + x_2^2.$$
6. Проведіть аналіз визначеності таких квадратних форм:  
$$f_1(x) = x_1^2 + 2x_2^2 - 3x_3^2 - 6x_1x_2 + 8x_1x_3 - 4x_2x_3,$$
  
$$f_2(x) = 2ax_1x_2 + 2bx_2x_3 + 2cx_3x_1.$$
7. Визначіть компоненти градієнта для функції  
$$f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2,$$
 та використовуючи метод найшвидшого спуску визначіть мінімум функції з заданою  $\varepsilon = 10^{-4}$ .
8. В чому сутність алгоритму методу найшвидшого спуску? Наведіть схему алгоритму методу.
9. Порівняйте алгоритми методу найшвидшого спуску та методу спряжених градієнтів.
10. В чому сутність алгоритму методу Флетчера-Рівса? Наведіть схему алгоритму методу.
11. Визначіть мінімум функції  $f(x) = 4x_1^2 + 3x_2^2 - 4x_1x_2 + x_1$  методом Флетчера-Рівса з заданою  $\varepsilon = 10^{-4}$ .
12. Поясніть сутність методу виключення областей. Наведіть приклади.
13. Що таке випадковий пошук?
14. Наведіть приклад біноміальної цільової функції.
15. Як визначається напрямок пошуку в методі Девідона-Флетчера-Пауелла?
16. Порівняйте метод Флетчера-Рівса і метод конфігурацій Хука-Дживса.

17. На чому ґрунтується метод Розенброка?
18. Як змінюється положення системи координат в методі Розенброка?
19. Поясніть суть симплекс-методу для двовимірного і тривимірного випадку.
20. Що таке симплекс? Дайте графічну інтерпретацію симплекса.
21. Сформулюйте умови виконання яких методом золотого перерізу може не привести до отримання правильного розв'язку.
22. За яких умов задача одновимірної оптимізації не має розв'язку? Наведіть приклади таких задач.
23. Поставте задачу на оптимізацію багатовимірної функції.
24. Мінімізуйте функцію  $f(x) = (1 - x_1)^2 + (2 - x_2)^2$  симплекс-методом.
25. Який критерій використовується для закінчення пошуку в симплекс-методі.
26. Що таке критерій оптимальності? Наведіть приклад.
27. Сформулюйте необхідні і достатні умови існування мінімуму функції декількох змінних.

## 5 ОПТИМІЗАЦІЯ СКЛАДНИХ СИСТЕМ УПРАВЛІННЯ

Найбільш раціональною складною системою (СС) є автоматизована система, використання якої суттєво зменшує кількість обслуговуючого персоналу, знижує вимоги до його кваліфікації, здешевлює вартість обладнання, сприяє швидкому пошуку місць пошкоджень з вказуванням необхідних заходів для усунення їх, а також надає можливість прогнозування надійності систем. Однак, перш ніж перейти до безпосереднього розв'язання задачі з автоматизації процесу функціонування складної системи, необхідно цей процес автоматизувати.

При цьому вирішуються такі основні задачі: на основі вибраного критерію оцінки ефективності якості, оптимізації і готовності систем розділити процес функціонування на його складові, встановити частоту контролю СС, обґрунтувати максимальне число вихідних координат, які найбільш повно характеризують стан СС, вибрати оптимальні алгоритми контролю, пошуку неполадок і відновлення як окремих підсистем, так і процесу функціонування в цілому, оптимізувати процес функціонування СС за узагальненим функціонально-статистичним критерієм, а також за деякими окремими критеріями.

Алгоритм розв'язування задачі включає в себе: усвідомлення проблеми, набір альтернатив і висування гіпотез, доведення оптимальності гіпотези або альтернативи, перевірку розв'язку, повтор і аналіз процесу розв'язування, висновки.

### 5.1 Оптимізація процесу за узагальненими функціонально-статистичним критерієм

Для оптимізації процесу контролю та управління, а також автоматизованих систем контролю та управління (АСКУ) за узагальненим статистичним критерієм можна використовувати метод академіка Л. С. Понтрягіна [20], метод динамічного програмування Белмана [21], або метод лінійного програмування, коли в першому наближенні можна прийняти, що критерій лінійно залежить від заданих параметрів.

Але для практичної оптимізації процесу контролю і управління за узагальненим статистичним критерієм потрібно застосувати метод нелінійного програмування професора Н.І. Андрєєва [24]. Необхідність застосування цього методу обумовлена тим, що узагальнений статистичний критерій є складним нелінійним функціоналом великої кількості параметрів, область зміни яких визначається нелінійними співвідношеннями.

### 5.1.1 Алгоритм оптимізації процесу і АСКУ

Складемо алгоритм оптимізації процесу і АСКУ методом нелінійного програмування за узагальненим статистичним критерієм [25]:

$$\mathcal{E}_K(t, \tau) = \frac{I_{\max}(t, \tau) C_{\min}(t, \tau)}{I_{\max \max}(t, \tau) C_{\Sigma}(t, \tau)} \quad (5.1)$$

1. Подамо узагальнений статистичний критерій  $\mathcal{E}$ , який є функцією апіорної ймовірності виконання системою задачі  $P_0(t, \tau)$ , помилок першого і другого роду  $P_{до}(t, \tau)$ ,  $P_{но}(t, \tau)$ , а також різних складових вартості у вигляді функціоналу  $\Phi$ , деяких функцій  $\varphi_i$ , лінійних форм  $L_i$ .

За лінійні форми виберемо лінійні дільниці ймовірності, тобто приймемо:

$$\left. \begin{aligned} L_1 &= P_0(t, \tau), \\ L_2 &= P_{до}(t, \tau), \\ L_3 &= P_{но}(t, \tau), \\ L_4 &= P(t, \tau), \\ L_5 &= C_i(t, \tau), \end{aligned} \right\} \quad (5.2)$$

або:

$$\begin{aligned} \varphi_1(L_1) &= \sum_{i=1}^m H_{oi}(t, \tau), \\ \varphi_2(L_2, L_3, L_4) &= \sum_{i=1}^m H_i(t, \tau), \\ \varphi_3(L_1, \dots, L_5) &= \sum_{i=1}^m C_i(t, \tau). \end{aligned}$$

Остаточно одержимо:

$$\mathcal{E}(t, \tau) = \Phi(\varphi_1, \dots, \varphi) = \frac{\varphi_1(L_1) - \varphi_2(L_2, L_3, L_4)}{\varphi_3(L_1, \dots, L_5)} \quad (5.3)$$

2. Визначимо область зміни змінних з умови проектування АСКУ, вони можуть мати границі:

$$0,5 \leq L_1 \leq 1$$

$$0,001 \leq L_2 \leq 0,1$$

$$0,001 \leq L_3 \leq 0,01$$

$$0,8 \leq L_4 \leq 1$$

$$L_5 = \min$$

Найбільше значення критерію  $\mathcal{E}(e, \tau)$ , в області  $\mathcal{E}$  змінних  $P_0(t, \tau), P_{10}(t, \tau), P(t, \tau), C(t, \tau)$  збігається з найбільшим значенням функціоналу  $\Phi$  в області  $Q$  зміни змінних  $L_i (i = 1, \dots, 5)$ .

3. Визначимо максимум критерію. При знаходженні максимуму функціоналу  $\Phi(L_1, \dots, L_5)$  необхідно розрізнати два випадки.

**Випадок 1.** Функціонал  $\Phi$ , досягає максимуму всередині області  $Q$ . Необхідною умовою досягнення функціоналом  $\Phi$  максимального значення є рівність нулю частинних похідних:

$$\frac{\Delta \Phi}{\Delta L_i} = 0, \quad i = 1, \dots, 5 \quad (5.4)$$

в деякій точці області  $Q$  простору  $L_i$ . Коли функціонал в деякій частині області  $Q$  немає похідної, тоді деякі з рівнянь системи (5.4) можуть бути замінені умовою:

$$\frac{\Delta \Phi}{\Delta L_i} \quad \text{не існує, } i = 1, 2, \dots, 5.$$

В загальному випадку система рівнянь (5.4) може мати декілька розв'язків, з яких вибирається  $L_i$ , при якому функціонал  $\Phi$  має найбільше значення. Підставимо знайдений розв'язок в (5.2), (5.3) знайдемо остаточно  $P_0, P_{10}, P_{10}, P$ , за яких критерій  $\mathcal{E}(t, \tau)$  досягає максимуму. Цей випадок найчастіше зустрічається на практиці. Однак більш загальним є другий випадок.

**Випадок 2.** Функціонал  $\Phi$  немає максимуму всередині області  $Q$ , а досягає найбільшого значення на границі  $S$ . В цьому випадку визначається границя  $S$  методом лінійного програмування [22], а потім максимум  $\Phi$  на границі  $S$  використанням, наприклад, методу найшвидшого спуску або методу градієнта. Коли функціонал  $\Phi$  має декілька максимумів, тоді знаходиться максимум максимумом.

Таким чином, оптимізація процесу контролю і управління, а також АСКУ зводиться до розв'язання максимальної задачі методом нелінійного програмування.

Для складних АСКУ цю задачу доцільніше вирішувати методом статистичного моделювання на ЕОМ, теорія і практичні доповнення якого розроблені в фундаментальних роботах [26-33].

### 5.1.2 Визначення оптимальних значень деяких параметрів АСКУ

Для визначення оптимального значення ймовірності виконання задачі і складових вартості необхідно розв'язати систему рівнянь відносно цих параметрів:

$$\frac{\partial \mathcal{E}(t, \tau)}{\partial P(t, \tau)} = 0, \frac{\partial \mathcal{E}(t, \tau)}{\partial C_c(t, \tau)} = 0, \frac{\partial \mathcal{E}(t, \tau)}{\partial \Delta C_T(t, \tau)} = 0, \frac{\partial \mathcal{E}(t, \tau)}{\partial \Delta C_G(t, \tau)} = 0, \frac{\partial \mathcal{E}(t, \tau)}{\partial \Delta C_r(t, \tau)} = 0, \quad (5.5)$$

Для визначення максимального значення ймовірності виконання задачі при мінімальному значенні часу необхідно розв'язати рівняння:

$$\frac{\partial T(t, \tau)}{\partial P(t, \tau)} = 0. \quad (5.6)$$

Для визначення максимального значення ймовірності виконання задачі процесу контролю та управління при мінімальній вартості необхідно розв'язати рівняння:

$$\frac{\partial C_c(t, \tau)}{\partial P(t, \tau)} = 0. \quad (5.7)$$

Для оптимізації процесу контролю та управління, а також проектування АСКУ за приведеними статистичними критеріями вибирають один з варіантів процесу і системи, а потім поступово їх параметри доводять до оптимальних або близьких до оптимальних.

## 5.2 Оптимізація процесу і системи контролю та управління

За інтегральну оцінку стану процесу контролю та управління і АСКУ вибираємо ентропію. Ентропія стану об'єкта обумовлена невизначеністю стану різних функціональних систем об'єкта і визначається як сума ентропій об'єкта і системи контролю та управління.



Процес контролю і управління об'єкта є процесом заміщення його невизначеності визначеною кількістю інформації.

З іншої точки зору, процес можливо розглядати як зміну вектора в інформаційному просторі і просторі часу. Алгоритм контролю і управління визначається з умови отримання оптимальної кількості інформації.

Об'єкт вмикається на функціонування при певному рівні ентропії, вибраному з умови отримання заданої ймовірності виконання задачі.

Ентропія процесу контролю та управління і середня кількість інформації є статичними характеристиками процесу контролю. Більший інтерес для практики мають динамічні характеристики: швидкість зміни ентропії, швидкість одержання інформації - пропускна спроможність процесу. Однак статичні і динамічні характеристики окреслюють процес лише з однієї сторони – з точки зору інформаційної спроможності.

Більш повною є узагальнена характеристика процесу контролю, яка оцінює надійність, інформаційність, складність і вартість процесів контролю.

Розглянемо функціонально-статистичні характеристики процесу, які далі дозволяють виводити критерії оцінки ефективності, а також синтезувати процес і пристрої АСКУ за цим критерієм.

### 5.2.1 Основні поняття та означення

Процес контролю об'єкта повинен забезпечити приведення об'єкта в стан готовності з визначеною ймовірністю виконати поставлену перед ним задачу.

Ступінь готовності об'єкта до виконання поставленої перед ним задачі оцінюється за наперед вибраним критерієм оцінки готовності.

З технічної точки зору, процес контролю і управління об'єктом передбачає основні операції контролю і управління і допоміжні операції.

До основних операцій контролю і управління відносяться: операції контролю, пошуку і усунення недоліків, операції налаштування і регулювання апаратури.

Чим вища надійність систем об'єкта, тим менше несправностей виникає в системах об'єкта, тим менш докладно можна проводити контроль цих систем.

До допоміжних операцій контролю та управління відносяться:

- операції відкриття і закриття люків,
- операції установки і зняття пристосувань,
- операції монтажу і демонтажу обладнання,
- операції під'єднання і роз'єднання засобів контролю та управління та ін.

Процес контролю та управління тим кращий, чим менше він має допоміжних операцій або, інакше кажучи, чим менше часу потрібно на їх виконання.

Технічну досконалість процесу контролю і управління з цієї точки зору можна характеризувати коефіцієнтом експлуатаційно-технічної досконалості.

$$K_{TC} = \frac{m_{осн}}{m_{осн} + m_{доп}}, \quad (5.8)$$

де  $m_{осн}$ ,  $m_{доп}$  - математичне очікування сумарного часу, який витрачено відповідно на основні та допоміжні операції під час контролю і управління об'єктом.

Цим коефіцієнтом можна характеризувати ступінь автоматизації. В цьому немає особливих протиріч, адже на практиці, як правило, всі допоміжні операції виконуються вручну.

В повністю автоматичних системах контролю та управління як допоміжні, так і основні операції носять однаковий характер і виконуються автоматично, тому відпадає необхідність поділу операцій.

Коефіцієнтом досконалості (5.8) зручно користуватися при оцінюванні автоматичних систем контролю та управління, а також при порівнянні автоматичних і неавтоматичних систем, коли складність системи контролю і управління не має суттєвого значення.

При оцінюванні досконалості автоматичних систем контролю та управління з урахуванням складності і вартості доцільно використовувати коефіцієнт економіко-технічної досконалості.

$$K_{TC} = \frac{m_{с.осн}}{m_{с.осн} + m_{с.доп}}, \quad \text{де } m_{с.осн}, m_{с.доп} - \text{математичне очікування}$$

сумарної вартості затрат на виконання відповідно основних і допоміжних операцій контролю.

Цей коефіцієнт дозволяє судити про можливість скорочення економічних затрат при введенні автоматизації контролю та управління. Характер його змін збігається з характером коефіцієнта технічної досконалості. Розглянуті оцінки процесу контролю і управління не дозволяють оптимізувати процес, вони можуть бути корисними тільки при вирішенні деяких окремих задач аналізу.

### 5.2.2 Ентропія об'єкта в процесі контролю та управління

В процесі контролю та управління об'єкт спільно з системою контролю та управління має певну невизначеність, кількісною мірою якої є ентропія процесу контролю та управління.

$$H_o(x, t, \tau) = H_{B.O}(x, t, \tau) + H_{П.O}(x, t, \tau),$$

де  $H_{B.O}(x, t, \tau), H_{П.O}(x, t, \tau)$  – ентропія, обумовлена наявністю відповідно раптових і поступових відмов систем об'єкта та систем контролю.

Дійсно, статичний стан об'єкта в будь-який момент часу характеризується ймовірністю виконання цим об'єктом задачі:

$$P(t, \tau) = P_{B.O}(t, \tau)P_{П.O}(t, \tau),$$

де  $P_{B.O}(t, \tau), P_{П.O}(t, \tau)$  – ймовірність відсутності відповідно раптових та поступових відмов.

Ентропія  $H_{B.O}(x, t, \tau)$  обумовлена багатьма незалежними процесами, що не переважають один одного, які відбуваються в системах з різними швидкостями і викликають раптові відмови, випадково розподіленні в часі.

Ентропію, обумовлену наявністю раптових відмов, можна визначити за формулою:

$$H_{B.O}(x, t, \tau) = -\{ P_{B.O}(x, t, \tau) \log_2 P_{B.O}(x, t, \tau) + [1 - P_{B.O}(x, t, \tau)] \log_2 [1 - P_{B.O}(x, t, \tau)] \}, \quad (5.9)$$

де  $P_{B.O}(x, t, \tau) = \prod_{j=1}^n \prod_{i=1}^{m-n} P_{CX}(x_j, t, \tau) P_{CO}(x_i, t, \tau)$  – ймовірність відсутності раптових відмов в об'єкті і АСКУ,  $P_{CX}(x_j, t, \tau), P_{CO}(x_i, t, \tau)$  – ймовірність відсутності раптових відмов відповідно в  $j$ -му пристрої АСКУ і в  $i$ -ій системі об'єкта.

Для експоненціального закону розподілу ймовірності безвідмовної роботи ймовірність того, що в нерезервному об'єкті спільно з АСКУ не станеться раптової відмови, визначається за формулою:

$$P_{B.O}(x, t, \tau) = \int_{-\infty}^{\infty} \dots \int f(z, t, \tau) \exp\left\{-\sum_{i=1}^m \lambda(x_i, t, \tau)\right\} dz, \quad (5.10)$$

де  $\lambda(x_i, t, \tau)$  – інтенсивність відмов  $i$ -ї системи, яка входить в ланцюг систем, що обумовлюють виконання об'єктом поставленої перед ним задачі,  $f(z, t, \tau)$  – закон розподілу інтенсивності зовнішніх подразнень, які діють на системи об'єкта і системи АСКУ.

Якщо позначити через

$$\Lambda(x, t, \tau) = \sum_{i=1}^m \lambda(x_i, t, \tau), \quad (5.11)$$

інтенсивність відмови об'єкта спільно з АСКУ, формулу (5.10) можна подати у вигляді:

$$P_{в.о}(x, t, \tau) = \int_{-\infty}^{\infty} \dots \int f(z, t, \tau) \exp\{-\Lambda(x, t, \tau)t\} dz. \quad (5.12)$$

Таким чином, ентропія, обумовлена наявністю в системах відмов, є функцією часу, яка залежить від умов експлуатації. За певних умов експлуатації і експоненціального закону розподілу ймовірності безвідмовної роботи системи ентропія зростає з часом. Рівень цієї ентропії можна змінювати, змінюючи технологію виготовлення систем, а також умови їх експлуатації. В реальному процесі контролю та управління ентропію, обумовлену наявністю раптових відмов, неможливо виміряти і перетворити в інформацію, її можна тільки заранню врахувати.

Ймовірність відсутності раптових відмов об'єкта і АСКУ, які містять  $m$  об'єктів системи, в тому числі  $\beta$  резервованих вузлів, з врахуванням закону розподілу зовнішніх факторів збурення в умовах контролю і управління можна визначити за формулою:

$$P_{в.о}(x, t, \tau) = \int_{-\infty}^{\infty} \dots \int f(z, t, \tau) \exp\left\{-\sum_{i=1}^m \lambda(x_i, t, \tau)\right\} \prod_{j=1}^m P_{в.о}(x_j, t, \tau) dz, \quad (5.13)$$

де  $P_{в.о}(x_j, t, \tau)$  – ймовірність відсутності раптових відмов  $j$ -ї резервованої системи.

Ентропія  $H_{п.о}(x, t, \tau)$  обумовлена поступовим виходом параметрів систем і різноманітними подразниками. На відміну від ентропії  $H_{в.о}(x, t, \tau)$  ентропія  $H_{п.о}(x, t, \tau)$  може змінюватися включно до рівня ентропії, обумовленої наявністю помилок приладів контролю та управління, може переводитися налагодженням та регулюванням в інформацію, необхідну в кінцевому рахунку для прийняття рішення на запуск об'єкта.

Ентропія  $H_{п.о}(x, t, \tau)$  визначається за формулою в [48]. Ентропії  $H_{в.о}(x, t, \tau)$  і  $H_{п.о}(x, t, \tau)$  безперервно змінюються, як це схематично і умовно показано на рис. 5.1

До початку процесу контролю та управління в момент  $t_0$  ентропія  $H_{в.о}(x, t, \tau)$ , а ентропія  $H_{опт}(x, t, \tau) = H_{п.о}(x, t_0, \tau)$ .

Через певний проміжок часу за відсутності контролю та управління невизначеність стану об'єкта і АСКУ безперервно зростає, внаслідок цього зростає і ентропія. Для простоти розгляду умовно припустимо, що зростання ентропії відбувається за лінійним законом, хоча насправді закони зміни ентропії в часі можуть бути іншими.

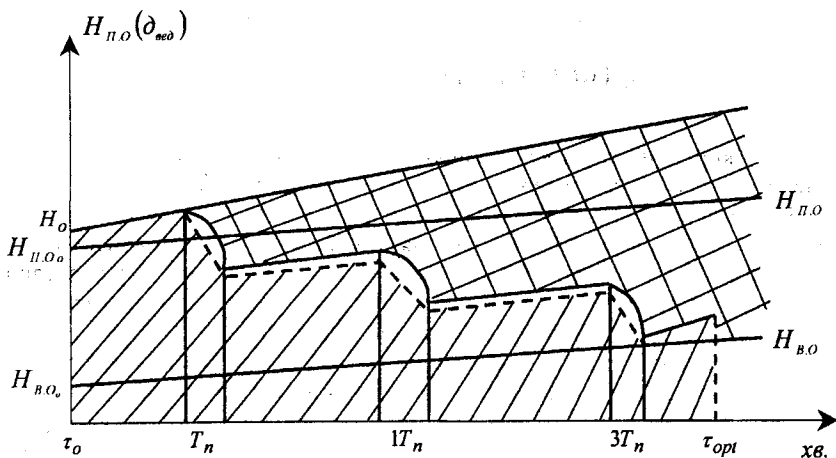


Рисунок 5.1 – Ентропія поступових і раптових відмов системи

В безпервному або в дискретному процесі контролю та управління ентропія стану об'єкта і АСКУ переводиться в певний стан інформації. Оптимальний рівень ентропії  $H_{опт}(x, t, \tau)$  вибирається з умови виконання об'єктом поставленої перед ним задачі.

Ентропію стану об'єкта і АСКУ можна також визначити з використанням статичних характеристик не лише вихідних параметрів, але і перехідної, імпульсної і передаточної функцій, а також частотних систем об'єкта і АСКУ.

Методика визначення ентропії об'єкта і АСКУ через динамічні характеристики аналогічна розглянутим в цьому розділі. Невизначеність стану об'єкта і АСКУ дозволяє розглядати процес контролю та управління як певне джерело інформації.

### 5.2.3 Кількість інформації, одержаної в процесі контролю та управління

Процес контролю та управління об'єктом звичайно розтягнутий в часі і може протікати або безпервно, або дискретно з певною частотою, вибраною з умови достатності часу для зниження невизначеності до заданого рівня при певній інтенсивності контролю та управління. Цей рівень вибирається з умови забезпечення з якоюсь достовірністю дієздатності системи об'єкта на певний час після перевірки, а також з умови збереження ресурсу системи з врахуванням умов їх роботи.

Час проведення перевірок  $t_n$  повинен бути зведений до мінімуму так, щоб за одиницю часу отримати максимальну кількість інформації, яка має найбільшу ймовірність. Але при цьому мінімум не повинен бути менший допустимого часу виведення об'єкта з стану готовності з врахуванням складності АСКУ.

Процес контролю та управління можна розглядати як процес виконання за певним алгоритмом деяких складних  $K$  дослідів  $G_j$  ( $i = 1, \dots, k$ ), ( $j = 1, \dots, l$ ), який складається не більше ніж із  $l$  піддослідів.

В процесі виконання цих дослідів частина невизначеності замінюється інформацією. Кількість інформації, яку одержують в процесі контролю і управління, дорівнює кількості ентропії:

$$I_{G_i}(x, t, \tau) = H_0(x, t, \tau) - H_{G_i}(x, t, \tau), \quad (5.14)$$

де  $H_{G_i}(x, t, \tau) = P_{G_1} H_{G_1}(x, t, \tau) + \dots + P_{G_l} H_{G_l}(x, t, \tau)$  – умовна ентропія стану об'єкта спільно з АСКУ під час здійснення складного дослідів;  $H_{G_i}$ , яка складається з  $l$  піддослідів,  $P_{G_i}$  – ймовірність результату дослідів.  $H_{G_i}(x, t, \tau)$  – ентропія об'єкта після здійснення дослідів.

Одержану кількість інформації в цьому процесі можна збільшити (рис. 5.1) підвищенням точності АСКУ і зменшенням тим самим ентропії  $H_{np}(x, t, \tau)$ , обумовленої наявністю помилок АСКУ, а також покращенням пристосованості алгоритму контролю до одержання інформації і зниження ентропії  $H_{av}(x, t, \tau)$ .

Якщо припинити процес контролю та управління, ентропія об'єкта і АСКУ знову зростає.

Мінімальний рівень невизначеності, який можна досягти в процесі контролю та управління,  $H_{\min}(x, t, \tau) = H_{uo}(x, t, \tau) + H_{np}(x, t, \tau) + H_{av}(x, t, \tau)$ .

На практиці процес контролю та управління необхідно проводити до досягнення рівня оптимальної невизначеності  $H_{opt}(x, t, \tau)$ . Цей рівень вибраний з умови забезпечення імовірності і якості виконання об'єктом задачі:

$$H_{opt}(x, t, \tau) = P(x, l_{knt}, \tau_{op}) \log_2 P(x, l_{knt}, \tau_{op}) + [1 - P(x, l_{knt}, \tau_{op})] \log_2 [1 - P(x, l_{knt}, \tau_{op})],$$

де  $P(x, l_{knt}, \tau_{op})$  – ймовірність виконання об'єктом та АСКУ поставленої перед ним задачі з заданою якістю з моменту кінця контролю та управління  $l_{knt}$  і до моменту кінця роботи системи об'єкта  $\tau_{op}$ .

Оптимальний рівень ентропії може знижуватися до нуля в процесі отримання відомостей про виконану об'єктом задачу.

## 5.2.4 Кількість інформації, отриманої в процесі контролю та управління за нормального спільного закону розподілу вхідних та вихідних параметрів

Процес контролю та управління складного об'єкта звичайно зводиться до одночасного паралельного або зсунутого в часі переведення багатовимірної вектора  $X(x, t, \tau)$  в багатовимірний вектор  $Y(y, t, \tau)$  [37]:  $X \rightarrow Y$ . При цьому допускається, що вектор  $x(t, \tau)$  характеризує апіорний стан об'єкта.

Система контролю та управління шляхом формування, вимірювання та регулювання переводить об'єкт в новий апостеріорний стан, який визначається вектором  $Y(t, \tau)$ . Таким чином, головне призначення процесу контролю та управління з цієї точки зору полягає в тому, щоб перевести невизначений випадковий процес  $X(t, \tau)$  в більш визначений  $Y(t, \tau)$ .

До початку контролю та управління, а також після його завершення випадкові процеси  $X(t, \tau)$ ,  $Y(t, \tau)$  протікають в часі до певної міри однорідно і можуть в певному наближенні вважатися стаціонарними, адже окремі реалізації процесу коливаються навколо постійних в часі середніх значень з постійною в часі дисперсією коливань.

Найбільш цікаві для практики випадки, коли відмічається точне, але з запізненням на  $\tau$ , відтворення стаціонарного складного процесу:

$$Y(y, \dots, t, \tau) = X(x, \dots, t - \tau, \tau) \quad (5.15)$$

або коли вихідний процес відтворюється без запізнення, але діє вектор завад  $\nu$ , який не залежить від вхідного процесу (рис. 5.2):

$$Y(y, t, \tau) = X(x, t, \tau) + \nu(\nu, t, \tau), \quad j = 1, 2, \dots, m. \quad (5.16)$$

Під завадами в цьому випадку можна розуміти шуми каналу зв'язку, помилки приладів, помилки алгоритмів вихідних параметрів та ін.

Середня кількість інформації, яку отримують при переведеннях векторів (5.15), (5.16), можна визначити за формулою:

$$I(X \rightarrow Y, t, \tau) = \int_{-\infty}^{\infty} \dots \int f(X, Y, t, \tau) \log_2 \frac{f(X, Y, t, \tau)}{f(X, t, \tau) f(Y, t, \tau)} dX dY, \quad (5.17)$$

де  $f(X, Y, t, \tau)$  – спільний диференціальний закон розподілу ймовірностей векторів  $X$  і  $Y$ ,  $f(X, t, \tau)$ ,  $f(Y, t, \tau)$  – диференціальний закон розподілу ймовірності відповідно векторів  $X$  і  $Y$ .

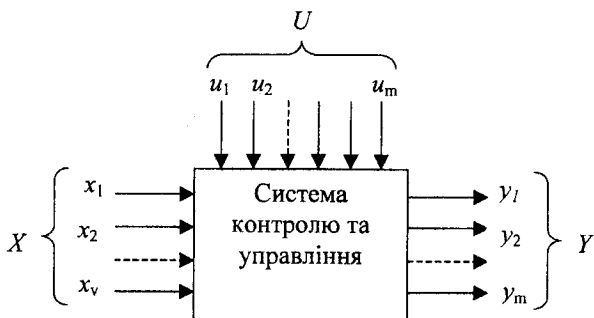


Рисунок 5.2 – Вектори вхідних і вихідних сигналів системи контролю і управління

Виконавши деякі перетворення, формулу (5.17) покажемо в вигляді

$$I(X \rightarrow Y, t, \tau) = H(X, t, \tau) - m_y [H_y(X, t, \tau)], \quad (5.18)$$

або

$$I(X \rightarrow Y, t, \tau) = H(Y, t, \tau) - m_x [H_x(Y, t, \tau)], \quad (5.19)$$

де  $H(X, t, \tau) = -\int_{-\infty}^{\infty} \dots \int f(X, Y, t, \tau) \log_2 f(X, t, \tau) dx$  – апіорна ентропія вхідного вектора  $X$ :

$$m_y [H_y(X, t, \tau) = -\int_{-\infty}^{\infty} \dots \int f(X, Y, t, \tau) \log_2 f_y(X, t, \tau) dx dy \quad (5.20)$$

середнє по вихідних параметрах значення ентропії після одержання сигналів на виході,  $f_y(X, t, \tau)$  – умовний диференціальний закон розподілу вектора  $X$  при одержанні вектора  $Y$ .

При цьому:

$$f_y(X, t, \tau) = \frac{f(x, Y, t, \tau)}{f(Y, t, \tau)}. \quad (5.21)$$

Формули, за якими визначають середню кількість інформації при одновимірній передачі  $X_i \rightarrow Y_i X_i$ , як за виглядом, так і за змістом аналогічні (5.16) – (5.21).

При нормальному законі розподілу випадкових векторів на вході і виході АСКУ



$$f(X, t, \tau) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left\{ -\frac{(X - m_x)^2}{2\sigma_x^2} \right\}, \quad (5.22)$$

$$f(Y, t, \tau) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp \left\{ -\frac{(Y - m_y)^2}{2\sigma_y^2} \right\}. \quad (5.23)$$

Знаходимо :

$$\begin{aligned} \log \frac{f(X, Y, t, \tau)}{f(X, t, \tau) f(Y, t, \tau)} &= \log_2 \frac{1}{\sqrt{1-r^2}} - \frac{r^2 \log_2 l}{2(1-r^2)} \times \\ &\times \left[ \frac{(X - m_x)^2}{\sigma_x^2} - \frac{2(X - m_x)(Y - m_y)}{2\sigma_x \sigma_y} + \frac{(Y - m_y)^2}{\sigma_y^2} \right]. \end{aligned} \quad (5.24)$$

Підставимо (5.22), (5.23), (5.24) в (5.17), отримаємо після деяких перетворень:

$$I(X \rightarrow Y, t, \tau) = \log_2 \frac{1}{\sqrt{1-r_{xy}^2(t, \tau)}} = -\frac{1}{2} \log_2 [1 - r_{xy}^2(t, \tau)], \quad (5.25)$$

де  $r_{xy}(t, \tau) = \frac{1}{\sigma_x \sigma_y} \int_{-\infty}^{\infty} \dots \int (X - m_x)(Y - m_y)(X, Y, t, \tau) -$  коефіцієнт кореляції векторів  $X, Y$ ;

$\sigma_x^2(t, \tau), \sigma_y^2(t, \tau)$  - дисперсії векторів  $X, Y$ .

Неважко показати, що ентропія випадкового  $m$ -вимірною вектора  $X(t, \tau)$ , який має нормальний закон розподілу, може бути подана в вигляді

$$H(X, t, \tau) = \log_2 \sqrt{(2\pi)^m |R^m|}.$$

Середню кількість інформації при передачі  $X \rightarrow Y$  можна виразити через кореляційні матриці

$$I(X \rightarrow Y, t, \tau) = -\frac{1}{2} \log_2 \frac{|R^{xy}|}{|R^x| + |R^y|}, \quad (5.26)$$

$$\text{де } R^x = \begin{vmatrix} R_{11}^{xx} & \dots & R_{1m}^{xx} \\ R_{21}^{xx} & \dots & R_{2m}^{xx} \\ \dots & \dots & \dots \\ R_{m1}^{xx} & \dots & R_{mm}^{xx} \end{vmatrix},$$

$$R^{yy} = \begin{vmatrix} R_{11}^{yy} & \dots & R_{1m}^{yy} \\ R_{21}^{yy} & \dots & R_{2m}^{yy} \\ \dots & \dots & \dots \\ R_{m1}^{yy} & \dots & R_{mm}^{yy} \end{vmatrix},$$

$$|R^{xy}| = \begin{vmatrix} R_{11}^{xx} & \dots & R_{1m}^{xx} & R_{11}^{xy} & \dots & R_{1m}^{xy} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ R_{m1}^{xx} & \dots & R_{mm}^{xx} & R_{m1}^{xy} & \dots & R_{mm}^{xy} \\ R_{11}^{yy} & \dots & R_{1m}^{yy} & R_{11}^{yx} & \dots & R_{1m}^{yx} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ R_{m1}^{yy} & \dots & R_{mm}^{yy} & R_{m1}^{yx} & \dots & R_{mm}^{yx} \end{vmatrix} \text{ — кореляційні показники, які}$$

відповідають кореляційним матрицям, складеним із кореляційних моментів:

$$\begin{aligned} R_{\partial\mu}^{xx} &= m[(x_0 - m_{x0})(x_\mu - m_{x\mu})] \\ R_{\partial\mu}^{yy} &= m[(y_0 - m_{y0})(y_\mu - m_{y\mu})] \\ R_{\partial\mu}^{xy} &= m[(x_0 - m_{x0})(y_\mu - m_{y\mu})] \end{aligned} \quad (5.27)$$

При незалежних компонентах вектора  $X(t, \tau)$

$$R_{\partial\mu}^{xx} = \begin{cases} \sigma_0^2, & \text{якщо } \partial = \mu; \\ 0, & \text{якщо } \partial \neq \mu, \end{cases} \quad (5.28)$$

а ентропія вектора

$$H(X, t, \tau) = \log_2 \sigma_1 \sqrt{2\pi e} + \dots + \log_2 \sigma_m \sqrt{2\pi e}. \quad (5.29)$$

Система контролю не дає інформації про стан об'єкта, якщо вона подає вихідний вектор  $Y(t, \tau)$  без зв'язку з характеристиками вхідного

вектора  $X(t, \tau)$ . Із теорії статистики відомо, що ступінь зв'язку між випадковими процесами характеризується кореляційною функцією.

Якщо вихідний вектор  $Y(t, \tau)$  не зв'язаний з характеристиками вхідного на АСКУ вектором  $X(t, \tau)$ , то  $R_{yx} = 0$ , отже кількість інформації, одержаної під час контролю,  $I(X \rightarrow Y, t, \tau) = 0$ .

Система контролю спроможна дати як завгодно багато інформації про стан об'єкта, якщо вона подає вихідний вектор  $Y(t, \tau)$  з урахуванням всіх характеристик вхідного вектора  $X(t, \tau)$ . При врахуванні всіх характеристик вектора  $X(t, \tau)$ , коефіцієнт кореляції  $r_{xy}(t, \tau) \rightarrow 1$ .

Для одновимірному випадковому процесу формула (5.26) перетворюється в формулу

$$I(X \rightarrow Y, t, \tau) = -\frac{1}{2} \log_2 [1 - r^2(t, \tau)], \quad (5.30)$$

де  $r(t, \tau) = \frac{R_{11}(t, \tau)}{\sigma_x(t, \tau)\sigma_y(t, \tau)}$  – коефіцієнт кореляції.

Розглянемо найбільш цікаві випадки при передачі інформації (5.15), (5.16), обмежуючись при цьому розглядом одновимірної передачі.

При запізненні:

$$y(t, \tau) = x(t + \tau, \tau),$$

$$r(t, \tau) = \frac{1}{\sigma^2} \int \dots \int [x(t) - m_x][x(t - \tau) - m_x] f(X, Y, t, \tau) dx dy$$

або

$$r(t, \tau) = \frac{R(\tau)}{R(0)},$$

де  $R(\tau)$  – кореляційна функція центрування випадкової величини;

$R(0) = \sigma^2$  – дисперсія.

Кількість інформації, яка передається в цьому випадку

$$I(X \rightarrow Y, t, \tau) = -\frac{1}{2} \log_2 \left[ 1 - \frac{R^2(\tau)}{R^2(0)} \right]. \quad (5.31)$$

Оскільки  $0 \leq R(\tau)/R(0) < 1$  якщо  $\tau > 0$ , то  $\infty > I(X \rightarrow Y, t, \tau) \geq 0$ .

Система контролю не дає практичної інформації про стан об'єкта, якщо вона являє собою вихідний вектор  $Y(t, \tau)$  з необмежено великою затримкою  $\tau$ , відносно протікання вхідного процесу  $X(t, \tau)$ .

Із теорії статистики відомо, що якщо  $\tau \rightarrow \infty$ , то  $R(\tau) \rightarrow 0$ . Із (5.31) випливає, що якщо  $\tau \rightarrow \infty$ , то  $R(\tau) \rightarrow 0$  і  $I(X \rightarrow Y, t, \tau) \rightarrow 0$ .

Система контролю спроможна дати необмежено велику кількість інформації про стан об'єкта, якщо вона являє собою вихідний вектор  $Y(t, \tau)$  при необмежено малій затримці  $\tau$ , з урахуванням всіх характеристик вхідного вектора  $X(t, \tau)$ .

За наявності шумів та розподілів координат:

$$y(t, \tau) = x(t, \tau) + u(t, \tau);$$

$$\sigma_y^2(t, \tau) = \sigma_x^2(t, \tau) + \sigma_u^2(t, \tau);$$

$$r(t, \tau) = \frac{1}{\sigma_x \sigma_y} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x, y, t, \tau) dx dy = \frac{\sigma_x^2}{\sigma_x \sqrt{\sigma_x^2 + \sigma_u^2}} = \frac{1}{\sqrt{1 + \frac{\sigma_u^2}{\sigma_x^2}}}$$

Середня кількість інформації, отримана при передачі  $x \rightarrow y$ :

$$I(x \rightarrow y, t, \tau) = -\frac{1}{2} \log_2 (1 - r^2) = \frac{1}{2} \log_2 \frac{\sigma_x^2 + \sigma_u^2}{\sigma_u^2} \quad (5.32)$$

Система контролю не дає інформації про стан об'єкта але як завгодно малої точності переведення вхідного вектора  $X(t, \tau)$  в вихідний вектор  $Y(t, \tau)$ .

Якщо  $\sigma_u \gg \sigma_x$ ,  $\frac{\sigma_x^2 + \sigma_u^2}{\sigma_u^2} \rightarrow 1$ , отже  $I(X \rightarrow Y, t, \tau) \rightarrow 0$ .

Система здатна дати скільки завгодно інформації про стан об'єкта при будь-якій високій точності переведення вхідного вектора  $X(t, \tau)$  в вихідний  $Y(t, \tau)$ .

Якщо  $\sigma_u \ll \sigma_x$ , то  $\frac{\sigma_x^2}{\sigma_u^2} + 1 \rightarrow \infty$ , то, згідно з (5.28)

$$I(X \rightarrow Y, t, \tau) \rightarrow \infty.$$

Таким чином, середня кількість інформації, отримана в процесі контролю та управління для нормального закону розподілу параметрів та завд залежить від самих завд та точності приладів, що використовуються

для контролю, а також від ступеня зв'язку векторів та затримки відтворення. За законами розподілу вихідних параметрів та завад, які різняться від нормальних, середня кількість інформації, отримана в процесі контролю, визначається за формулою (5.17).

### 5.2. 5 Ентропія та інформація в системах регулювання та управління

Системи регулювання та управління є, як правило, замкнутими (рис.5.3).



Рисунок 5.3 – Структура системи регулювання

Процес, що проходить в таких системах, – джерело інформації. Параметри процесу  $x_1, \dots, x_m$  змінюються системою отримання інформації.

Система контролю з цієї точки зору є системою передачі та обробки інформації з метою вироблення координат управління  $y_1, \dots, y_m$ , координати управління перетворюються виконавчою системою в управлінські  $z_1, \dots, z_m$  дії, які діють на процес управління, змінюючи його ентропію, та виходячи з цього, сам процес в потрібному напрямі.

Контур циркуляції інформації в АСКУ може бути поданий в загальному випадку для довільних законів розподілу в векторній формі

$$X(t, \tau) \rightarrow Y(t, \tau) \rightarrow Z(t, \tau) \rightarrow X(t, \tau) \quad (5.33)$$

або для одновимірної системи

$$x(t, \tau) \rightarrow y(t, \tau) \rightarrow z(t, \tau) \rightarrow x(t, \tau). \quad (5.34)$$

Припустимо, що інформація в період часу  $\tau$ , накопичується, перетворюється, а лиш після цього за інтервал  $\tau$ , один раз виробляється координата управління. Це припущення не суперечить фізичним процесам в реальних системах переробки інформації, адже в реальних системах всі процеси квантовані.

Для отримання загальних кількісних співвідношень при довільних законах розподілу розробляється замкнений контур в місці з'єднання процесу управління з системою отримання інформації та подається на вихід останньої деяке достатньо визначене значення  $X$  від постійного джерела.

Тоді в незамкнутій системі буде реалізована передача

$$X \rightarrow Y \rightarrow Z \rightarrow X'$$

Кількість інформації при передачі  $X \rightarrow X'$

$$I(X \rightarrow X', t, \tau) = H(X', t, \tau) - m_x [H_x(X', t, \tau)], \quad (5.35)$$

де  $H(X', t, \tau)$  - ентропія вектора  $X'$  до постачання вектора  $X$ ;

$m_x [H_x(X', t, \tau)]$  - усереднене по  $X$  значення ентропії після отримання сигналу  $X'$ .

Рівність (5.35) подається у вигляді:

$$I \rightarrow (X \rightarrow X', t, \tau) = H(X, t, \tau) - H_x(X', t, \tau) + S,$$

де  $S = H_x(X', t, \tau) - m_x [H_x(X')]$  - приріст ентропії шумів, приведених до виходу процесу управління.

Замкнемо контур первинного контролю. Тоді на вихід системи отримання інформації будуть діяти координати процесу управління в деякий  $(\nu - 1)$ -й момент; а величиною  $X'$  будуть координати того ж процесу управління в  $\nu$ -й момент, тобто  $X = X_{\nu-1}$ ;  $X' = X_\nu$ .

Рівняння (5.32) для визначення середньої кількості інформації в  $\nu$ -й момент часу в замкненому контурі перепишеться у вигляді

$$I_\nu(X, t, \tau) = H_{\nu-1}(X, t, \tau) + H_\nu(X, t, \tau) + S_\nu(t, \tau), \quad (5.36)$$

де  $H_{\nu-1}(X, t, \tau)$ ,  $H_\nu(X, t, \tau)$  - значення ентропії процесу управління відносно в  $\nu$ -й та  $(\nu - 1)$ -й момент часу.

Таким чином, приріст ентропії в системі, що управляється дискретним чином, під час процесу за інтервал часу рівний різниці приросту ентропії приведених шумів і кількості інформації, що передається в замкненому контурі за той же інтервал часу.

З врахуванням ентропії для початкового моменту часу на базі рівності (5.36) можна записати

$$H_v(X, t, \tau) = H_0(X, t, \tau) - \sum_{i=0}^v I(X, t, \tau) + \sum_{i=0}^v S_i(X, t, \tau), \quad (5.37)$$

тобто ентропія процесу управління дорівнює початковій ентропії приведених шумів мінус сумарна кількість, що передана через контур ( $X_{v-1} \rightarrow Xv$ ).

Для стаціонарного процесу

$$H_v(X, t, \tau) = H_{v-1}(X, t, \tau); I_v(X, t, \tau) = S(X, t, \tau). \quad (5.38)$$

В замкненій системі протікає стаціонарний процес управління, якщо кількість інформації, що передається за крок дискретизації при управлінні, рівне відповідному приросту ентропії приведених шумів.

Дійсно при стаціонарному процесі  $H_v(X, t, \tau) = H_0(X_0, t, \tau)$ .

Згідно з (5.36)  $I_v(X, t, \tau) = S(X, t, \tau)$ .

В замкненій системі відсутнє управління, якщо приріст ентропії приведених шумів у відповідний інтервал часу рівний приросту ентропії процесу контролю і управління.

Для процесу, що не управляється, кількість інформації рівна 0, тому згідно з (5.36)

$$H_v(X, t, \tau) - H_{v-1}(X) = S(X, t, \tau).$$

Тобто приріст ентропії приведених шумів у відповідному інтервалі часу рівний приросту ентропії процесу, що не управляється.

Два процеси ідентичні, якщо їх поточні координати рівні:  $X' = X''$ .

Напевно, що ентропії ідентичних процесів також рівні. Крім того, необхідною (але недостатньою) умовою ідентичності процесів є рівність

$$I_v(X', t, \tau) - S_v(X', t, \tau) = I_v(X'', t, \tau) - S_v(X'', t, \tau). \quad (5.39)$$

Всі ці співвідношення для багатовимірних процесів контролю та управління справджуються для одновимірних.

## 5.2.6 Ентропія степеня вільності процесу контролю та управління

Кожен параметр процесу контролю та управління, незалежний від інших параметрів, дає один степінь вільності процесу. Число степенів вільності процесу рівне числу незалежних вихідних параметрів.

При оптимізації процесу контролю і управління всі степені вільності можна умовно розділити на три види: особливо важливі, основні і допоміжні.

До особливо важливих слід віднести степені вільності, що визначають сталий хід процесу. Відхилення параметрів особливо важливих степенів вільності процесу контролю і управління назвемо небезпечною невизначеністю, кількісною мірою якої є небезпечна ентропія, що визначається одним ступенем вільності за формулою:

$$H(x, t, \tau) = - \int_{-\infty}^{\infty} f(x, t, \tau) \log_2 f(x, t, \tau) dx, \quad (5.40)$$

де  $f(x, t, \tau)$  – закон розподілу вихідного параметра процесу  $x$ , який може викликати катастрофу чи важку аварію.

В процесі контролю і управління складним об'єктом необхідно перш за все слідкувати за тим, щоб небезпечна ентропія не перевищувала певний рівень.

Для оцінки ступеня небезпеки процесу контролю і управління введемо коефіцієнт небезпеки

$$k_0(x, t, \tau) = \frac{H(x, t, \tau)}{H_0(x, t, \tau)}, \quad (5.41)$$

де  $H(x, t, \tau)$  – небезпечна ентропія, що визначається за формулою (5.40).

$H_0(x, t, \tau)$  – рівень ентропії, при якому є ймовірність катастрофи чи аварії.

В практичних випадках  $0 \leq k_0(x, t, \tau) \leq 1$ .

До основних відносяться степені вільності, стан яких визначає стан якості всього процесу. До цих степеней відносяться також і особливо важливі степені вільності. Невизначеність параметрів основних степеней вільності оцінюється як ентропія параметрів кожного степеня вільності окремо, що визначається за формулою, що подібна (5.40), а також сумарною ентропією.

$$H(x, t, \tau) = \sum_{i=1}^m H(x, t, \tau). \quad (5.42)$$



Ступінь готовності об'єкта з урахуванням (5.40), (5.42) можна оцінити коефіцієнтом готовності

$$k_r(x_i, t, \tau) = \frac{H(x_i, t, \tau)}{H_0(x_i, t, \tau)}, \quad (5.43)$$

$$k_r(x, t, \tau) = \frac{\sum_{i=1}^n H(x_i, t, \tau)}{\sum_{i=1}^n H_0(x_i, t, \tau)}, \quad (5.44)$$

де  $H_0(x_i, t, \tau)$  – максимально можлива ентропія  $i$ -го параметра до початку процесу контролю і управління.

До допоміжних відносяться степені вільності, що характеризують допоміжні операції процесу контролю і управління.

Для кількісної оцінки невизначеності параметрів допоміжних степенів вільності застосуємо ентропію параметрів допоміжних степенів вільності  $H_a(x, t, \tau)$ ,  $H_n(x, t, \tau)$ .

Степінь вільності допоміжних операцій можна визначити коефіцієнтами:

$$k_m(x, t, \tau), k_n(x, t, \tau).$$

Ентропія параметрів допоміжних степенів вільності і відповідні коефіцієнти визначаються за формулами, що аналогічні (5.40) – (5.44).

В окремих випадках при оптимізації процесу контролю і управління не цікавляться невизначеністю параметрів кожного степеня вільності окремо, а середньою ентропією, що припадає на один параметр:

$$H_m(X, t, \tau) = \lim_{m \rightarrow \infty} \frac{1}{m} \int \dots \int f(X, t, \tau) \log_2 f(X, t, \tau) dx. \quad (5.45)$$

Вибір степенів вільності процесу контролю і управління являє собою досить важку задачу, яка вирішується на стадії як графічного, так і технічного проектування з урахуванням досвіду експлуатації.

### 5.2.7 Динамічні статистичні характеристики процесу контролю і управління

В попередніх розділах розглядалися в основному статистичні характеристики процесу контролю і управління, хоча і враховувались

зміни цих характеристик в даний час  $t$  і відносно якогось майбутнього моменту часу  $\tau$ .

З практичної точки зору більш інтенсивними є динамічні статистичні характеристики процесу. До динамічних статистичних характеристик процесу контролю і управління слід віднести приріст ентропії і кількості інформації, що припадає на одиницю часу, а також "пропускну здатність" процесу контролю.

Приріст ентропії процесу за одиницю часу характеризується швидкістю зміни невизначеності процесу контролю і управління і оцінюється за формулою:

$$U_H(t, \tau) = \frac{\Delta H_T(x, t, \tau)}{T}, \quad (5.46)$$

де  $\Delta H_T(x, t, \tau)$  - приріст ентропії процесу за інтервал часу  $T$ .

Якщо невизначеність процесу зростає, то приріст ентропії має від'ємний знак, в разі спаду невизначеності процесу - додатний.

Кількість інформації, що отримується за одиницю часу, характеризує швидкодію процесу контролю і управління і визначається за формулою:

$$U_I(t, \tau) = \frac{I_T(t, \tau)}{T}, \quad (5.47)$$

де  $I_T(t, \tau)$  - середня кількість інформації, що отримується за час  $T$ .

Максимальну середню кількість інформації, що отримується за одиницю часу, назвемо пропускну здатністю процесу контролю і управління:

$$C_{nc}(t, \tau) = \frac{I_{Tmax}(t, \tau)}{T}. \quad (5.48)$$

Для нормального розподілу вірогідності значень вихідних параметрів:

$$C_{nc}(t, \tau) = F_s \log_2 \frac{\sigma_x^2 + \sigma_u^2}{\sigma_u^2}, \quad (5.49)$$

де  $F_s$  - смуга "пропускання процесу" контролю і управління або половина дискретних значень вихідних параметрів процесу, які можуть бути виміряні в  $I$  с;  $\sigma_x^2$ ,  $\sigma_u^2$  - дисперсія відповідно вихідних параметрів і шумів процесів.

Таким чином, динамічні статистичні характеристики процесу контролю і управління характеризують його не тільки з точки зору стану об'єкта, що контролюється, точності ведення контролю і управління, але і з точки зору швидкодії процесу.

До динамічних характеристик процесу контролю і управління також можна віднести час, що витрачається на виконання операцій.

### 5.2.8 Критерій оцінки готовності складних об'єктів

Функціонально-статистичну оптимізацію деяких параметрів процесу контролю складного об'єкта можна провести на основі вибраного критерію готовності об'єкта [34, 35].

В процесі контролю стану складного об'єкта і його безвідмовність в роботі оцінюється шляхом безпосередньої або непрямой оцінки характеристик вихідних параметрів, які, зокрема, можуть бути статистичними та динамічними характеристиками об'єкта, а також інші узагальнені показники стану об'єкта.

При оцінці стану об'єкта перш за все:

- встановлюється безвідмовність і працездатність об'єкта з визначенням якісних показників в деякий момент часу  $t_n$ , в який виконується перевірка;
- шляхом прогнозування встановлюється безвідмовність і працездатність об'єкта з визначенням якісних показників на заданий інтервал часу  $\tau_{op}$ .

Якщо об'єкт відмовив, виконують пошук і усунення несправностей.

Для встановлення працездатності об'єкта з визначенням якісних показників в деякий момент часу перевіряється виконання умов:

$$\Delta P(X, X', t_n, \tau) \leq \Delta P_{\text{д}}(X_0, X'_0, t_n, \tau), \quad (5.50)$$

де  $\Delta P(X, X', t_n, \tau)$  - дійсне значення відхилення деякого функціонала в момент часу  $t_n$  від його номінального значення;

$\Delta P_{\text{д}}(X_0, X'_0, t_n, \tau)$  - допустиме значення відхилення функціоналу в момент часу  $t_n$  від номінального значення;  $X$  -  $m$ -вимірний вектор вихідних параметрів об'єкта;

$X'$  -  $k$ -вимірний вектор якісних показників працездатності об'єкта;

$k$  - число якісних показників.

Функціонал  $P(X_0, X'_0, t_n, \tau)$  - деяка узагальнювальна функція вихідних параметрів об'єкта. В якості функціонала можна в загальному випадку вибрати вірогідність виконання об'єктом задачі з певними якісними показниками, в окремих випадках вірогідність надходження деяких узагальнених параметрів (характеристик) або тільки об'єкта і тільки систем управління в заданій заздалегідь області, оскільки при

певному допустимому відхиленні функціонала від номінального значення це можливо методом статистичних і динамічних характеристик об'єкта і систем, що в нього входять

При статистичній оптимізації процесу контролю і управління в якості функціонала можна прийняти ентропію стану об'єкта або більш загальний критерій ефективності. Готовність об'єкта до пуску в цьому випадку необхідно визначити, вирішуючи нерівність:

$$H(t_n, \tau) \leq H_{оп}(t_n, \tau),$$

де  $H(t_n, \tau)$  – ентропія об'єкта після контролю;

$H_{оп}(t_n, \tau)$  – допустима ентропія об'єкта, вибрана в умовах виконання об'єктом задачі з певною вірогідністю для певної якості; або нерівність:

$$\mathcal{E}(t_n, \tau) \geq \mathcal{E}_{оп}(t_n, \tau),$$

де  $\mathcal{E}_{оп}(t_n, \tau)$  – ефективність контролю і застосування об'єкта;

$\mathcal{E}(t_n, \tau)$  – задана ефективність, за якої об'єкт з певною заданою вірогідністю виконує поставлену перед ним задачу.

Для встановлення працездатності об'єкта з певними якісними показниками, починаючи з певного моменту  $t_n$  до моменту  $\tau_{оп}$ , перевіряється виконання умов:

$$\Delta P(X, X', t_n, \tau_{оп}) \leq \Delta P_{д}(X_0, X_0', t_n, \tau_{оп}), \quad (5.51)$$

де  $\Delta P(X, X', t_n, \tau_{оп})$  – дійсні з урахуванням прогнозування значення відхилень функціонала на інтервалі часу  $t_n \dots \tau_{оп}$ ;

$\Delta P_{д}(X_0, X_0', t_n, \tau_{оп})$  – допустимі з урахуванням прогнозування значення відхилень функціонала на інтервалі часу  $t_n \dots \tau_{оп}$ .

Прогнозування функціоналу можна здійснити статистичним методом [33], методом граничних випробовувань [36], а також методами екстраполювання випадкових процесів [37-38]. Статистичний метод оснований на прогнозуванні функціоналів за статистичними даними множини однотипних об'єктів, і тому його застосування для прогнозування функціонала одиничного об'єкта обмежено. Методи "індексу дрейфу", граничних випробовувань [37] дозволяють здійснити прогнозування функціонала для одиничного об'єкта, проте їх застосування також обмежене або порівняно малою точністю, або великою трудністю.

Найбільш цікавим з вказаних методів прогнозування є метод, що був розроблений В. С. Красовським [38]. Цей метод базується на використанні

функції максимальної правдоподібності та може бути використаний для прогнозування функціонала.

Уявімо собі, що прогнозування функціонала здійснюється за параметрами прогнозування  $x_1, \dots, x_m, x_1, \dots, x_m$ , які вимірюються точно та незалежно в деякі нерівно відносні моменти часу  $t_{ij} (j = 1, 2, \dots, n)$ .

Математичне очікування кожного параметра прогнозування може бути подане з будь-якою точністю у вигляді поліномів степеня.

$$x_{ij}(t, \tau) = \sum_{i=0}^{iv} a_{iv} t^i, (v = 1, \dots, m), \quad (5.52)$$

де  $a_{iv}(t, \tau)$  – невідомі не випадкові коефіцієнти.

Уявімо також, що виміряне значення  $v$ -го параметра прогнозування  $X_{ij}^u$  є випадковою величиною, що розподілена за нормальним законом

$$f(x_{ij}^u, t_j) = \frac{1}{\sigma_{x_{ij}^u} \sqrt{2\pi}} \exp \left\{ -\frac{(x_{ij}^u - x_{ij})^2}{2\sigma_{x_{ij}^u}^2} \right\}. \quad (5.53)$$

З врахуванням рівності (5.52) для незалежних  $n$ -вимірних величин густина ймовірності або функція правдоподібності сукупності вимірювань  $x_{ij}^u$  отриманих в момент часу  $j = 1, \dots, n$ :

$$f(x_{i1}^u, \dots, x_{in}^u) = \sigma_{x_{ij}^u}^{-n} (2\pi)^{-n/2} x;$$

$$x_{\text{exp}} \left\{ -\frac{1}{2} \sum_{j=1}^n \left( \frac{x_{ij}^u - \sum_{i=0}^{iv} a_{iv} t^i}{\sigma_{x_{ij}^u}} \right)^2 \right\}. \quad (5.54)$$

В виразі (5.54) необхідно вибрати такі значення коефіцієнта  $a_{iv}$  при яких досягається максимум функції правдоподібності  $f_{\text{max}}$ . Звісно, що максимум досягається при мінімальному значенні показника експоненти:

$$\min \sum_{j=1}^n (x_{ij}^u - \sum_{i=0}^{iv} a_{iv} t^i)^2. \quad (5.55)$$

Коефіцієнти  $a_{i\alpha}$ , що забезпечують мінімум виразу (5.55), визначаємо за формулою:

$$a_{i\alpha} = \frac{\Delta v_i}{\Delta v_i},$$

де  $\Delta v_i$ ,  $\Delta v$  - відповідно головний визначник та визначник, який отримали з  $\Delta v$  заміною  $i+1$ -го стовпця на праві частини системи рівняння:

$$\sum_{j=1}^n t_j^i (a_{v_0} + a_{v_i} t_j + \dots + a_{v_i} t_j^{iv}) = \sum_{j=1}^n x_{v_j}^u t_j^i. \quad (5.56)$$

Рівняння (5.56) отримуємо, взявши частинні похідні від чисельника показника експоненти в виразі (5.54) по кожному з коефіцієнтів  $a_{v_i}$  та прирівнявши ці похідні до нуля.

Значення параметра  $x_{v_i} t_{op} = t_n + T$  - в момент часу визначається з виразу:

$$x_{v_i} (t_{op}) = \sum_{i=1}^{iv} \frac{\Delta v_i}{\Delta v} t_{op}^i.$$

Функціонал в даному випадку вироджується в деякий простір вихідних параметрів  $x_1, \dots, x_m, x'_1, \dots, x'_m$ , що визначають прогнозований стан об'єкта. У випадку виконання нерівності (5.52) вважається, що об'єкт готовий до пуску.

Метод прогнозування з використанням функції правдоподібності, хоча і є цікавим, але його використання пов'язане з деякими труднощами.

Практично, як правило, прогнозування вихідних параметрів необхідно виконувати після одного вимірювання параметра системи рівнянь. Тому не вдається використовувати інформацію попередніх вимірювань. Крім того, цей метод не передбачає обліку всіх випадкових збурень, що діють після вимірювання параметра і в основному таких, що визначають кінцевий стан об'єкта.

Більш прийнятним є метод розв'язання системи диференціальних рівнянь, що описують рух об'єкта, або методи наближених оцінок стану об'єкта.

Розглянемо найбільш загальний метод оцінки готовності об'єкта до пуску за ймовірністю виконання об'єктом задачі з заданими якісними показниками. При відомій ймовірності виконання задачі пусковим обладнанням  $P_{акв}(t, \tau)$  ймовірність виконання задачі об'єктом:

$$P(X, X', t, \tau) = P_{акв}(t, \tau) P_{в.о.}(X, X', t, \tau) P_{п.о.}(X, X', t, \tau), \quad (5.57)$$

де  $P_{B.O.}(X, X', t_n, \tau_\phi) = \prod_{\nu=1}^m \prod_{\mu=1}^k P_{B.O.}(x_\nu, x'_{\nu\mu}, t_n, \tau_{\phi})$  – ймовірність відсутності несподіваних відмов у  $m$  системах управління об'єктом з  $k$  якісними показниками, починаючи з деякого моменту часу  $t_n$ , до моменту закінчення роботи системою управління  $\tau_\phi$ ;

$P_{П.О.}(X, X', t_n, \tau_{op}) = \prod_{\nu=1}^m \prod_{\mu=1}^k P_{П.О.}(x_\nu, x'_{\nu\mu}, t_n, \tau_{op})$  – ймовірність відсутності поступових відмов у  $m$  системах управління об'єктом з  $k$  якісними показниками, починаючи з деякого моменту часу  $t_n$  до моменту закінчення роботи системою управління  $\tau_{op}$ .

При експоненціальному законі розподілу ймовірності безвідмовної роботи ймовірність відсутності несподіваних відмов об'єкта з врахуванням резервування систем визначається за формулою:

$$P_{B.O.}(x_\nu, x'_{\nu\mu}, t_n, \tau_{op}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\zeta_1, \dots, \zeta_k, t_n, \tau_{op}) f(\zeta_1, \dots, \zeta_n, t_n, \tau_{op}) \times \exp\left[-\sum_{\mu=1}^k \lambda(x_{\nu\mu}, t_n, \tau_{op}) t\right] \times \prod_{\mu=1}^k P_{B.O.}(x_{\nu\mu}, t_n, \tau_{op}) d\zeta_1, \dots, d\zeta_k, d\xi_1, \dots, d\xi_n. \quad (5.58)$$

Ймовірність відсутності поступових відмов у  $\nu$ -й системі управління об'єктом з  $k$  якісними показниками визначаються за формулою:

$$P_{П.О.}(x_{\nu\mu}, t_n, \tau_{op}) = \int_{x_{\nu\mu\min}}^{x_{\nu\mu\max}} \dots \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (n) f(x_{\nu\mu}, t_n, \tau_{op}) f(\zeta_1, \dots, \zeta_k, f_n, \tau_{op}) \times f(\xi_1, \dots, \xi_n, f_n, \tau_{op}) dx_{\nu\mu}, \dots, dx_{\nu k}, d\zeta_k, d\xi_1, \dots, d\xi_n, \quad (5.59)$$

де  $f(x_{\nu\mu}, \zeta_1, \dots, \zeta_k, t_n, \tau_{op})$  – спільний диференціальний закон розподілу ймовірностей показників якості роботи системи;

$f(\zeta_1, \dots, \zeta_k, t_n, \tau_{op})$  – спільний диференціальний закон розподілу ймовірностей збурень, що діють на систему в проміжок часу;

$f_n, \dots, \tau_{op}; f(\xi_1, \dots, \xi_n, t_n, \tau_{op})$  – спільний диференціальний закон розподілу ймовірностей внутрішніх збурень.

Для ілюстрації методики оцінки здатності систем управління об'єктом розглянемо приклад [39].

Припустимо, що один із вихідних параметрів об'єкта є деякий якісний показник  $x_{\nu\mu}$  (рис. 5.4). На об'єкт контролю окрім дії  $z(t)$ , що керує, діють зовнішні збурення  $\zeta_1, \dots, \zeta_k$  та внутрішні збурення  $\xi_1, \dots, \xi_n$  у

вигляді випадкового виходу в часі первинних параметрів системи автоматичного управління.

Для оцінки здатності об'єкта до запуску за формулою (5.60) необхідно оцінити закон розподілу  $x_{\text{вп}}$ . Для спрощення припустимо, що малі відхилення первинних параметрів незалежні між собою і від зовнішніх збурень. Крім цього, внутрішні і зовнішні збурення, а також відхилення якісних показників виконання завдання системою мають математичне очікування, рівне нулю, і розподілені за нормальним законом. В цьому випадку для визначення закону розподілу якісних показників необхідно знайти дисперсію цих показників в момент часу  $\tau_{\text{оп}}$ .

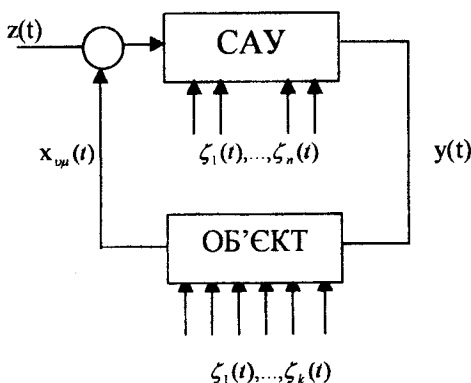


Рисунок 5.4 – Структурна схема системи управління

При незалежних випадкових внутрішніх збуреннях дисперсію відхилення кожного показника якості можна визначити як суму дисперсій, викликаних дією кожного збурення окремо:

$$\sigma^2(x_{\text{вп}}, \xi_1, \dots, \xi_n, t, \tau_{\text{оп}}) = \sum_{i=1}^n \sigma^2(x_{\text{вп}}, \xi_i, t, \tau).$$

З врахуванням дії нестационарних випадкових зовнішніх збурень з кореляційними функціями  $R_i(r, \rho)$ :

$$\sigma^2(x_{\text{вп}}, t, \tau_{\text{оп}}) = \int_0^t \dots \int_0^t \omega_{\zeta_i}(t, r) \omega_{\zeta_i}(t, \rho) R_i(r, \rho) dr, d\rho, \quad (5.60)$$

де  $\omega_{\zeta_i}(t, r)$  та  $\omega_{\zeta_i}(t, \rho)$  – вагомі функції об'єкта, що характеризують передачу сигналу від точки прикладення зовнішніх збурень  $\zeta_i$  на показник якості, що нас цікавить.



При дії малих збурень дисперсію, визначену за формулою (5.60), можна розкласти в ряд Тейлора і обмежитись, з певним наближенням, розглядом перших двох членів:

$$\sigma^2(x_{\text{ви}}, t, \tau) = \sigma_0^2(x_{\text{ви}}, t, \tau) + \sum_{i=1}^n \left[ \frac{\partial \sigma^2(x_{\text{ви}}, t, \tau)}{\partial \zeta_i} \right] \Delta \zeta_i. \quad (5.61)$$

Перший доданок цього виразу характеризує дисперсію при номінальних показниках якості, друге – деякий приріст дисперсії показника якості за рахунок дії збурень.

Припускаючи, що від збурень у виразі (5.61) залежать тільки вагомі функції  $\omega_{\zeta_i}(t, \tau)$  і диференційовані вирази (5.60), отримуємо формулу для визначення частинної похідної:

$$\begin{aligned} \frac{\partial \sigma^2(x_{\text{ви}}, t, \tau)}{\partial \zeta_i} = \\ = \int_0^t \int_0^{\rho} \left[ \omega_{\zeta_i}(t, \tau) \frac{\partial \omega_{\zeta_i}(t, \tau)}{\partial \zeta_i} R_i(\rho, r) + \omega_{\zeta_i}(t, \rho) \frac{\omega_{\zeta_i}(t, \rho)}{\zeta_i} R_i(r, \rho) \right] dr d\rho_i. \end{aligned} \quad (5.62)$$

Враховуючи відомі властивості кореляційної функції формулу (5.57) перепишемо у вигляді  $R_i(r, \rho) = R_i(\rho, r)$ ,

$$\frac{\partial \sigma^2(x_{\text{ви}}, t, \tau)}{\partial \zeta_i} = 2 \int_0^t \int_0^{\rho} \frac{\partial \omega_{\zeta_i}(t, \tau)}{\partial \zeta_i} \omega_{\zeta_i}(t, r) R_i(r, \rho) dr d\rho_i. \quad (5.63)$$

Таким чином визначення частинних похідних зводиться до визначення функціоналів виду (5.63) яке можна провести заздалегідь.

При  $n$  внутрішніх збуреннях та  $k$  зовнішніх для повного розв'язку задачі необхідно розрахувати  $nk$  функціоналів. Методика розрахунку подібних функціоналів та неперервних моделях та ЦВМ достатньо розроблена.

Для оцінки ймовірності виконання задачі об'єктом за формулою (5.60) необхідно знайти інтервал:

$$P_{\text{п.о.}}(x_{\text{ви}}, \tau_{\text{оп}}) = \int_{x_{\text{ви, min}}}^{x_{\text{ви, max}}} \dots \int_{x_{\text{оп, min}}}^{x_{\text{оп, max}}} f(x_{\text{ви}}, \tau_{\text{оп}}) dx_{\text{ви}}, \quad (5.64)$$

де  $f(x_{оп}, \tau_{оп})$  – спільний диференціальний закон розподілу показників якості виконання задачі об'єктом:  $x_{оп, \min}, x_{оп, \max}$  – область допустимих значень показників якості.

Апаратура контролю об'єкта, що використовується на даний час, як правило, не виконує безпосереднього прогнозування при оцінці здатності об'єкта до виконання поставленого перед ним завдання. Оцінка здатності об'єкта в сучасній апаратурі виробляється за спрощеними критеріями "придатний - не придатний", "менше- норма - більше". Однак найпростіші критерії оцінки здатності надмірно посилюють вимоги, висунуті до надійності об'єкту, оскільки допустимі значення функціоналів в цьому випадку задаються з вимог дії наймаксимальніших збурень в найнесприятливіших комбінаціях. Крім цього апаратура контролю, виробляючи оцінку здатності за простішими критеріями, не дозволяє накопичувати інформацію для прогнозування відмов в системах об'єкта, а також не видає ймовірність виконання завдання даним об'єктом, яка характеризує ефективність його застосування і якою цікавляться для визначення ефективності дії групи об'єктів.

Для здійснення можливості оцінки здатності об'єкта з врахуванням прогнозного стану необхідно при розробці систем управління отримати всі необхідні вагомі коефіцієнти з врахуванням характеристик всіх можливих збурень, які використовуються в процесі прогнозування.

### 5.2.9 Визначення частоти контролю систем управління об'єктом та системи контролю

Частота контролю систем управління об'єктом та систем контролю визначається необхідністю застосування об'єкта а також готовністю об'єкта до застосування. Основне призначення контролю складається в тому, щоб або підготувати об'єкт до застосування, або утримувати його в безвідмовному стані тобто в такому стані, в якому він в будь-який момент в заданому інтервалі із заданою ймовірністю та з визначеною якістю виконає поставлене перед ним завдання.

Контроль об'єкта можна умовно розділити на регламентний контроль, контроль під час ремонту, безпосередній контроль об'єкта перед пуском, контроль в процесі роботи.

Розглянемо питання про оптимізацію процесу контролю з точки зору вибору оптимального значення частоти проведення регламентного контролю. При цьому будемо розуміти, що частота проведення регламентного контролю з системами управління об'єктом збігається з частотою проведення регламентних робіт із системами контролю. Методика вибору частоти проведення регламентного контролю може бути

розповсюджена на інші види контролю.

При виборі частоти проведення регламентного контролю необхідно враховувати такі основні фактори:

- час готовності об'єкта після виконання регламентного контролю;
- можливість відновлення готовності об'єкта при безпосередньому контролі перед пуском в процесі роботи.;
- інтенсивність відмов об'єкта в робочому та неробочому стані, а також під час його вмикання та вимикання;
- вартість системи контролю та технічного обслуговування;
- ресурс роботи систем управління об'єктом;
- наслідки пуску об'єкта в несправному стані;
- періоди роботи об'єкта (початковий, нормальний, старіння чи втомленості);
- умови роботи об'єкта, різні види зберігання і т.д.

Уявімо, що основний об'єм робіт, а також усі перевірки виконуються в процесі проведення регламентних робіт. При цьому природним є вимога щоб об'єкт мав задану готовність перед початком проведення регламентного контролю.

Якщо під готовністю розуміти ймовірність виконання об'єктом з визначеними якісними показниками поставленого перед ним завдання, то перед проведенням регламентного контролю :

$$P\left(X, X', t_{np}, \tau_{op}\right) \geq P_{\partial}\left(X, X', t, \tau_{op}\right), \quad (5.65)$$

де  $P\left(X, X', t_{np}, \tau_{op}\right)$  – ймовірність виконання об'єктом завдання в момент початку регламенту  $t_{np}$  до моменту закінчення виконання об'єктом завдання  $\tau_{op}$ ;  $P_{\partial}\left(X, X', t, \tau_{op}\right)$  – задана ймовірність виконання об'єктом завдання в момент  $\tau_{op}$ .

Підставивши в нерівність (5.65) вираз ймовірності виконання об'єктом завдання, отримаємо :

$$P(X, X', t_{np}, \tau_{op}) P_{ACKV}(t_{np}, \tau_{op}) \prod_{v=1}^m \prod_{\mu=1}^k P_{B.O}(x'_{v\mu}, t_{np}, \tau_{op}) * \\ * \prod_{v=1}^m \prod_{\mu=1}^k P_{П.О}(x'_{v\mu}, t_{np}, \tau_{op}) P_{\partial}(X, X', t, \tau_{op}) \quad (5.66)$$

Розв'язавши рівняння (5.66) відносно  $t_{np}$ , знайдемо час початку проведення регламентного контролю.

Розв'язок завдання значно спрощується, якщо завчасно відомі ймовірності

$$P_{ACKV}(t_{np}, \tau_{op}), P_{B.O}(x'_{v\mu}, t_{np}, \tau_{op}), P_{П.О}(x'_{v\mu}, t_{np}, \tau_{op}), \quad (5.67)$$

а також, якщо враховувати, що загальна ймовірність  $P$  буде менша найнадійнішої системи, що входить в об'єкт чи в систему контролю.

З врахуванням останнього величину  $t_{np}$  наближено можна знайти з рівняння

$$P(x'_{e\mu}, t_{np}, \tau_{op}) = P_{\partial}(x'_{v\mu}, t, \tau_{op}). \quad (5.68)$$

Частота перевірок визначається за відомим співвідношенням

$$F_p = \frac{1}{t_{np}}$$

Частоту регламентного контролю з врахуванням вартості його виконання можливо визначити з рівняння вартості регламентних робіт [21]

$$t_{np.c} = \left[ \frac{2T}{C_n} \left( \frac{C_{c.o}(F_p)}{m} + T_c \bar{P} C_p + T_H \bar{P} t_p \right) \right]^{1/2}, \quad (5.69)$$

де  $t_{нр.с}$  – середній час між відмовами в неробочому стані, діб;

$C_n$  – вартість втрат, які можуть мати місце в результаті невідомої відмови, крб.;

$C_{с.о}(F_p)$  – вартість регламентного обладнання, що змінюється в залежності від частоти регламентного контролю, крб.;

$m$  – число систем, з якими проводиться регламентний контроль;

$T_c$  – корисний строк служби підготовленої системи, діб;

$C_p$  – вартість заміни та ремонту системи, що вийшла з ладу, крб.;

$t_p$  – час, необхідний на ремонт та заміну, діб.

Частота регламентного контролю з врахуванням вартості його виконання визначається з виразу

$$F_{р.с.} = \frac{1}{t_{нр.с}}$$

Вибрана частота проведення регламентного контролю з врахуванням ймовірності виконання об'єктом завдання, а також з врахуванням вартості робіт може виявитися неоптимальною.

Для вибору оптимальної частоти виконання регламентного контролю необхідно використати узагальнений функціонально-статистичний критерій оцінки ефективності:

$$\Xi(t_n, \tau) = \frac{K_I(t_n, \tau)}{K_{I_0}(t_n, \tau)}. \quad (5.70)$$

Оптимальну частоту регламентного контролю за узагальненим критерієм можливо вибрати за таким алгоритмом:

1. Критерій є функцією лінійних форм.
2. Визначаються області зміни змінних:

$$P_o(t_n, \tau) \geq P_{озад.}(t, \tau),$$

$$P(t_n, \tau) \geq P_{зад.}(t, \tau),$$

$$\Delta C_\phi \leq C_{Зад.},$$

$$\Delta C_c \leq \Delta C_{с.о.},$$

$$\Delta C_G \leq \Delta C_{Годо.},$$



Надалі замість перших двох станів вводимо один працездатний, чи безвідмовний стан, а під контролем будемо розуміти всі види робіт, включаючи контроль, пошук несправностей та поновлень.

Якщо під вихідними параметрами розуміють реакції на вхідні сигнали систем або блоків, з точністю до несправностей яких відбувається контроль та пошук несправностей, то в загальному випадку число непрацездатних станів може виявитися більшим.

Практично виправдане припущення про вихід з ладу одночасно тільки одного блока чи системи. Число несправних станів в цьому випадку однакове.

Коли вірогідність відмови окремих блоків або систем мала, стани, що обумовлені цими відмовами можна не враховувати [41].

Параметри, що визначають стан об'єкта, обираються на основі аналізу структурної схеми з  $m$  блоків та систем, складеної для оцінки ефективності з урахуванням надійності та якісних показників.

Стан кожного  $i$ -го блока  $S_i$ , має ймовірність  $P_i$ , яку визначаємо з урахуванням інтенсивності відмов цього блока, (причому  $\sum P_i = 1$ ).

В якості міри апіорної невизначеності об'єкту приймається ентропія, яка змінюється в часі та може прогнозуватися до моменту часу  $t_{op}$ .

$$H(S_1, S_2, \dots, S_n, t, \tau) = -\sum P_i(t, \tau) \log_2 P_i(t, \tau). \quad (5.72)$$

Визначимо через  $F_i = F_i(y_1, \dots, y_m, t, \tau)$  множину результатів контролю, яка пов'язана з множиною  $y_i = y_i(x_1, \dots, x_m, t, \tau)$ . Складемо таблицю відповідності параметрів та ознак стану об'єкта для дискретних значень часу процесу контролю.

В будь-якому об'єкті, як правило, є заборонені стани, які призводять до аварій або катастроф. І не дивлячись на те, що ймовірність виявлення таких станів мала, необхідно вибрати в таблицю та виключити з розгляду ознаки цих станів, попередньо відбираючи їх для контролю.

Крім того, іноді виявляється можливим на основі простих висновків виключити з розглядання неможливі стани об'єкта та ознаки цих станів. Все це приводить до спрощення процедури подальшого відбору ознак станів.

Параметр і ознака	Номер стану							
	1	2	3	4	5	6	7	8
$x_1$	1	0	1	1	1	1	1	0
$x_2$	1	1	0	1	1	1	1	0
$x_3$	1	1	1	0	1	1	1	1
$x_4$	1	1	1	1	0	0	0	0
$x_5$	1	1	1	1	1	0	1	1
$x_6$	1	1	1	1	1	1	0	1
$y_1$	1	0	1	1	1	1	1	1
$y_2$	1	0	1	1	1	1	0	1
$y_3$	1	0	0	0	1	1	1	1
$y_4$	1	0	0	0	0	0	1	1
$y_5$	1	0	0	1	0	1	1	1
$y_6$	1	1	0	0	1	0	0	0
$y_7$	0	1	1	0	0	0	1	1
$y_8$	1	1	1	1	0	0	0	0
$y_9$	0	0	0	1	0	1	0	0

Число ознак стану, що залишилися, можна мінімізувати за ймовірністю [22] або по інформаційним методом [42; 43].

Мінімізуємо число ознак стану інформаційним методом [42]. Для того, щоб знайти, в якому  $i$ -му стані знаходиться об'єкт, необхідно визначити  $N$  ознак  $y_1, \dots, y_n$ . Кількість інформації, яка отримується при вимірюванні  $k$ -ої ознаки:

$$I(Y_k, t, \tau) = H(S, t, \tau) - H(S/Y_k, t, \tau), \quad (5.73)$$

де

$$H(S/Y_k, t, \tau) = \sum P(Y_k = i) H_{Y_k} = i(s, t, \tau) - \quad (5.74)$$

середня умовна ентропія стану об'єкта за умови визначення ознаки  $Y_k$ :

$$H_{Y_k} = j(S, t, \tau) = - \sum_{i=1}^N P(S_i/Y_{k=j}) \log_2 P(S_i/Y_{k=j}) - \quad (5.75)$$

середня умовна ентропія  $i$ -ого стану об'єкта за умови визначення.

Підставляючи в формулу (5.73) значення другого члена (5.74) з урахуванням (5.75), загалом отримуємо:



$$I(Y_k, t, \tau) = H(S, t, \tau) + \sum_{j=0}^1 P(Y_{k=j}) \sum_{i=1}^N (S_i / Y_{k=j}) \log_2 P(S_i / Y_{k=j}).$$

Вибір ознак стану має починатися з ознаки, яка несе максимальну кількість інформації:

$$I(Y_k, t, \tau) = H(S, t, \tau) - H(S/Y_k, t, \tau) = \max. \quad (5.76)$$

Наступна ознака обирається за цим же критерієм, але виходячи зі стану об'єкта, характеризується ентропією  $H(S/Y_k, t, \tau)$ .

Вибір ознак стану відбувається до тих пір, доки повна кількість інформації, яка містить багато відібраних ознак  $F_i(Y_1, Y_2, \dots, Y_n)$  відносно об'єкта, не буде рівною ентропії об'єкта  $I(F, S, t, \tau) = H(S, t, \tau)$ . При цьому стан об'єкта визначається одночасно за відібраними ознаками.

Під час мінімізації ознак стану об'єкта за загальним функціонально-статистичним критерієм спочатку обирають ознаку, яка має найбільшу ефективність, потім з тих, що залишилися, знову обирають ознаку з найбільшою ефективністю і т.д.

Звичайно вихідні параметри (ознаки), які характеризують стан об'єкта, обирають розробки систем управління об'єктом, тому на стадії проектування систем управління необхідно мінімізувати ознаки за загальним статистичним критерієм.

Крім цього критерію, мінімізацію ознак також можна виконувати на будь-якому з окремих критеріїв.

### 5.2.11 Вибір оптимального алгоритму контролю об'єкта

Для спрощення статистичної оптимізації процес контролю умовно розбіємо на процеси первинної та вторинної обробки інформації [44].

Первина обробка інформації призначена для заміщення невизначеності кожної конкретної системи об'єкта інформації за визначеним алгоритмом і для перетворення неперервної інформації в дискретну форму, а також для попереднього відбору та накопичення інформації.

Вторинна обробка інформації призначена для оптимізації регламентування всього процесу контролю та пуску об'єкта, реєстрації відхилень вихідних параметрів, вибору оптимального алгоритму та переробки керуючого сигналу підстройки несправного приладу; періодичної реєстрації вихідних параметрів та прогнозування відмов блоків та систем; оцінки готовності об'єкта в цілому та його окремих систем управління, зокрема за прийнятим критерієм оцінки підготовки; обирання характеру зіставлення з іншими машинами та приладами;

відпрацювання та прийняття рішень про включення резервних систем управління; вибір блокування, аварійного захисту обладнання, а також реєстрації порушень регламенту часу.

На основі такого розподілу оптимізацію процесу контролю можна також розділити на оптимізацію процесу первинної обробки інформації та оптимізації процесу вторинної обробки інформації. Розглянемо загальний випадок, який має відношення до оптимізації як одного, так і іншого процесу.

Критерії оптимізації алгоритму пошуку несправностей «час-вірогідність», «половинного розбивання», «комбінований метод», а також критерії оптимізації алгоритму контролю та пошуку несправностей, оснований тільки на використанні теорії інформації [45], є неповними та можуть бути використані тільки в окремих випадках.

Оптимальний алгоритм контролю об'єкта оберемо на основі узагальненого функціонально-статистичного критерію оптимізації

$$\mathfrak{E}(x_1, t, \tau) = \frac{K_I(x_1, t, \tau)}{K_{I0}(x_1, t, \tau)}, \quad (5.77)$$

де

$$K_I(x_1, t, \tau) = \frac{I_{\max}(x_1, t, \tau)}{C(x_1, t, \tau)} - \quad (5.78)$$

реальна ефективність алгоритму контролю та управління;

$$K_{I0}(x_1, t, \tau) = \frac{I_{\max \max}(x_1, t, \tau)}{C_{\min}(x_1, t, \tau)} - \quad (5.79)$$

ідеалізована ефективність алгоритму контролю та управління за максимальної його невизначеності та мінімальної вартості приладу контролю.

При оптимізації алгоритму контролю за критерієм будемо вважати оптимальним такий алгоритм, який має найбільшу ефективність. Розглянемо методику визначення чисельників та знаменників виразів (5.78), (5.79) при оптимізації алгоритму.

При складанні алгоритму оптимального процесу контролю необхідно розрізняти два специфічних види робіт, а відповідно, й два специфічних алгоритми їх виконання [45; 46]:

- алгоритм контролю, призначений для встановлення факту готовності об'єкта в цілому або однієї з його систем управління;
- алгоритм пошуку причин несправності чи невідповідності у випадку її встановлення.

Для обрання оптимального алгоритму контролю будемо вважати, що відома ймовірність безпомилкової роботи об'єкта  $P_o(x, t, \tau)$ , що складається з  $m$  систем, та його несправностей  $1 - P_o(x, t, \tau)$ ; крім того, відомо, що в об'єкті може відмовити тільки одна  $k$ -а з систем з ймовірністю  $\bar{P}_k(t, \tau)$ ,

$$\left( \sum_{i=1}^m \bar{P}_i(t, \tau) = 1 \right).$$

Ентропія об'єкта до початку контролю

$$H_o(x, t, \tau) = -\{P_o(x, t, \tau) \log_2 P_o(x, t, \tau) + [1 - P_o(x, t, \tau)] \log_2 [1 - P_o(x, t, \tau)] + [1 - P_o(x, t, \tau)] \sum_{i=1}^m \bar{P}_i(t, \tau) \log_2 \bar{P}_i(t, \tau)\}.$$

Будемо вважати, що при першому дослідженні контролюється  $n$  систем. Тоді ентропія об'єкта після першого дослідження

$$H(x, t, \tau, a) = H_o(x, t, \tau) - \left\{ 1 - \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \right\} \times \log_2 \left\{ 1 - \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \right\} + \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \log_2 [1 - P_o(x, t, \tau)] \sum_{i=1}^n \bar{P}_i(t, \tau).$$

Кількість інформації, яка отримується за дослідження:

$$I(x, t, \tau, a) = -\left\{ \left[ 1 - \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \right] \times \log_2 \left[ 1 - \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \right] + \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \log_2 \left[ \sum_{i=1}^n \bar{P}_i(t, \tau) [1 - P_o(x, t, \tau)] \right] \right\}.$$

Кількість інформації  $I(x, t, \tau, a)$  буде максимальною якщо

$$\sum_1^n \bar{P}_i = \frac{t}{2[1 - P(x, t, \tau)]}$$

Це видно на рис. 5.5.

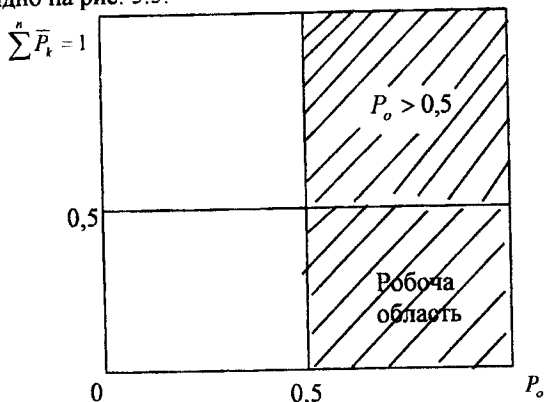


Рисунок 5.5 – Залежність кількості інформації від ймовірності правильної роботи об'єкта контролю

Максимальна кількість інформації, яка отримується за оптимальним алгоритмом контролю та знаходиться в числівнику виразу (5.78)

$$I_{\max}(x, t, \tau) = H_o(x, t, \tau) \quad (5.80)$$

Таким чином, алгоритм процесу контролю об'єкта буде оптимальним, якщо одночасно контролюється весь об'єкт. Якщо об'єкт знаходиться в стані максимальної невизначеності, тобто  $P(x, t, \tau) = 0,5$ , отримаємо кількість інформації  $I_{\max \max}(x, t, \tau)$ , що входить в (5.80).

Для обґрунтування оптимального алгоритму пошуку несправностей будемо вважати, що в об'єкті відмовила з вірогідністю  $\bar{P}_i(t, \tau)$  одна з  $m$  систем.

Ентропія об'єкта під час відмови  $m$  систем

$$H(x, t, \tau) = -\sum_1^m \bar{P}_i(t, \tau) \log_2 \bar{P}_i(t, \tau) \quad (5.81)$$

Умовна ентропія об'єкта після здійснення дослідження пошуку  $a_i$

$$H(x_1, x_2, a_1, t, \tau) = - \left\{ (\bar{P}_1 + \bar{P}_2) \left[ \frac{\bar{P}_1}{\bar{P}_1 + \bar{P}_2} \times \right. \right. \\ \times \log_2 \frac{\bar{P}_1}{\bar{P}_1 + \bar{P}_2} + \frac{\bar{P}_2}{\bar{P}_1 + \bar{P}_2} \log_2 \frac{\bar{P}_2}{\bar{P}_1 + \bar{P}_2} \left. \right] + \\ \left. + \left[ 1 - (\bar{P}_1 + \bar{P}_2) \right] \sum_3^m \frac{\bar{P}_i}{[1 - \bar{P}_1 + \bar{P}_2]} \log_2 \frac{\bar{P}_i}{[1 - \bar{P}_1 + \bar{P}_2]} \right\}.$$

Кількість інформації, яка отримується за дослідження:

$$I(x_1, x_2, a_1, t, \tau) = - \left\{ (\bar{P}_1 + \bar{P}_2) \log_2 (\bar{P}_1 + \bar{P}_2) + \right. \\ \left. + \left[ 1 - (\bar{P}_1 + \bar{P}_2) \right] \log_2 \left[ 1 - (\bar{P}_1 + \bar{P}_2) \right] \right\}.$$

Максимальну кількість інформації  $I(x_1, x_2, \dots, x_n, t, \tau, a)$  отримаємо якщо  $\sum_1^m P_i = 1/2$ , тобто якщо, наприклад, при рівномірних відмовах за перше дослідження розглянемо число систем  $n = m/2$ . Кількість інформації  $I_{\max \max}(x, t, \tau)$  отримується тоді, коли системи мають максимальну невизначеність.

Таким чином, максимальну кількість інформації під час кожного дослідження можна отримати шляхом обирання максимальної кількості одночасно досліджуваних систем.

Однак максимальна ефективність алгоритмів контролю та пошуку досягається не тільки їх максимальною інформаційною можливістю, але й вартістю апаратури, яка реалізує алгоритми:

$$C(t, \tau) = aT + bC_{ACKV} + dG + eV.$$

де  $T(t, \tau)$  – час, необхідний для проведення процесу контролю за оптимальним алгоритмом з точки зору отримання  $I_{\max}$ ;

$C(t, \tau), C(t, \tau)$  – відповідно вартість та вага апаратури контролю за оптимальним алгоритмом з точки зору отримання  $I_{\max}$ ;

$V(t, \tau)$  – обсяг апаратури контролю за оптимальним алгоритмом з точки зору отримання  $I_{\max}(t, \tau)$ ;

$a, b, d, e$  – вагові коефіцієнти, які враховують роль та вартість різних факторів.

Звісно, що за вартість  $C_{\min}(t, \tau)$  можна прийняти, як і раніше, початкову вартість найпростішої, наприклад, нерезервованої апаратури, яка реалізує алгоритм.

### Контрольні запитання і завдання

1. Що таке функціонально-статистичний критерій?
2. Алгоритм оптимізації процесу контролю в складних системах.
3. Наведіть основні умови знаходження максимуму функціонала.
4. Що таке ентропія процесу контролю та управління?
5. Які операції відносяться до основних операцій контролю і управління в складних системах?
6. Що таке ентропія об'єкта контролю і управління? Наведіть математичну і графічну модель.
7. Як визначити кількість інформації, одержану в процесі контролю і управління?
8. Як визначити кількість інформації, одержану в процесі контролю і управління для нормального спільного закону розподілу вхідних та вихідних параметрів?
9. Як визначається ентропія в системах регулювання та управління?
10. Як визначається ентропія степеня свободи в системах регулювання та управління?
11. Наведіть математичну модель критерію готовності складних об'єктів.
12. Як визначається частота контролю системи управління об'єктом.
13. Наведіть алгоритм вибору вихідних параметрів об'єкта для найбільшої ефективності контролю.
14. Наведіть математичну модель вибору оптимальної частоти виконання регламентного контролю.

## ЛИТЕРАТУРА

1. Банди Б. Методы оптимизации. Вводный курс. – М.: "Радио и связь", 1988.
2. Реклейтис Г., Рейвиндран А., Регсдел К. Оптимизация в технике. Кн. 1; – М.: Мир, 1986. – 350с.
3. Реклейтис Г., Рейвиндран А., Регсдел К. Оптимизация в технике. Кн. 2; – М.: Мир, 1986. – 405с.
4. Ю. И. Дегтярёв. Методы оптимизации – М.: "Советское радио", 1980. – 272с.
5. Щуп Т. Решение инженерных задач на ЭВМ. – М.: Мир, 1982. – 235с.
6. Васильев Ф. П. Лекции по методам решения экспериментальных задач. – М.: Изд-во Моск. Ун-та, 1974. – 374 с.
7. Полак Э. Численные методы оптимизации. – М.: Мир, 1974.
8. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах. – М.: Наука, 1975.
9. Моделирование и оптимизация на ЭВМ радиоэлектронных устройств/ Под. ред. З. М. Бененсона. – М.: Радио и связь, 1981. – 272 с.
10. Чуа. Л. А., Пен-Мин Лин. Машинный анализ электронных схем. – М.: Энергия, 1980. – 638 с.
11. Бахвалов Н. С. Численные методы. Т.1. Анализ, алгебра, обыкновенные дифференциальные уравнения. – М.: Наука, 1975.- 631 с.
12. Турчак Л. И. Основы численных методов. - М.: Наука, 1987.
13. Форсайт Дж., Малькольм М., Моулер Р. Машинные методы математических вычислений. - М.: Мир, 1980.
14. Краскевич В. Е.; Зеленский К. Х., Гречко В. И. Численные методы в инженерных исследованиях. - К.: Выща шк., 1986.- 263 с.
15. Демидович Б. П., Марон И. А. Основы вычислительной математики. – М.: Наука, 1970. – 664 с.
16. Иванов В. В. Методы вычислений на ЭВМ: Справочное пособие. – Киев: Наук. думка, 1986. – 584с.
17. Мэтьюз Джон, Финк Куртис. Численные методы. Использование MatLab. :Изд. дом Вильямс, 2001. – 720 с.
18. Кудрицкий Б. Поиск оптимальных решений средствами Excel 7.0 в примерах.- СПб.: ВHV – Санкт-Петербург, 1997. – 384 с.
19. Вильям Дж. Орвис. Excel для ученых, инженеров и студентов. – К.: "Юниор", 1999 – 528 с.
20. Понтрягин Л.С. и др. Математическая теория оптимальных процессов.-М.: Физматгиз, 1963.
21. Беллман Р. Динамическое программирование. - М.: Издательство иностранной литературы, 1960.
22. Гасс С. Линейное программирование. - М.: Физматгиз, 1961.

23. Шанталов А.С. Инженерные методы исследования регулируемых систем. Т.1, 2.- М.: ВИАУ, 1958.
24. Андреев Н.И. Нелинейное программирование в задачах исследования оптимальных систем автоматического управления М.: Издательство АН СССР, 1963.
25. Кузьмин И.В. и др. Методические указания по проблемному чтению курса ОТСС. Ч. 2. Вывод критериев оценки эффективности, качества и оптимизации СС.- Винница: ВПИ, 1983.
26. Бусленко Н.П., Шрейдер Ю.А. Метод статистических испытаний (Монте-Карло) и его реализация на ЦВМ.- М.: Физматгиз, 1961.
27. Бусленко Н.П., Голенко Д.П., Соболев И.М. и др. Метод статистических испытаний Монте-Карло.- М.: Физматгиз, 1962.
28. Бусленко Н.П. Решение задачи теории массового обслуживания методом моделирования на ЭЦВМ.- Проблемы передачи информации, 1961, №9.
29. Коган Б.М.; Тер – Микаэлян Т.М. Решение инженерных задач на ЦВМ.- М.: Госэнергоиздат, 1964.
30. Коган Б.М. Общие вопросы построения систем автоматизации с управляющими вычислительными машинами. // Электричество, 1962. - №11.
31. Фельдбаум А.А. Вычислительные устройства в автоматических системах. - М.: Физматгиз, 1959.
32. Гнедов Г.М., Кудрявцев В.Б. Основы автоматизации систем контроля и управления. - Л.: ЛВИКА, 1964.
33. Корн Ф.С. Тенденции развития проверочного оборудования сложных систем оружия. // В сб. статей под ред. В.А. Боднера.- М.: 1962.
34. Кузьмин И.В. Оценка эффективности автоматических систем контроля и управления. - Харьков: ХВКИУ, 1966.
35. Кузьмин И.В. Оценка эффективности АСКУ.- М.: Сов. Радио, 1971.
36. Дружинин Г.В. Статистическая теория износа и разрегулирования аппаратуры. Труды ВВИА им. проф. Н.Е. Жуковского, вып. 898, 1961.
37. Семенов В.М. К теории экстраполирования случайных процессов.- Труды ВВИА им. проф. Н.Е. Жуковского, т. 1, 1954.
38. Красовский В.С. Прогнозирование состояния объектов автоматического контроля. - В сб. статей / Под ред. И.М. Синдеева.- М.: МВВИА, вып. 1020, 1963.
39. Богданов Ю.В. Применение теории операторов к задачам исследования динамических систем.- Харьков: ХВКИУ, 1969.
40. Маст Л.Т., Паулсен Ф.Д. Выбор оптимального интервала между проверками систем, находящихся в состоянии статистической готовности к применению по тревоге.- В сб. статей / Под ред. В.А. Боднера.- М.: Изд-во иностр. лит-ры, 1962.



41. Зарубежная радиоэлектроника, 1960,-№ 4, С. 15.
42. Сиднеев И.М. О выборе параметров, определяющих состояние технического устройства при автоматическом контроле.- Труды ВВИА им. проф. Н.Е. Жуковского, вып. 1020, 1963.
43. Голдман С. Теория информации.- М.: Изд-во иностр. лит-ры, 1957.
44. Кузьмин И.В. Теоритические основы информационной техники.- Харьков: ХВКИУ, 1969.
45. Хоун А.Д., Солнц Е. Методы определения эффективной последовательности отыскания неисправностей. - В сб. статей / Под ред. В.А. Боднера.- М.: Изд-во иностр. лит-ры, 1962.
46. Сиднеев И.М. Некоторые общие методы автоматического контроля состояния динамических систем.- Труды ВВИА им. проф. Н.Е. Жуковского, 1962.
47. Месяцев П.П. Надежность производства ЭВМ.- М.: Машгаз, 1963.
48. Кузьмин И. В. и др. Методические указания по проблемному чтению курса ОТСС. Ч.2. Вывод критериев оценки эффективности, качества и оптимизации СС. – Винница: ВПИ, 1983.

## Додаток А

### Програма 1. Пошук мінімуму методом золотого перерізу

Програма призначена для пошуку чисельного наближення мінімуму функції  $f(x)$  на інтервалі  $[a,b]$  методом золотого перерізу в середовищі MatLab. Метод застосовується, тільки якщо функція  $f(x)$  унімодальна на інтервалі  $[a,b]$ .

```
function [S,E,G]=golden (f, a, b, delta, epsilon)
%Вхід - f – функція, яка вводиться як рядок 'f'
%      - a і b – крайні точки інтервалу
%      - delta – допустиме відхилення для абсцис
%      - epsilon – допустиме відхилення для ординат
%Вихід - S=(p, up) – містить абсцису p і ординату up мінімуму
%      - E=(dp, dy) – містить межі помилки для p і up
%      - G – матриця розміром n x 4: k-й рядок містить
%      [ak ck dk bk]; a, c, d і b на k-й ітерації
r1=(sqrt(5)-1)/2;
r2=r1^2;
h=b-a;
ya=feval (f, a);
yb=feval (f, b);
c=a+r2*h;
d=a+r1*h;
yc=feval (f, c);
yd=feval (f, d);
k=1;
A(k)=a; B(k)=b; C(k)=c; D(k)=d;
while (abs(yb-ya)>epsilon)|(h>delta)
    k=k+1;
    if (yc<yd)
        b=d;
        yb=yd;
        d=c;
        yd=yc;
        h=b-a;
        c=a+r2*h;
        yc=feval (f, c);
    else
        a=c;
        ya=yc;
        c=d;
        yc=yd;
        h=b-a;
```

```

d=a+r1*h;
yd=feval (f, d);
end
A(k)=a; B(k)=b; C(k)=c; D(k)=d;
end
dp=abs(b-a);
dy=abs(yb-ya);
p=a;
yp=ya;
if (yb<ya)
    p=b;
    yp=yb;
end
G=[A' C' D' B'];
S=[p yp];
E=[dp dy];

```

В програмах 2 і 4 необхідно, щоб функція  $f$  записувалась, як М-файл. Аргумент функції  $f$  повинен бути матрицею розміром  $1 \times n$ . Для ілюстрації розглянемо запис функції із прикладу 3 в вигляді М-файлу.

```

function z=f(V)
z=0; x=V(1); y=V(2);
z=x.^2-4x+y.^2-y-x.*y;

```

**Програма 2. Метод мінімізації Нелдера-Міда.** Програма призначена для наближеного знаходження локального мінімуму функції  $f(x_1, x_2, \dots, x_N)$ , де  $f$  – неперервна функція від  $N$  дійсних змінних, і задана  $N+1$  початкова точка  $V_k=(v_{k,1}, \dots, v_{k,N})$  для  $k=0, 1, \dots, N$ , в середовищі MatLab.

```

function [V0, y0, dV, dy]=nelder (F, V, min1, max1, epsilon, show)
%Вхід - F – функція, яка вводиться як рядок 'F'
%      - V – матриця розміром 3 x n, яка містить початковий симплекс
%      - min1 & max1 – мінімальна і максимальна кількість ітерацій
%      - epsilon – допустиме відхилення
%      - show == 1 – показує число ітерацій (P і Q)
%Вихід - V0 – вершина для мінімуму
%      - y0 – значення функції F(V0)
%      - dV – розмір кінцевого симплексу
%      - dy – межа помилки для мінімуму
%      - P – матриця, яка містить ітерації вершин
%      - Q – масив, який містить ітерації для F(P)
if nargin == 5;
    show=0;

```

```

end
[mm n]=size (V);
% Послідовність вершин
for j=1:n+1
    Z=V(j, 1:n);
    Y(j)=feval (F, Z);
end
[mm lo]= min (Y);
[mm hi]= max (Y);
li = hi;
ho = lo;
for j=1:n+1
    if (j ~= lo & j ~= hi & Y(j)<Y(li))
        li=j;
    end
    if (j ~= hi & j ~= lo & Y(j)>=Y(ho))
        ho=j;
    end
end
end
cnt=0;
% Початок алгоритму Нелдера-Міда
while (Y(hi)>Y(lo)+epsilon&cnt<max1) | cnt<min1
    S=zeros (1, 1:n);
    for j=1:n+1
        S=S+V(j, 1:n);
    end
    M=(S-V(hi, 1:n))/n;
    R=2*M-V(hi, 1:n);
    yR=feval (F, R);
    if (yR<Y(ho))
        if (Y(li)<yR)
            V(hi, 1:n)=R;
            Y (hi)=yR;
        else
            E=2*R-M;
            yE=feval (F, E);
            if (yE<Y(li))
                V(hi, 1:n)=E;
                Y(hi)=yE;
            else
                V(hi, 1:n)=R;
                Y(hi)=yR;
            end
        end
    end
end
end

```

```

else
    if (yR<Y(hi))
        V(hi, 1:n)=R;
        Y(hi)=yR;
    end
    C=(V(hi, 1:n)+M)/2;
    yC=feval (F, C);
    C2=(M+R)/2;
    yC2=feval (F, C2);
    if (yC2<yC)
        C=C2;
        yC=yC2;
    end
    if (yC<Y(hi))
        V(hi, 1:n)=C;
        Y(hi)=yC;
    else
        for j=1:n+1
            if (j ~ = lo)
                V(j, 1:n)=(V(j, 1:n)+V(lo, 1:n))/2;
                Z=V(j, 1:n);
                Y(j)=feval (F, Z);
            end
        end
    end
end
end
[mm lo]=min (Y);
[mm hi]=max (Y);
li=hi;
ho=lo;
for j=1:n+1
    if (j ~ = lo&j ~ = hi&Y(j)<=Y(li))
        li=j;
    end
    if (j ~ = hi & j ~ = lo&Y(j)>=Y(ho))
        ho=j;
    end
end
end
cnt=cnt+1;
P(cnt, :)=V(lo, :);
Q(cnt)=Y(lo);

```

```

end
% Кінець алгоритму Нелдера-Міда
% Визначення розміру симплексу

```

```

snorm=0;
for j=1:n+1
    s=norm(V(j)-V(lo));
    if (s>=snorm)
        snorm=s;
    end
end
Q=Q';
V0=V(lo, 1:n);
y0=Y(lo);
dV=snorm;
dy=abs(Y(hi)-Y(lo));
if (show == 1)
    disp (P);
    disp (Q);
end

```

**Програма 3. Пошук локального мінімуму, який використовує квадратичне інтерполювання.** Програма призначена для знаходження локального мінімуму функції  $f(x)$  на інтервалі  $[a; b]$  в середовищі MatLab. Обирається початкове наближення  $p_0$  і потім відбувається пошук на інтервалах  $[a; p_0]$  і  $[p_0; b]$ .

```

function [p, up, dp, dy, P]=guadmin (f, a, b, delta, epsilon)
%Вхід      - f – функція, яка вводиться як рядок 'f'
%          - a і b – крайні точки інтервалу
%          - delta – допустиме значення для абсцис
%          - epsilon – допустиме значення для ординат
%Вихід     - p – абсциса мінімуму
%          - up – ордината мінімуму
%          - dp – межа помилки для p
%          - dy – межа помилки для up
%          - P – вектор ітерацій
p0=a;
maxj=20;
maxk=30;
big=1e6;
err=1;
k=1;
P(k)=p0;
cond=0;
h=1;
if (abs(p0)>1e4), h=abs(p0)/1e4; end
while (k<maxk&err>epsilon&cond~=5)
    i1=(feval (f, p0+0.00001)-feval (f, p0-0.00001))/0.00002;

```

```

if (f1>0), h= - abs (h); end
p1=p0+h;
p2=p0+2*h;
pmin=p0;
y0=feval (f, p0);
y1=feval (f, p1);
y2=feval (f, p2);
ymin=y0;
cond=0;
j=0;
%Визначення такого h, що  $y_1 < y_0 & y_1 < y_2$ 
while (j<maxj&abs(h)>delta&cond == 0)
    if (y0<=y1)
        p2=p1;
        y2=y1;
        h=h/2;
        p1=p0+h;
        y1=feval (f, p1);
    else
        if (y2<y1);
            p1=p2;
            y1=y2;
            h=2*h;
            p2=p0+2*h;
            y2=feval (f, p2);
        else
            cond= - 1;
        end
    end
    j=j+1;
    if (abs(h)>big | abs (p0)>big), cond=5; end
    end
    if (cond == 5),
        pmin=p1;
        ymin=feval (f, p1);
    else
        %Квадратичне інтерполювання для знаходження УР
        d=4*y1-2*y0-2*y2;
        if (d<0),
            hmin=h*(4*y1-3*y0-y2)/d;
        else
            hmin=h/3;
            cond=4;
        end
    end
end

```

```

pmin=p0+hmin;
ymin=feval (f, pmin);
h=abs (h);
h0=abs (hmin);
h1=abs (hmin-h);
h2=abs (hmin-2*h);
%Визначення величини наступного h
if (h0<h), h=h0; end
if (h1<h), h=h1; end
if (h2<h), h=h2; end
if (h == 0), h=hmin; end
if (h<delta), cond=1; end
if (abs(h)>big | abs (pmin)>big), cond=5; end
%Критерій зупинки ітераційного процесу мінімізації
e0=abs (y0-ymin);
e1=abs (y1-ymin);
e2=abs (y2-ymin);
if (e0~=0 & e0<err), err=e0; end
if (e1~=0 & e1<err), err=e1; end
if (e2~=0 & e2<err), err=e2; end
if (e0~=0 & e1 == 0 & e2 == 0), error=0; end
if (err<epsilon), cond=2; end
p0=pmin;
k=k+1;
P(k)=p0;
end
if (cond == 2&h<delta), cond=3; end

```

```

end
p=p0;
dp=h;
yp=feval (f, p);
dy=err;

```

В програмі 4 необхідно, щоб функція  $f$  записувалась, як М-файл. Для ілюстрації розглянемо функцію  $f$ , для якої градієнт  $f$  дорівнює  $(2x - 4 - y, 2y - 1 - x)$ . Відповідний М-файл для цієї функції має такий вигляд:

```

function z=G(V)
z=zeros (1, 2);
x=V(1); y=V(2);
g=[2x-4-y 2*y-1-x];
z= - (1/norm(g))*g;

```



#### Програма 4. Метод найшвидшого спуску або градієнтний метод

Програма призначена для знаходження чисельного наближення локального мінімуму  $f(X)$ , де  $f$  - неперервна функція від  $N$  дійсних змінних і  $X=(x_1, \dots, x_N)$ , починаючи з однієї точки  $P_0$  в середовищі MatLab методом найшвидшого спуску.

```
function [P0, y0, err]=grads(F, G, P0, max1, delta, epsilon, show)
%Вхід
% - F - функція, яка вводиться як рядок 'F'
% - G= - (1/norm(gradF))*gradF; вибір напрямку
% вводитьься, як рядок 'G'
% - P0 - початкова точка
% - max1 - максимальне число ітерацій
% - delta - допустиме відхилення для hmin в єдиному
% параметрі мінімізації в обраному напрямку
% - epsilon - допустиме відхилення для помилки в y0
% - show; якщо show == 1, ітерації виводяться на дисплей
%Вихід
% - P0 - точка мінімуму
% - y0 - значення функції F(P0)
% - err - межа помилки для y0
% - P - вектор, який містить ітерації
if nargin == 5, show=0; end
[mm n]=size(P0);
maxj=10; big=1e8; h=1;
P=zeros(maxj, n+1);
Len=norm(P0);
y0=feval (F, P0);
if (len>e4), h=len/1e4; end
err=1; cnt=0; cond=0;
P(cnt+1, :)= [P0 y0];
While (cnt<max1 & cond~=5 & (h>delta | err>epsilon))
% Обчислення напрямку пошуку
S=feval (G, P0);
% Початок вибору параметра квадратичної мінімізації
P1=P0+h*S;
P2=P0+2*h*S;
y1=feval (F, P1);
y2=feval (F, P2);
cond=0; j=0;
while (j<maxj & cond == 0)
len=norm (P0);
if (y0<y1)
P2=P1;
y2=y1;
h=h/2;
P1=P0+h*S;
```

```

        y1=feval (F, P1);
else
    if (y2<y1)
        P1=P2;
        y1=y2;
        h=2*h;
        P2=P0+2*h*S;
        y2=feval (F, P2);
    else
        cond= - 1;
    end
end
j=j+1;
if (h<delta), cond=1; end
if (abs(h)>big | len>big), cond=5; end
end
if (cond == 5)
    Pmin=P1;
    ymin=y1;
else
    d=4*y1-2*y0-2*y2;
    if (d<0)
        hmin=h*(4*y1-3*y0-y2)/d;
    else
        cond=4;
        hmin=h/3;
    end
    % Побудова наступної точки
    Pmin=P0+hmin*S;
    ymin=feval (F, Pmin);
    % Визначення величини наступного h
    h0=abs (hmin);
    h1=abs (hmin-h);
    h2=abs (hmin-2*h);
    if (h0<h), h=h0; end
    if (h1<h), h=h1; end
    if (h2<h), h=h2; end
    if (h == 0), h=hmin; end
    if (h<delta), cond=1; end
    % Критерій зупину для мінімізації
    e0=abs (y0-ymin);
    e1=abs (y1-ymin);
    e2=abs (y2-ymin);
    if (e0~=0 & e0<err), err=e0; end

```

```
if (e1~=0 & e1<err), err=e1; end
if (e2~=0 & e2<err), err=e2; end
if (e0 == 0 & e1 == 0 & e2 == 0), err=0; end
if (err<epsilon), cond=2; end
if (cond == 2 & h<delta), cond=3; end
end
cnt=cnt+1;
P(cnt+1, :)= [Pmin ymin];
P0=Pmin;
y0=ymin;
end
if (show==1)
    disp(P);
end
```

ВСТУП .....	3
<b>1 СУТНІСТЬ МЕТОДІВ МАТЕМАТИЧНОЇ ОПТИМІЗАЦІЇ І ЇХ КЛАСИФІКАЦІЯ</b> .....	<b>4</b>
1.1 Основні поняття і положення оптимізації.....	4
1.2 Класифікація математичних методів оптимізації.....	6
1.3 Векторні математичні моделі оптимізації.....	9
1.4 Моделі оптимізації методом дослідження операцій.....	12
1.5 Моделі оптимізації задач обчислювальної математики.....	13
1.6 Економіко-математичні моделі.....	15
1.7 Задачі оптимізації в технічному проектуванні.....	17
1.8 Військові моделі оптимізації.....	19
1.9 Ціна математичної моделі.....	20
Контрольні запитання і завдання.....	21
<b>2 АНАЛІЗ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ</b> .....	<b>23</b>
2.1 Мінімум і максимум функцій за відсутності обмежень.....	23
2.2 Стаціонарні точки функцій цілі і необхідні й достатні умови екстремуму.....	26
2.3 Дослідження цільових функцій $W_k$ , $k=1, 2, \dots$ й множини.....	26
2.4 Методи пошуку мінімуму і максимуму функцій за наявності обмежень.....	29
2.4.1 Метод множників Лагранжа.....	29
2.4.2 Теорема Куна-Таккера. Подвійність задачі.....	31
2.4.3 Метод штрафних функцій.....	33
2.4.3.1 Загальні поняття і положення.....	33
2.4.3.2 Функції штрафу для обмежень у вигляді нерівностей.....	34
2.4.3.3 Штрафні функції для обмеження рівностей.....	38
2.5 Засоби пошуку екстремуму.....	40
2.5.1 Способи визначення градієнтів.....	41
2.5.2 Методи організації процесу пошуку.....	45
Контрольні запитання і завдання.....	47
<b>3 МЕТОДИ ОПТИМІЗАЦІЇ ОДНОВИМІРНИХ ЗАДАЧ</b> .....	<b>48</b>
3.1 Метод загального пошуку.....	51
3.2 Метод половинного ділення (розділення відрізка навпіл).....	51
3.3 Метод дихотомії.....	52
3.4 Метод “золотого перерізу”.....	54
3.5 Метод Фібоначчі.....	56
3.6 Порівняння методів одновимірного пошуку.....	61
Контрольні запитання і завдання.....	63
<b>4 МЕТОДИ ОПТИМІЗАЦІЇ БАГАТОВИМІРНИХ ЗАДАЧ</b> .....	<b>64</b>
4.1 Критерії оптимальності.....	66
4.2 Методи прямого пошуку.....	70
4.2.1 Метод пошуку по симплексу ( $S^2$ -метод).....	72

4.2.2	Метод покоординатного підйому.....	81
4.2.3	Метод виключення областей.....	81
4.2.4	Метод випадкового пошуку.....	85
4.3	Градієнтні методи .....	86
4.3.1	Найшвидший підйом з 3 використанням одномірного пошуку.....	88
4.3.2	Метод найшвидшого спуску.....	89
4.3.3	Метод Флетчера-Рівса.....	90
4.3.4	Метод Девідона-Флетчера-Пауела.....	93
4.3.5	Метод конфігурацій Хука-Дживса .....	97
4.3.6	Метод конфігурацій Розенброка .....	98
4.4	Аналіз сучасних пакетів програм для одновимірної та багатовимірної оптимізації .....	99
	Контрольні запитання і завдання .....	107
<b>5</b>	<b>ОПТИМІЗАЦІЯ СКЛАДНИХ СИСТЕМ УПРАВЛІННЯ .....</b>	<b>109</b>
5.1	Оптимізація процесу за узагальненим функціонально-статистичним критерієм.....	109
5.1.1	Алгоритм оптимізації процесу і АСКУ .....	110
5.1.2	Визначення оптимальних значень деяких параметрів АСКУ.....	112
5.2	Оптимізація процесу і системи контролю та управління.....	112
5.2.1	Основні поняття та визначення.....	113
5.2.2	Ентропія об'єкта в процесі контролю та управління.....	114
5.2.3	Кількість інформації, одержаної в процесі контролю та управління.....	117
5.2.4	Кількість інформації, отриманої в процесі контролю та управління за нормального спільного закону розподілу вхідних та вихідних параметрів.....	119
5.2.5	Ентропія та інформація в системах регулювання та управління....	125
5.2.6	Ентропія ступеня свободи процесу контролю та управління.....	128
5.2.7	Динамічні статистичні характеристики процесу контролю і управління.....	129
5.2.8	Критерій оцінки готовності складних об'єктів .....	131
5.2.9	Визначення частоти контролю систем управління об'єктом та системи контролю .....	138
5.2.10	Вибір вихідних параметрів об'єкту, які найповніше характеризують його стан та є ефективнішими під час контролю.....	142
5.2.11	Вибір оптимального алгоритму контролю об'єкта.....	145
	Контрольні запитання і завдання .....	150
	Література .....	151
	Додаток А Приклади програм мінімізації цільової функції.....	154

*Навчальне видання*

Іван Васильович Кузьмін  
Микола Максимович Биков  
Світлана Михайлівна Москвіна  
Андрій Іванович Кузьмін

## **Методи оптимізації складних систем**

**Навчальний посібник**

Оригінал-макет підготовлено авторами

Редактор С.А. Малішевська

Навчально-методичний відділ ВДТУ  
Свідоцтво Держкомінформу України  
серія ДК № 746 від 25.12.2001  
21021, м. Вінниця, Хмельницьке шосе, 95, ВДТУ

Підписано до друку ~~11/02~~ Гарнітура Times New Roman  
Формат 29,7x42  $\frac{1}{4}$  Папір офсетний  
Друк різнографічний Ум. друк. арк. 6,96  
Тираж 85 прим.  
Зам. № 2003-118

Віддруковано в комп'ютерному інформаційно-видавничому центрі  
Вінницького державного технічного університету  
Свідоцтво Держкомінформу України  
Серія ДК № 746 від 25.12.2001  
21021 м. Вінниця, Хмельницьке шосе, 95, ВДТУ, ГНК, 9-й поверх,  
тел. (0432) 44-01-59