МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

О.М. Кочевський, О.Г. Гусак

КОНСПЕКТ ЛЕКЦІЙ
З ДИСЦИПЛІНИ
"МОДЕЛІ І МЕТОДИ ОПТИМІЗАЦІЇ"

для студентів спеціальності 7.000008 "Енергетичний менеджмент" денної та заочної форм навчання

СУМИ ВИД-ВО СУМДУ 2005

ВСТУП

Процес оптимізації є в основі всієї інженерної діяльності, оскільки функції спеціаліста полягають у тому, щоб, з одного боку, проектувати нові, більш ефективні та менш дорогі технічні системи, а з іншого боку, розробляти методи підвищення якості функціонування існуючих систем.

У практичній діяльності часто з багатьох можливих розв'язань завдань необхідно вибрати оптимальне. Наприклад, із декількох варіантів перевезення сировини споживачам необхідно вибрати найбільш дешевий, але такий, що враховує обмеження на допустимі терміни поставок; із можливих планів розкрою матеріалу вибрати такий, який би дозволив виконати план при якнайменшій кількості відходів і т.п.

Зокрема, вибір оптимальної схеми теплових та енергетичних мереж, а також найбільш придатних установок для використання у цих мережах, та інші проблеми такого характеру, є за своєю суттю головною задачею спеціаліста з енергоменеджменту. У своїй повсякденній трудовій діяльності він проводить техніко-економічний аналіз різних варіантів схем мереж, що пов'язано з використанням відповідних критеріїв для комплексної порівняльної оцінки якості схем, що розглядаються, та необхідного математичного апарата розв'язання оптимізаційних задач.

У багатьох випадках задача пошуку оптимального рішення може бути формалізована і розв'язана точно або приблизно відомими методами.

Необхідність чіткого розуміння шляхів підвищення ефективності роботи теплових та енергетичних мереж наполегливо потребує вивчення курсу "Моделі і методи оптимізації", змістовна основа якого здебільшого базується на спеціальному розділі математиці – теорії оптимізації, а також на відповідних чисельних методах.

Курс "Моделі і методи оптимізації" ставить за мету:

- ознайомити студентів з теоретичними аспектами чисельних методів розв'язання задач оптимізації і їхньої комп'ютерної реалізації;
- показати можливості цих методів на прикладах розв'язання прикладних задач;
- навчити застосовувати відповідний математичний апарат під час розв'язання проблем, пов'язаних з оптимізацією.

У результаті вивчення дисципліни студент повинен знати основні, найбільш ефективні чисельні методи розв'язання задач оптимізації та їх особливості.

Студент повинен уміти:

- правильно формулювати і класифікувати завдання оптимізації;
- вибирати або розробляти методи для їх розв'язання та реалізовувати ці методи у вигляді комп'ютерних програм;
- застосовувати сучасні моделі та методи оптимізації для розв'язання практичних задач вибору та вдосконалення енергетичного обладнання.

Подяка При створенні даного конспекту лекцій автори значною мірою спиралися на матеріали дистанційного курсу “Методи оптимізації” Любчака В.О. та Острівної О.Г., з яким можна ознайомитися за адресою http://dl.sumdu.edu.ua/mo/ukr/ukr.html.

РОЗДІЛ 1 ОСНОВНІ ПОНЯТТЯ ТА ВИЗНАЧЕННЯ

1.1 Загальні поняття

1.1.1 Поняття про оптимізацію

Оптимізація – цілеспрямована діяльність, що полягає в отриманні найкращих результатів за відповідних умов.

Пошуки оптимальних розв'язків привели до створення спеціальних математичних методів і вже в XVIII столітті були закладені математичні основи оптимізації (варіаційне числення, чисельні методи й інші). Однак до другої половини XX століття методи оптимізації в багатьох галузях науки і техніки застосовувалися дуже рідко, оскільки практичне використання математичних методів оптимізації вимагало величезної обчислювальної роботи, що без ЕОМ реалізувати було досить важко, а у ряді випадків – неможливо. Особливо великі труднощі виникали під час розв'язання задач оптимізації процесів у хімічній технології через велике число параметрів та їх складний взаємозв'язок між собою. За наявності ЕОМ розв'язання задач помітно спрощуються.

Як відомо, інженерні розрахунки, що проводяться на етапі конструювання, завжди багатоваріантні, тобто задача створення будь-якої технічної системи завжди має безліч розв'язань. У міру ускладнення конструкції проектувальнику все важче вирішувати, якою саме слід створити технічну систему, щоб вона виявилася найбільш досконалою. Звичайно число можливих варіантів проекту під час введення у розгляд кожного нового параметра зростає у експоненціальній залежності.

Наведемо кілька прикладів поставлення конструкторських задач [1].

1 Так, перед конструктором гідравлічної турбіни постає задача вибору оптимальних співвідношень між коловою швидкістю, питомою вагою та розмірами ротора. У певному діапазоні, що більше діаметр турбіни, то більший коефіцієнт корисної дії може бути досягнуто, але відповідно стрімко збільшується вартість турбіни.

2 Перед авіаконструкором постійно постають задачі створення максимально легких та компактних конструкцій при забезпеченні необхідної міцності та надійності.

3 Вибір конструктивних параметрів усіх типів гідравлічних об'ємних насосів визначається прагненням отримати максимальну продуктивність при мінімальних розмірах, вазі, вартості. При проектуванні шестеренних, шиберних, гвинтових та поршневих насосів необхідно враховувати цілий ряд аналогічних обмежувальних умов: в усіх типах насосів колова швидкість ротора обмежується вимогою відсутності кавітації; для кожного розмірного типу насоса є оптимальна кількість рухомих елементів, їх оптимальні розміри і т.п.

4 При конструюванні гідравлічних та пневматичних циліндрів-підйомників ставиться задача забезпечення максимальної вантажопідйомності

при мінімальних габаритах, вазі та вартості. При цьому необхідно отримати оптимальні значення діаметра циліндра та довжини ходу при заданій потужності насоса або компресора та забезпеченні надійності та стійкості. В телескопічних циліндрах до цього переліку шуканих оптимальних величин додається ще і оптимальна кількість ланок.

Поставлення задачі оптимізації під час розроблення будь-якої інженерної конструкції припускає існування конкуруючих показників якості виробу:

- економічність пристрою – потужність пристрою (при фіксованих габаритних розмірах);
- економічність пристрою – вартість пристрою (при фіксованих номінальних параметрах роботи).

Вибір компромісного варіанта для забезпечення зазначених показників якості і являє собою процедуру розв'язання оптимізаційнної задачі.

Поставлення задачі оптимізації вимагає вибрати лише один показник якості досліджуваної конструкції, що якнайкраще виражає міру її досконалості. Цей показник якості називають критерієм оптимальності.

Критерій оптимальності – прийнятий показник якості (ефективності) досліджуваної конструкції, або показник, що дозволяє кількісно оцінити ефективність проектного розв'язання.

Одночасно об'єкту оптимізації не повинно приписуватися два та більше критеріїв оптимальності, тому що практично завжди экстремум одного критерію не відповідає екстремуму іншого.

Типовий приклад неправильного поставлення задачі оптимізації:

"Одержати максимальну продуктивність при мінімальній собівартості". Помилка полягає в тому, що ставиться завдання пошуку оптимуму двох показників, що суперечать один одному по своїй суті.

Правильне поставлення задачі могло бути таким:

а) одержати максимальну продуктивність при заданій собівартості;

б) одержати мінімальну собівартість при заданій продуктивності;

У першому випадку критерій оптимізації – продуктивність, а в другому – собівартість.

Поставлення задачі оптимізації вимагає також наявність можливості варіювати значеннями деяких параметрів, від яких залежить значення критерію оптимальності об'єкта. Іншими словами, об'єкт повинен мати ступені свободи – *параметри оптимізації*. В задачах оптимізації інженерних пристроїв та систем як параметри оптимізації можуть виступати насамперед геометричні розміри та кількість елементів конструкції. Звичайно в поставлення задачі оптимізації вводять лише ті параметри, що справляють найбільш значущий вплив на критерій оптимальності. Відзначимо, що діапазон варіювання цих параметрів звичайно обмежений з міркувань міцності або економічності.

Вплив, який здійнюють ці параметри на значення критерію оптимальності, описується цільовою функцією. *Цільова функція* – функція, що пов'язує критерій оптимальності з параметрами оптимізації. Як цільову функцію, так і параметри оптимізації слід вибирати таким чином, щоб їх можна було оцінити кількісно, оскільки тільки в цьому випадку можна порівнювати ефекти від вибору тих або інших значень параметрів оптимізації.

1.1.2 Математичне поставлення задач оптимізації. Види обмежень

Незважаючи на те, що прикладні задачі належать до цілком різних галузей, вони мають загальну форму. Усі ці задачі можна класифікувати як задачі мінімізації дійснозначущої функції f ($\boldsymbol{x}$) n-вимірного векторного аргумента $\boldsymbol{x} = (x_1, x_2, ..., x_n)$, компоненти якого задовольняють системі рівнянь h_k ($\boldsymbol{x}$) = 0, набору нерівностей g_i ($\boldsymbol{x}$) $\geq$ 0, а також обмежені зверху і знизу, тобто $x_i (u) \geq x_i \geq x_i (l)$.

Надалі функцію f ($\boldsymbol{x}$) будемо називати *цільовою функцією*, рівняння $h_k (\boldsymbol{x}) = 0$ – обмеженнями-рівняннями, а нерівності $g_i (\boldsymbol{x}) \geq 0$ – обмеженнями-нерівностями. При цьому передбачається, що усі функції, що фігурують у задачі, є дійснозначними, а число обмежень є скінченна величина.

Завдання загального вигляду “мінімізувати функцію f ($\boldsymbol{x}$) при обмеженнях h_k ($\boldsymbol{x}$) = 0, k = 1, ..., K, g_i ($\boldsymbol{x}$) $\geq$ 0, j = 1, ..., J, $x_i (u) \geq x_i \geq x_i (l)$, i = 1, ..., N” називається задачею оптимізації з обмеженнями або задачею умовної оптимізації.

Задачі, у яких немає обмежень, тобто $J = K = 0$; $x_i (u) = \infty$, $x_i (l) = -\infty$, i = 1, ..., N, називаються оптимізаційними задачами без обмежень або задачами безумовної оптимізації.

Задачі оптимізації класифікують відповідно до вигляду функцій f ($\boldsymbol{x}$), h_k ($\boldsymbol{x}$), g_i ($\boldsymbol{x}$) і розмірності вектора $\boldsymbol{x}$. Задачі без обмежень, у котрих $\boldsymbol{x}$ являє собою одномірний вектор, називаються задачами з однією змінною і складають найпростіший, але водночас дуже важливий підклас оптимізаційнних задач. Завдання умовної оптимізації, у яких функції h_k ($\boldsymbol{x}$) і g_i ($\boldsymbol{x}$) є лінійними, називаються задачами із лінійними обмеженнями.

1.1.3 Про критерій оптимальності та його визначення

Що звичайно оптимізується величина, пов’язана з економічністю роботи об’єкта, що розглядається (апарат, цех, завод). Варіант роботи, що оптимізує об’єкт, повинен оцінюватися деякою кількісною мірою – критерієм оптимальності.

Критерієм оптимальності називається кількісна оцінка якості об’єкта, що оптимізується.

На підставі обраного критерію оптимальності складається цільова функція, що являє собою залежність критерію оптимальності від параметрів, що впливають на її значення. Вид критерію оптимальності або цільової функції визначається конкретною задачею оптимізації.

Таким чином, задача оптимізації зводиться до знаходження екстремуму цільової функції.

Найбільш загальним поставленням оптимальної задачі є вираження критерію оптимальності у вигляді економічної оцінки (продуктивність, собівартість продукції, прибуток, рентабельність).Однак у окремих задачах оптимізації, коли об'єкт є частиною технологічного процесу, не завжди вдасться або не завжди доцільно виділяти прямий економічний показник, який би цілком характеризував ефективність роботи розглянутого об'єкта. У таких випадках критерієм оптимальності може бути технологічна характеристика, що побічно оцінює економічність роботи агрегата (час контакту, вихід продукту, ступінь перетворення, температура). Наприклад, встановлюється оптимальний температурний профіль, тривалість циклу "реакція – регенерація". Але в будь-якому випадку будь-який критерій оптимальності має економічну природу.

Розглянемо більш докладно вимоги, що повинні висуватися до критерію оптимальності.

1 Критерій оптимальності повинен виражатися кількісно.

2 Критерій оптимальності повинен бути єдиним.

3 Критерій оптимальності повинен відбивати найбільш істотні боки процесу.

4 Бажано, щоб критерій оптимальності мав ясний фізичний зміст і легко розраховувався.

При поставленні конкретних задач оптимізації цільова функція, що пов'язує критерій оптимальності з параметрами оптимізації, має бути записана у вигляді аналітичного виразу – математичної моделі. Якщо об'єктом оптимізації є добре відомий об'єкт або процес, цей аналітичний вираз отримати нескладно.

На жаль, у багатьох випадках, зокрема, в задачах конструювання, вид цільової функції є невідомим або відомим лише наближено. Тоді отримання цільової функції передбачає проведення тривалого експериментального дослідження (або розрахункового, шляхом застосування сучасних чисельних методів моделювання фізичних процесів). У цьому випадку для планування дослідження використовується теорія планування експерименту, що також базується на методах оптимізації, які ми будемо розглядати далі. Метою застосування теорії планування експерименту є відшукання таких значень параметрів оптимізації, при яких критерій оптимальності сягає максимуму, з забезпеченням найбільшої точності та найменшої можливої кількості дослідів.

Отже, для розв'язання задачі оптимізації необхідно:

а) вибрати критерій оптимальності та параметри оптимізації;

б) встановити можливі обмеження, що повинні накладатися на параметри оптимізації;

в) скласти, якщо можливо, математичну модель об'єкта оптимізації – цільову функцію; якщо це неможливо, значення цільової функції при кожному наборі значень параметрів оптимізації доведеться отримувати шляхом трудомісткого дослідження;

г) вибрати метод оптимізації, що дозволить знайти екстремальні значення шуканих величин.

1.2 функції однієї змінної. необхідні і достатні умови

Функція $f(x)$ має локальний мінімум у точці x_0, якщо існує деяка додатна величина δ, така, що якщо $|x - x_0| < \delta$, то $f(x) \geq f(x_0)$, тобто якщо існує окіл точки x_0, такий, що для всіх значень x у цьому околі $f(x)$ більше $f(x_0)$. Функція $f(x)$ має глобальний мінімум у точці x^*, якщо для всіх x справедлива нерівність $f(x) \geq f(x^*)$ [2].

На рис. 1.1 подане графічне зображення функції $f(x)$, що має локальний мінімум у точці x_0 і глобальний мінімум у точці x^*.

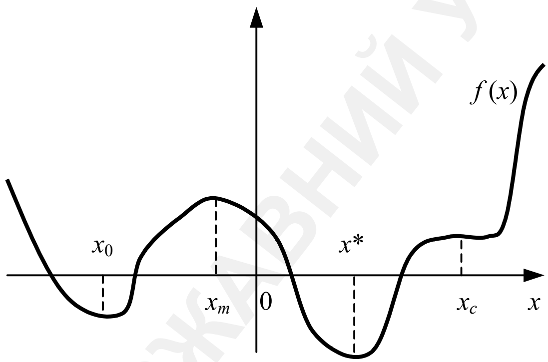


Рисунок 1.1 – Графічне зображення функції

Класичний підхід до задачі знаходження значень x_0 і x^* полягає у пошуку рівнянь, які вони повинні задовольняти. Подана на рис. 1.1 функція та її похідні неперервні, і бачимо, що в точках x_0 і x^* похідна $f'(x)$ (градієнт функції) дорівнює нулю. Отже, x_0 та x^* будуть розв'язками рівняння

$$f'(x) = 0. \tag{1.1}$$

Точка x_m, в якій досягається локальний мінімум, і точка x_c, в якій є точка горизонтального перегину функції, також задовольняє це рівняння. Отже, рівняння (1.1) є тільки *необхідною* умовою мінімуму, але не є *достатньою* умовою мінімуму.

Відмітимо, однак, що в точках x_0 і x^* похідна $f'(x)$ змінює знак з від'ємного на додатний. У точці x_m знак змінюється з додатного на від'ємний, у той час як у точці x_c він не змінюється. Отже, похідна в мінімумі є зростаючою функцією, а оскільки ступінь зростання $f'(x)$ вимірюється другою похідною, можна очікувати, що $f''(x_0) > 0$, $f''(x^*) > 0$, тоді як $f''(x_m) < 0$.

Якщо, однак, друга похідна дорівнює нулю, ситуація залишається невизначеною.

Отримані вище результати можуть знайти надійне обґрунтування, якщо розглянути розкладання функції $f(x)$ у ряд Тейлора в околі точки x_0 (чи x^*, чи x_m), що звичайно, вимагає неперервності функції $f(x)$ та її похідних:

$$f(x_0 + h) - f(x_0) = hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \ldots \qquad (1.2)$$

Якщо в точці x_0 досягається мінімум, то ліва частина (1.2) буде невід'ємною для будь-якого досить малого h ($|h| < \delta$). Отже, перша похідна $f'(x_0)$ повинна дорівнювати нулю, і це є достатня умова (див. рівняння 1.1). Якби вона була додатною, то достатньо мале від'ємне значення h робило б праву частину (1.2) від'ємною, а якби вона була від'ємною, то достатньо мале додатне значення h робило б праву частину від'ємною.

Оскільки в наступному члені (1.2) завжди $h^2 > 0$, то, якщо

$$f''(x_0) > 0, \qquad (1.3)$$

у точці x_0 досягається мінімум. Якщо $f'(x_m) = 0$ і $f''(x_m) < 0$, то з аналогічних міркувань у точці x_m досягається максимум. Для визначення розходження між локальним і глобальним мінімумами необхідно порівняти значення функцій $f(x_0)$ і $f(x^*)$.

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

$f'(x) = 3x^2 - 4x + 1 = 0$,

тоді $(3x - 1)(x - 1) = 0$, 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.

1.3 функції *N* змінних. необхідні і достатні умови

Розглянемо функцію n дійсних змінних

$$f(x_1, x_2, x_3, \ldots, x_n) = f(\boldsymbol{x}).$$

Точка в n-мірному евклідовому просторі з координатами $x_1, x_2, x_3, \ldots, x_n$ позначається вектором-стовпцем $\boldsymbol{x}$. Градієнт функції, тобто вектор з компонентами $\partial f / \partial x_1$, $\partial f / \partial x_2$, …, $\partial f / \partial x_n$, позначається grad $f(\boldsymbol{x})$ або, іноді, $g(x)$. Матриця Гессе (гессіан) функції $f(\boldsymbol{x})$ позначається як $\boldsymbol{G}(\boldsymbol{x})$ і являє собою симетричну матрицю n х n елементів виду

$$G_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} .$$

Тобто, матриця $\boldsymbol{G}(\boldsymbol{x})$ набуде вигляду:

$$G(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Функція $f(\boldsymbol{x})$ має локальний мінімум у точці $\boldsymbol{x}_0$, якщо існує окіл точки $\boldsymbol{x}_0$ такий, що $f(\boldsymbol{x}) > f(\boldsymbol{x}_0)$ у всіх точках цього околу, тобто існує додатна величина δ, така, що для $|\boldsymbol{x} - \boldsymbol{x}_0| < \delta$ справедлива нерівність $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$.

У випадку глобального мінімуму в точці $\boldsymbol{x}^*$ для всіх $\boldsymbol{x}$ справедливою є нерівність $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$.

При таких визначеннях і очевидних припущеннях щодо диференційованості можна узагальнити рівняння (1.2) і отримати:

$$\begin{aligned} f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) &= \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x_1, ..., x_n) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, ..., x_n) + ... \\ &= \boldsymbol{h}^{\mathrm{T}} \operatorname{grad} f(\boldsymbol{x}) + 0.5\, \boldsymbol{h}^{\mathrm{T}} \boldsymbol{G}(\boldsymbol{x}_0)\, \boldsymbol{h} + \ldots . \qquad (1.4) \end{aligned}$$

Тоді, якщо $\boldsymbol{x}_0$ є точкою мінімуму функції $f(\boldsymbol{x})$, то кожна перша часткова похідна $\partial f / \partial x_i$ ($i = 1, …, n$) повинна обертатися в нуль у точці $\boldsymbol{x}_0$. Якщо це не так, то відповідним вибором h_i можна домогтися того, що різниця $f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0)$ буде від'ємною.

Отже, необхідною умовою мінімуму в точці $\boldsymbol{x}_0$ є рівняння

$$\operatorname{grad} f(\boldsymbol{x}_0) = 0, \qquad (1.5)$$

тобто,

$$\frac{\partial f(x_0)}{\partial x_i} = 0 \qquad (i = 1, \ldots, n). \tag{1.6}$$

Тоді знак різниці $f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0)$ визначається членом

$$0.5\, \boldsymbol{h}^{\mathrm{T}}\, \boldsymbol{G}(\boldsymbol{x}_0)\, \boldsymbol{h}. \tag{1.7}$$

Якщо матриця $\boldsymbol{G}(\boldsymbol{x}_0)$ додатно визначена, то цей член є додатним для всіх $\boldsymbol{h}$. Отже, необхідними і достатніми умовами мінімуму є:

$$\operatorname{grad} f(\boldsymbol{x}_0) = 0,\ \boldsymbol{G}(\boldsymbol{x}_0) \text{ додатно визначена.} \tag{1.8}$$

Матриця називається додатно визначеною ($\boldsymbol{G}(\boldsymbol{x}_0) > 0$), якщо для будь-якого ненульового $\boldsymbol{h}$ виконується нерівність $\boldsymbol{h}^{\mathrm{T}}\, \boldsymbol{G}(\boldsymbol{x}_0)\, \boldsymbol{h} > 0$.

Необхідними і достатніми умовами максимуму є:

$$\operatorname{grad} f(\boldsymbol{x}_m) = 0,\ \boldsymbol{G}(\boldsymbol{x}_m) \text{ від'ємно визначена [2].} \tag{1.9}$$

Приклад Дослідити екстремальну точку (точки) функції $f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2 - 4x_1 - 8x_2 - 12x_3 + 100$:

$$\operatorname{grad} f(\boldsymbol{x}) = \begin{pmatrix} 2x_1 - 4 \\ 2x_2 - 8 \\ 2x_3 - 12 \end{pmatrix} = 0 \quad \text{при } x_1 = 2,\ x_2 = 4,\ x_3 = 6.$$

Маємо матрицю $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Визначимо, чи є вона додатно визначеною:

$$\boldsymbol{h}^{\mathrm{T}}\, \boldsymbol{G}(\boldsymbol{x}_0)\, \boldsymbol{h} = [h_1\ h_2\ h_3] \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 2\,h_1^2 + 2\,h_2^2 + 2\,h_3^2 > 0.$$

Тобто матриця $\boldsymbol{G}(\boldsymbol{x}_0)$ дійсно є додатно визначеною. Всі власні значення додатні і дорівнюють 2.

Отже, в точці (2; 4; 6) функція $f(\boldsymbol{x})$ сягає мінімуму.

РОЗДІЛ 2 МЕТОДИ ПОШУКУ ДЛЯ ФУНКЦІЙ ОДНІЄЇ ЗМІННОЇ

Незважаючи на те, що безумовна оптимізація функції однієї змінної – найбільш простий тип оптимізаційних задач, вона займає центральне місце в теорії оптимізації як з теоретичної, так і з практичної точки зору. Це пов’язано з тим, що задачі однопараметричної оптимізації досить поширені в інженерній практиці і, крім того, застосовуються при реалізації більш складних ітеративних процедур багатопараметричної оптимізації.

2.1 Методи виключення інтервалів

2.1.1 Загальний підхід

Методи пошуку, що дозволяють визначити оптимум функції однієї змінної шляхом зменшення інтервалу пошуку, називаються методами виключення інтервалів.

Усі методи одновимірної оптимізації базуються на припущенні, що досліджувана цільова функція в припустимій області принаймні має властивість *унімодальності*, тобто має лише один локальний мінімум, тому що для унімодальної функції $f(x)$ порівняння значень у двох різних точках інтервалу пошуку дозволяє визначити, у якому із заданих двома зазначеними точками підінтервалів точки оптимуму відсутні.

Правило виключення інтервалів. Припустимо, точки a та b окреслюють інтервал, який містить справжню точку мінімума, і всередині цього інтервалу $[a, b]$ функція є $f(x)$ *унімодальною*, тобто має один мінімум у точці x^*. Розглянемо точки x_1 і x_2, що розміщені в інтервалі таким чином, що $a < x_1 < x_2 < b$. Порівнюючи значення функції в точках x_1 та x_2, можна зробити такі висновки.

- Якщо $f(x_1) > f(x_2)$, то точка мінімуму $f(x)$ не є в інтервалі (a, x_1), тобто x^* належить (x_1, b).
- Якщо $f(x_1) < f(x_2)$, то точка мінімуму $f(x)$ не є в інтервалі (x_2, b), тобто x^* належить (a, x_2).

Це правило дозволяє реалізувати процедуру пошуку шляхом послідовного виключення частин вихідного обмеженого інтервалу. Пошук завершується тоді, коли підінтервал, що залишився, зменшується до досить малих розмірів.

Головна перевага таких пошукових методів – вони ґрунтуються на обчисленні тільки значень функції і, отже, не вимагають виконання умови диференційованості і запису в аналітичному вигляді. Остання властивість особливо цінна при імітаційному моделюванні.

Процес застосування методів пошуку на основі виключення інтервалів вміщує два етапи:

1) *етап встановлення границь інтервалу*, на якому реалізується процедура пошуку границь достатньо широкого інтервалу, що містить точку оптимуму;

2) *етап зменшення інтервалу*, на якому реалізується кінцева послідовність перетворень початкового інтервалу з тим, щоб зменшити його довжину до заздалегідь встановленої величини.

1) *Етап встановлення границь інтервалу*

На цьому етапі спочатку вибирається вихідна точка, а потім на основі правила виключення будується відносно широкий інтервал, що містить точку оптимуму. Звичайно використовується евристичний метод, наприклад, Свенна, згідно з яким $(k + 1)$ спробна точка визначається за рекурентною формулою

$$x_{k+1} = x_k + 2^k \Delta, \quad k = 0, 1, 2, \ldots, \tag{2.1}$$

де x_0 – довільно обрана початкова точка;

Δ – величина кроку, що підбирається певним чином.

Знак Δ визначається шляхом порівняння значень $f(x_0)$, $f(x_0 + |\Delta|)$ та $f(x_0 - |\Delta|)$.

- Якщо $f(x_0 - |\Delta|) > f(x_0) > f(x_0 + |\Delta|)$, то, згідно з припущенням про унімодальність, точка мінімуму має розміщуватися правіше точки x_0, і величина Δ має мати додатне значення.
- Якщо $f(x_0 - |\Delta|) < f(x_0) < f(x_0 + |\Delta|)$, то Δ має від'ємне значення.
- Якщо ж $f(x_0 - |\Delta|) > f(x_0) < f(x_0 + |\Delta|)$, то точка мінімуму лежить між $x_0 - |\Delta|$ і $x_0 + |\Delta|$, і пошук граничних точок завершено.
- Якщо $f(x_0 - |\Delta|) < f(x_0) > f(x_0 + |\Delta|)$, то маємо протиріччя припущенню про унімодальність.

2) *Етап зменшення інтервалу*

Після того, як встановлено межі інтервалу, що містить точку оптимуму, можна застосувати більш складну процедуру зменшення інтервалу пошуку для отримання уточнених оцінок координат оптимуму. Величина підінтервалу, що виключається на кожному кроці, залежить від розміщення пробних точок x_1 та x_2 всередині інтервалу пошуку. Оскільки розміщення точки оптимуму апріорі невідоме, доцільно припустити, що розміщення пробних точок має забезпечувати зменшення інтервалу у одному й тому самому співвідношенні. Крім того, для підвищення ефективності алгоритму, необхідно прагнути, щоб це співвідношення було максимальним.

Приклад Застосувавши формулу Свенна, визначити інтервал пошуку мінімуму функції $f(x) = (100 - x)^2$ при заданій початковій точці $x_0 = 30$ та величині кроку $|\Delta| = 5$.

Обчислюємо значення функції $f(x_0)$, $f(x_0 + |\Delta|)$ та $f(x_0 - |\Delta|)$:

$$f(x_0) = f(30) = 4900; \quad f(x_0 + |\Delta|) = f(35) = 4225; \quad f(x_0 - |\Delta|) = f(25) = 5625.$$

Оскільки $f(x_0 - |\Delta|) > f(x_0) > f(x_0 + |\Delta|)$, то величина Δ має бути додатною, а координата точки мінімуму x^* має бути більше 30.

Застосовуємо формулу (2.1).

$$x_1 = x_0 + \Delta = 30 + 5 = 35; \quad f(x_1) = f(35) = 4225 < f(x_0);$$
$$x_2 = x_1 + 2\Delta = 35 + 2 * 5 = 45; \quad f(x_2) = f(45) = 3025 < f(x_1);$$
$$x_3 = x_2 + 2^2\Delta = 45 + 4 * 5 = 65; \quad f(x_3) = f(65) = 1225 < f(x_2);$$
$$x_4 = x_3 + 2^3\Delta = 65 + 8 * 5 = 105; \quad f(x_4) = f(105) = 25 < f(x_3);$$
$$x_5 = x_4 + 2^4\Delta = 105 + 16 * 5 = 185; \quad f(x_5) = f(185) = 7225 > f(x_4).$$

У останній точці значення функції вже перевищує попереднє. Отже, точка мінімуму міститься в такому інтервалі невизначеності: $65 < x^* < 185$.

2.1.2 Метод ділення інтервалу навпіл

Цей метод дозволяє виключити з розглядання в точності половину інтервалу на кожній ітерації. Алгоритм розрахунку наведений нижче.

Завдання Знайти мінімальне значення функції $f(x)$ на відрізку $[a, b]$.

Крок 1 Положити $x_m = (a + b) / 2$; довжина інтервалу $L = b - a$. Обчислити значення $f(x_m)$.

Крок 2 Покласти $x_1 = a + L / 4$; $x_2 = b - L / 4$. Обчислити значення $f(x_1)$ та $f(x_2)$.

Крок 3 Порівняти $f(x_1)$ та $f(x_m)$.

Якщо $f(x_1) < f(x_m)$, то виключити інтервал $(x_m, b]$, далі покласти $b = x_m$, $x_m = x_1$. Отже, середньою точкою нового інтервалу пошуку стає колишня точка x_1.

Перейти до кроку 5.

Якщо $f(x_1) \geq f(x_m)$, то перейти до кроку 4.

Крок 4 Порівняти $f(x_2)$ та $f(x_m)$.

Якщо $f(x_2) < f(x_m)$, то виключити інтервал $[a, x_m)$, далі покласти $a = x_m$, $x_m = x_2$.

Перейти до кроку 5.

Якщо $f(x_2) \geq f(x_m)$, то виключити $[a, x_1)$ та $(x_2, b]$, покласти $a = x_1$, $b = x_2$.

Перейти до кроку 5.

Крок 5. Обчислити $L = b - a$. Якщо величина |L| менше заздалегідь встановленої точності ε, тобто достатньо мала, то закінчити пошук. У протилежному випадку повернутися до кроку 2.

Зауваження

1 Середня точка інтервалів, що послідовно виключаються, завжди збігається з однією з пробних точок x_1, x_2 або x_m, що знайдені на попередній ітерації. Отже, на кожній ітерації потрібно не більше двох обчислень значень функції.

2 Як бачимо з алгоритму, з кожних трьох значень цільової функції f, обчислених в інтервалі пошуку, надалі використовується тільки два, а третє не дає додаткової інформації і надалі не використовується.

3 Якщо проведено n обчислень значень функції, то довжина отриманого інтервалу невизначеності складає $(1 / 2)^{n / 2}$ довжини вихідного інтервалу.

Приклад Знайти мінімум функції $f(x) = (100 - x)^2$ методом ділення інтервалу навпіл. Інтервал пошуку $60 \le x \le 150$.

Ітерація 1

Крок 1 Тут $a = 60$; $b = 150$; $L = 150 - 60 = 90$;

$x_m = (60 + 90) / 2 = 105$; $f(x_m) = 25$.

Крок 2 Обчислюємо $x_1 = 60 + 90 / 4 = 82.5$; $x_2 = 150 - 90 / 4 = 127.5$.

Крок 3 Маємо $f(x_1) = 306.25 < f(x_m)$; $f(x_2) = 756.25 > f(x_m)$

Крок 4 – 5 Отже, виключаємо з розгляду інтервали (60; 82.5) та (127.5; 150). Довжина інтервалу пошуку зменшується з 90 до 45.

Ітерація 2

Беремо $a = 82.5$; $b = 127.5$; $x_m = 105$.

Обчислюємо $x_1 = 82.5 + 45 / 4 = 93.75$; $x_2 = 127.5 - 45 / 4 = 116.25$.

Маємо $f(x_1) = 39.06 > f(x_m)$; $f(x_2) = 264.06 > f(x_m)$

Новий інтервал невизначеності складає від 93.75 до 116.25. Довжина інтервалу складає 22.5.

Ітерація 3

Беремо $a = 93.75$; $b = 116.25$; $x_m = 105$.

Обчислюємо $x_1 = 93.75 + 22.5 / 4 = 99.38$; $x_2 = 116.25 - 22.5 / 4 = 110.63$.

Маємо $f(x_1) = 0.39 < f(x_m)$; $f(x_2) = 113.00 > f(x_m)$

Новий інтервал невизначеності складає від 99.38 до 110.63. Довжина інтервалу складає 11.25, тобто 1/8 початкового інтервалу невизначеності.

2.1.3 Метод Фібоначчі

Назва методу походить від відомого ряду Фібоначчі, тобто ряду натуральних чисел, в якому кожний наступний член ряду є сумою двох попередніх членів: 1, 1, 2, 3, 5, 8, 13, …, і т.д. [2].

Припустимо, що потрібно визначити мінімум якомога точніше, тобто з найменшим можливим інтервалом невизначеності, але при цьому можна виконати тільки n обчислень функції. Як варто вибрати n точок, у яких обчислюється функція? З першого погляду здається зрозумілим, що не слід шукати розв'язання для всіх точок, отриманих у результаті експерименту. Навпаки, треба спробувати зробити так, щоб значення функції, отримані в попередніх експериментах, визначали положення подальших точок. Дійсно, знаючи значення функції, ми тим самим маємо інформацію про саму функцію і положення її мінімуму і використовуємо цю інформацію надалі.

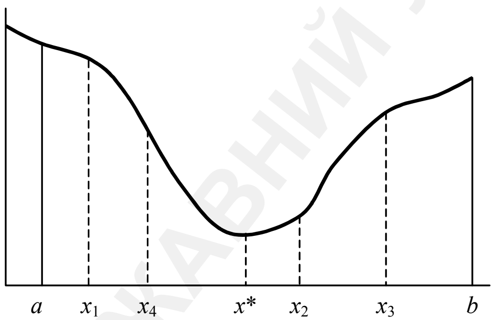


Рисунок 2.1 – Графічне представлення функції

Припустимо, що існує інтервал невизначеності (x_1, x_3) і відоме значення функції f (x_2) всередині цього інтервалу (рис. 2.1). Якщо можна обчислити функцію всього один раз у точці x_4, то де варто помістити точку x_4, для того, щоб отримати найменший можливий інтервал невизначеності?

Будемо вважати $x_2 - x_1 = L$ і $x_3 - x_2 = R$, причому $L > R$, як показано на рис. 2.2, 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 $x_3 - x_4$.

Оскільки невідомо, яка з цих ситуацій буде мати місце, виберемо x_4 таким чином, щоб мінімізувати найбільшу з довжин $x_3 - x_4$ і $x_2 - x_1$. Досягти цього можна, зробивши довжини $x_3 - x_4$ і $x_2 - x_1$ рівними, тобто помістивши x_4 усередині інтервалу симетрично відносно точки x_2, що вже є усередині інтервалу. Будь-яке інше положення точки x_4 може призвести до того, що отриманий інтервал буде більшим за L. Поміщаючи x_4 симетрично відносно x_2, ми нічим не ризикуємо в будь-якому випадку.

Якщо виявиться, що можна виконати ще одне обчислення функції, то варто застосувати описану процедуру до інтервалу (x_1, x_2), у якому вже є значення функції, обчислене в точці x_4, чи до інтервалу (x_4, x_3), у якому вже є значення функції, обчислене в точці x_2. Отже, стратегія зрозуміла із самого початку. Потрібно помістити наступну точку усередині інтервалу невизначеності симетрично відносно точки, що вже знаходиться там. Парадоксально, але щоб зрозуміти, як варто починати обчислення, необхідно розібратися у тому, як його варто закінчити.

На n-му обчисленні n-у точку варто помістити симетрично відносно (n – 1)-ї точки. Положення цієї останньої точки в принципі залежить від нас. Для того щоб отримати найбільше зменшення інтервалу на даному етапі, варто розділити навпіл попередній інтервал. Тоді точка x_n буде збігатися з точкою x_{n-1}. Однак при цьому ми не одержуємо ніякої нової інформації. Звичайно точки x_{n-1} і x_n віддалені одна від одної на достатню відстань, щоб визначити, у якій половині, лівій чи правій, знаходиться інтервал невизначеності. Вони розміщуються на відстані ε / 2 по обидва боки від середини відрізка L_{n-1}; можна самим задати величину ε чи вибрати цю величину рівною мінімально можливій відстані між двома точками.

Інтервал невизначеності буде мати довжину L_n, отже,

$L_{n-1} = 2\,L_n - \varepsilon$ (рис. 2.2, нижня частина).

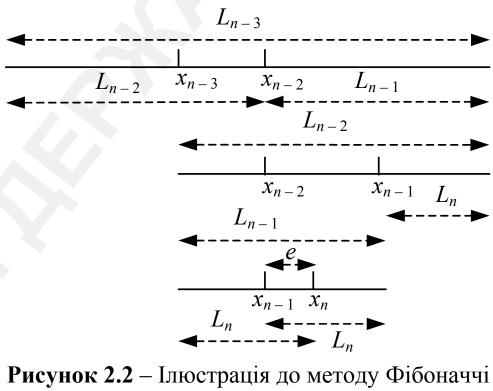


Рисунок 2.2 – Ілюстрація до методу Фібоначчі

На попередньому етапі точки x_{n-1} і x_{n-2} повинні бути розміщені симетрично усередині інтервалу L_{n-2} на відстані L_{n-1} від кінців цього інтервалу. Отже,

$L_{n-2} = L_{n-1} + L_n$ (рис. 2.2, середня частина).

Зауваження З рисунка зрозуміло, що на передостанньому етапі x_{n-2} залишається як внутрішня точка.

Аналогічно

$L_{n-3} = L_{n-2} + L_{n-1}$ (рис. 2.2, верхня частина)

У загальному випадку

$$L_{j-1} = L_j + L_{j+1} \quad \text{при } 1 < j < n. \tag{2.2}$$

Таким чином,

$$\begin{aligned} &L_{n-1} = 2\,L_n - \varepsilon, \\ &L_{n-2} = L_{n-1} + L_n = 3\,L_n - \varepsilon, \\ &L_{n-3} = L_{n-2} + L_{n-1} = 5\,L_n - 2\,\varepsilon, \\ &L_{n-4} = L_{n-3} + L_{n-2} = 8\,L_n - 3\,\varepsilon \quad \text{і т.д.} \end{aligned}$$

Якщо визначити послідовність чисел Фібоначчі в такий спосіб:

$F_0 = 1$, $F_1 = 1$ та $F_k = F_{k-1} + F_{k-2}$ для $k = 2, 3, ...$, то

$$L_{n-j} = F_{j+1}\,L_n - F_{j-1}\,\varepsilon, \quad j = 1, 2, ..., n-1. \tag{2.3}$$

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

$L_1 = F_n\,L_n - \varepsilon\,F_{n-2}$, тобто

$$L_n = \frac{L_1}{F_n} + \varepsilon\,\frac{F_{n-2}}{F_n}\,. \tag{2.4}$$

Отже, зробивши n обчислень функції, ми зменшимо початковий інтервал невизначеності в $1\,/\,F_n$ раз порівняно з його початковою довжиною (зневажаючи ε), і це – найкращий результат.

Якщо пошук почато, то його нескладно продовжити, використовуючи описане вище правило симетрії. Отже, необхідно знайти положення першої точки, що міститься 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:

$$L_2 = F_{n-1}\,L_n - \varepsilon\,F_{n-3} =$$

$$= F_{n-1}\frac{L_1}{F_n} + \varepsilon\frac{(F_{n-1}F_{n-2} - F_nF_{n-3})}{F_n} = \frac{F_{n-1}}{F_n}L_1 + \frac{(-1)^n\varepsilon}{F_n}. \qquad (2.5)$$

Після того як знайдене положення першої точки, числа Фібоначчі більше не потрібні. Використовуване значення ε може визначатися з практичних міркувань. Воно має бути менше L_1 / F_{n+1}, у протилежному разі ми будемо дарма витрачати час на обчислення функції.

Таким чином, пошук методом Фібоначчі, що названий так через появу під час пошуку чисел Фібоначчі, є ітераційною процедурою. У процесі пошуку інтервалу (x_1, x_2) з точкою x_2, що вже лежить у 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

$$x_4 = x_1 - x_2 + x_3. \qquad (2.6)$$

Якщо $f(x_2) = f_2$ і $f(x_4) = f_4$, то можна розглянути чотири випадки (рис. 2.3):

а)
$x_4 < x_2$
$f_4 < f_2$
Новий інтервал (x_1, x_2)
містить точку x_4;

б)
$x_4 > x_2$
$f_4 < f_2$
Новий інтервал (x_2, x_3)
містить точку x_4;

в)
$x_4 < x_2$
$f_4 > f_2$
Новий інтервал (x_4, x_3)
містить точку x_2;

г)
$x_4 > x_2$
$f_4 > f_2$
Новий інтервал (x_1, x_4)
містить точку x_2.

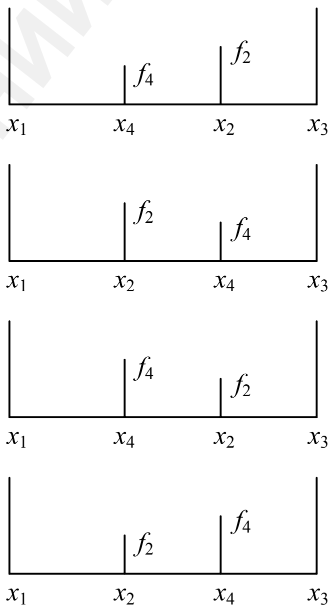


Рисунок 2.3 – Ілюстрація до методу Фібоначчі

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.

Приклад Знайти мінімум функції $f\,(x) = (100 - x)^2$ методом Фібоначчі. Інтервал пошуку $60 \le x \le 150$. Дозволяється зробити 6 обчислень значень функції (з урахуванням значень функції на границях інтервалу пошуку).

Для зручності розрахунків перейдемо до інтервалу одиничної довжини. Введемо змінну $x = 90\omega + 60$, тобто $\omega = (x - 60)\,/\,90$. Тоді задача полягає у відшуканні мінімуму функції $f\,(\omega) = (40 - 90\omega)^2$ при обмеженні $0 \le \omega \le 1$.

Ряд Фібоначчі $F_0 = F_1 = 1$; $F_2 = 2$; $F_3 = 3$; $F_4 = 5$; $F_5 = 8$; $F_6 = 13$ і т.д.

Оскільки дозволяється зробити 6 обчислень, $n = 6$, а шостий член ряду Фібоначчі F_n дорівнює 13. Початкова довжина інтервалу невизначеності L_1 дорівнює 1.

Спочатку отримуємо значення функції на краях інтервалу

$$f\,(0) = 1600, \quad f\,(1) = 2500.$$

Ітерація 1 (обчислення 3)

За формулою (2.5), вважаючи $\varepsilon = 0$, відшукуємо положення нової точки, в якій треба обчислити значення функції

$$L_2 = \frac{F_5}{F_6} L_1 = \frac{8}{13} \cdot 1 = 0.615.$$

Відповідне значення функції $f\,(0.615) = 236.7$.

Введемо позначення $\omega_1 = 0$, $\omega_2 = 0.615$, $\omega_3 = 1$.

Ітерація 2 (обчислення 4)

Згідно з формулою (2.6) та рис. 2.3, положення нової точки ω_4 вибираємо таким чином, щоб вона розміщувалась симетрично відносно точки ω_2 (щоб нові інтервали невизначеності були однаковими) (див. рис. 2.4):

$$\omega_2 - \omega_1 = \omega_3 - \omega_4.$$

$\omega_4 = \omega_3 - \omega_2 + \omega_1 = 1 - 0.615 + 0 = 0.385$.

Відповідне значення функції $f\,(0.385) = 29.0$.

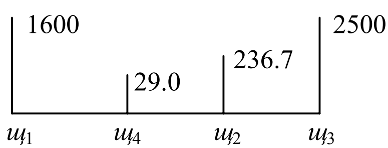


Рисунок 2.4 – Ілюстрація до прикладу методу Фібоначчі

Точку ω_3 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $L_2 = \omega_2 - \omega_1 = 0.615$.

Ітерація 3 (обчислення 5)

Точку ω_5 розміщуємо симетрично відносно точки ω_4, щоб нові інтервали невизначеності були однаковими

$$\omega_4 - \omega_1 = \omega_2 - \omega_5.$$

$\omega_5 = \omega_2 - \omega_4 + \omega_1 = 0.615 - 0.385 + 0 = 0.23.$

Відповідне значення функції $f(0.23) = 369.8$.

Точку ω_1 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $L_3 = \omega_2 - \omega_5 = 0.385$.

Ітерація 4 (обчислення 6)

Точку ω_6 розміщуємо симетрично відносно точки ω_4, щоб нові інтервали невизначеності були однаковими

$$\omega_6 - \omega_5 = \omega_2 - \omega_4.$$

$\omega_6 = \omega_2 - \omega_4 + \omega_5 = 0.615 - 0.385 + 0.23 = 0.46.$

Відповідне значення функції $f(0.46) = 2.37$.

Точку ω_5 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $0.385 \le \omega \le 0.615$. Довжина інтервалу невизначеності $L_4 = \omega_2 - \omega_4 = 0.23$.

Відповідно, довжина інтервалу невизначеності функції $f(x)$ після шести обчислень складає $(150 - 60) / F_6 = 90 / 13 = 6.92$. Отримана точка мінімального значення функції $x = 90 * 0.46 + 60 = 101.4$.

Завдання до підрозділу 2.1.3 Виконати пошук мінімуму або максимуму функції 1 додатка за методом Фібоначчі.

2.1.4 Метод “золотого перетину”

Не завжди можна заздалегідь визначити, скільки разів доведеться обчислювати функцію. У методі Фібоначчі це потрібно знати для визначення L_2, тобто положення початкової точки (див. рівняння (2.5)) [2].

Метод “золотого перетину” майже настільки ж ефективний, як і метод Фібоначчі, однак при цьому не потрібно знати n – кількість обчислень функції, що визначається спочатку. Після того як виконано j обчислень, виходячи з тих самих міркувань, що і раніше (див. рівняння (2.4)), записуємо

$$L_{j-1} = L_j + L_{j+1}. \tag{2.7}$$

Однак якщо n є невідомим, то ми не можемо використовувати умову $L_{n-1} = 2\,L_n - \varepsilon$. Якщо відношення наступних інтервалів буде постійним, тобто

$$\frac{L_{j-1}}{L_j} = \frac{L_j}{L_{j+1}} = \frac{L_{j+1}}{L_{j+2}} = ... = \tau\ , \tag{2.8}$$

то

$$\frac{L_{j-1}}{L_j} = 1 + \frac{L_{j+1}}{L_j}\,,$$

тобто $\tau = 1 + 1 / \tau$.

Таким чином, $\tau^2 - \tau - 1 = 0$, звідки $\tau = (1 + 5^{1/2}) / 2 \approx 1.618033989$. Тоді

$$\frac{L_{j-1}}{L_{j+1}} = \tau^2\,, \qquad \frac{L_{j-2}}{L_{j+1}} = \tau^3 \qquad \text{і т.д.}$$

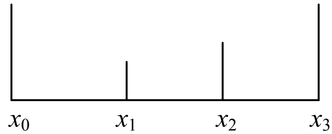
Отже, $\dfrac{L_1}{L_n} = \tau^{n-1}$, тобто

$$L_n = \frac{L_1}{\tau^{n-1}}\ . \tag{2.9}$$

У результаті аналізу двох розглянутих значень функції буде визначений той інтервал, що має досліджуватися надалі. Цей інтервал буде містити одну з попередніх точок і наступну точку, що розміщуються симетрично їй. Перша точка знаходиться на відстані L_1 / τ від одного кінця інтервалу, друга – на такій самій відстані від іншого. Оскільки $\lim\limits_{n\to\infty} F_{n-1} / F_n = 1/n$, то з рівняння (2.5) бачимо, що пошук методом “золотого перетину” є граничною формою пошуку методом Фібоначчі. Назва “золотий перетин” пішла від назви відношення в рівнянні (2.8). Бачимо, що L_{j-1} поділяється на дві частини так, що відношення цілого до більшої частини дорівнює відношенню більшої частини до меншої, тобто дорівнює так званому “золотому відношенню”.

Таким чином, якщо відшукується інтервал (x_0, x_3) і є два значення функції f_1 і f_2 у точках x_1 і x_2, то варто розглянути два випадки (рис. 2.5).

а)
$f_1 < f_2$
Новий інтервал (x_0, x_2);



б)
$f_1 > f_2$
Новий інтервал (x_1, x_3).

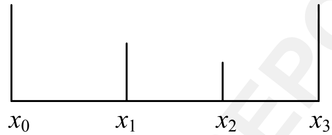


Рисунок 2.5 – Ілюстрація до методу “золотого перетину”

а) $f_1 < f_2$. Новий інтервал (x_0, x_2);
б) $f_1 > f_2$. Новий інтервал (x_1, x_3).

Покроковий алгоритм методу пошуку мінімуму на відрізку $[a, b]$:

Крок 1 Обчислюємо коефіцієнт дроблення відрізка $\tau \approx 1.618$.
Крок 2 $x_1 = a + (1 - 1 / \tau)\,(b - a)$, обчислити $f(x_1)$.
Крок 3 $x_2 = a + 1 / \tau\,(b - a)$, обчислити $f(x_2)$.
Крок 4
Якщо $|x_2 - x_1| < \varepsilon$, де ε – задане відхилення, то $x_m = (x_1 + x_2) / 2$, обчислити $f(x_m)$ і закінчити пошук.
Якщо $|x_2 - x_1| > \varepsilon$, то перейти до кроку 5.
Крок 5
Якщо $f(x_1) > f(x_2)$, то виключити з розгляду інтервал (a, x_1), далі встановити $a = x_1$, $x_1 = x_2$ і $f(x_1) = f(x_2)$. Перейти до кроку 3, потім до кроку 4.
Якщо $f(x_1) < f(x_2)$, то виключити з розгляду інтервал (x_2, b), далі встановити $b = x_2$, $x_2 = x_1$ і $f(x_2) = f(x_1)$. Перейти до кроку 2 і 4.

Відзначимо, що після перших двох обчислень значень функції кожне наступне обчислення дозволяє виключити підінтервал, величина якого складає $(1 - 1 / \tau)$-у частку довжини інтревалу пошуку. Отже, якщо початковий інтервал має одиничну довжину, то довжина інтервалу невизначеності після n обчислень значень функції складає $(1 / \tau)^{n-1}$.

Таким чином, застосування методів виключення інтервалів накладає єдине обмеження на досліджувану цільову функцію – унімодальність.

Отже, розглянуті методи можна використовувати для аналізу як неперервних, так і розривних, і дискретних функцій. Логічна структура пошуку заснована на простому порівнянні значень функції в двох пробних точках.

Приклад Знайти мінімум функції $f\ (x) = (100 - x)^2$ методом “золотого перетину”. Інтервал пошуку: $60 \le x \le 150$. Зробити 6 обчислень значень функції.

Як і в попередньому прикладі, для зручності розрахунків перейдемо до інтервалу одиничної довжини. Введемо змінну $x = 90\omega + 60$, тобто $\omega = (x - 60) / 90$. Тоді задача полягає у відшуканні мінімуму функції $f(\omega) = (40 - 90\omega)^2$ при обмеженні $0 \le \omega \le 1$.

Спочатку отримуємо значення функції на краях інтервалу:

$$f(0) = 1600, \quad f(1) = 2500.$$

Ітерація 1 та 2 (обчислення 3 та 4)

За формулою (2.9) відшукуємо довжину нового інтервалу невизначеності:

$$L_2 = \frac{L_1}{\tau} = \frac{1}{1.618} = 0.618.$$

Положення нових точок (див. крок 2 та крок 3)

$\omega_1 = a + (1 - 1 / \tau)\ (b - a) = 0 + (1 - 0.618) * 1 = 0.382.$

Відповідне значення функції $f(0.382) = 31.6.$

$\omega_2 = a + 1 / \tau\ (b - a) = 0 + 0.618 * 1 = 0.618.$

Відповідне значення функції $f(0.618) = 244.0.$

Введемо позначення $\omega_0 = 0$, $\omega_3 = 1$.

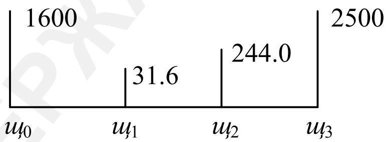


Рисунок 2.6 – Ілюстрація до прикладу методу “золотого перетину”

Див. крок 5. Точку ω_3 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $L_2 = \omega_2 - \omega_0 = 0.618$.

Ітерація 3 (обчислення 5)

Положення нової точки (див. крок 2)

$\omega_4 = a + (1 - 1 / \tau)\ (b - a) = 0 + (1 - 0.618) * 0.618 = 0.236.$

Відповідне значення функції $f(0.236) = 352.$

Див. крок 5. Точку ω_0 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $L_3 = \omega_2 - \omega_4 = 0.382$.

Ітерація 4 (обчислення 6)

Положення нової точки (див. крок 3)

$\omega_5 = a + 1 / \tau (b - a) = 0.236 + 0.618 * 0.382 = 0.472$.

Відповідне значення функції $f(0.472) = 6.15$.

Див. крок 5. Точку ω_4 відкидаємо, оскільки значення функції в цій точці є найбільшим. Отже, новий інтервал невизначеності $0.382 \le \omega \le 0.618$. Довжина інтервалу невизначеності $L_4 = \omega_2 - \omega_1 = 0.236$.

Відповідно, довжина інтервалу невизначеності функції $f(x)$ після шести обчислень складає $(150 - 60) / \tau^{6-1} = 90 / 11.1 = 8.12$. Отримана точка мінімального значення функції: $x = 90 * 0.472 + 60 = 102.5$.

Завдання до підрозділу 2.1.4 Виконати пошук мінімуму або максимуму функції 1 додатка за методом "золотого перетину".

2.1.5 Порівняння методів виключення інтервалів

Позначимо довжину початкового інтервалу невизначеності через L_1, а довжину інтервалу, отриманого в результаті n обчислень значень функції – через L_n. Як показник ефективності того чи іншого методу виключення інтервалів введемо в розгляд характеристику відносного зменшення вихідного інтервалу $Fr(n) = L_n / L_1$.

Нагадаємо, що при використанні методу ділення інтервалу навпіл, довжина отриманого інтервалу складає $L_1 (0.5)^{n/2}$, в методі "золотого перетину" – $L_1 / (1.618)^{n-1}$, у методі Фібоначчі – L_1 / F_n.

Отже, відносне зменшення інтервалу після n обчислень значень функції дорівнює:

$Fr(n) = (0.5)^{n/2}$ – метод ділення інтервалу навпіл;

$Fr(n) = L_1 / (1.618)^{n-1}$ – метод "золотого перетину";

$Fr(n) = L_1 / F_n$ – метод Фібоначчі.

Результати розрахунку за цими формулами наведено у таблиці 2.1.

Таблиця 2.1 – Величини відносного зменшення інтервалу невизначеності

Метод пошуку	Кількість обчислень функції				
	$n = 2$	$n = 4$	$n = 6$	$n = 8$	$n = 10$
Метод ділення інтервалу навпіл	0.5	0.25	0.125	0.063	0.031
Метод "золотого перетину"	0.618	0.236	0.090	0.034	0.013
Метод Фібоначчі	0.5	0.2	0.077	0.029	0.011

З таблиці витікає, що пошук за допомогою методу Фібоначчі забезпечує найбільше відносне зменшення вихідного інтервалу при одній і тій самій кількості обчислень значень функції.

2.2 Методи поліноміальної інтерполяції

У двох попередніх розділах була зроблена спроба знайти малий інтервал, у якому знаходиться мінімум функції. У наступних двох розділах застосовується інший підхід. Використовується кілька значень функції у визначених точках для апроксимації функції звичайним поліномом принаймні в невеликій області значень. Потім положення мінімуму функції апроксимується положенням мінімуму полінома, оскільки останній обчислити простіше.

2.2.1 Квадратична інтерполяція

Якщо відомі значення функції $f(x)$ у трьох різних точках a, b, c, що дорівнюють відповідно f_a, f_b, f_c, то функція $f(x)$ може бути апроксимована квадратичною функцією

$$\varphi(x) = A x^2 + B x + C, \tag{2.10}$$

де коефіцієнти A, B і C визначаються з рівнянь

$$\begin{aligned} A a^2 + B a + C &= f_a, \\ A b^2 + B b + C &= f_b, \\ A c^2 + B c + C &= f_c. \end{aligned} \tag{2.11}$$

Після перетворень цих рівнянь маємо

$$\begin{aligned} A &= [(c-b) f_a + (a-c) f_b + (b-a) f_c] / \Delta, \\ B &= [(b^2-c^2) f_a + (c^2-a^2) f_b + (a^2-b^2) f_c] / \Delta, \\ C &= [b\, c\, (c-b) f_a + c\, a\, (a-c) f_b + a\, b\, (b-a) f_c] / \Delta, \end{aligned} \tag{2.12}$$

де $\Delta = (a-b)(b-c)(c-a)$. Ясно, що $\varphi(x)$ буде мати мінімум у точці $x = -B / 2A$, якщо $A > 0$. Отже, можна апроксимувати точку мінімуму функції $f(x)$ значенням

$$\delta = \frac{1}{2}\left[\frac{(b^2-c^2) f_a + (c^2-a^2) f_b + (a^2-b^2) f_c}{(b-c) f_a + (c-a) f_b + (a-b) f_c}\right]. \tag{2.13}$$

Цей метод може безпосередньо застосовуватися до функцій однієї змінної. Припустимо, що задано унімодальну функцію однієї змінної $f(x)$, початкова апроксимація положення мінімуму і довжина кроку H, що є величиною того самого порядку, що і відстань від точки A до точки дійсного мінімуму x^* (умова, яку не завжди просто задовольнити).

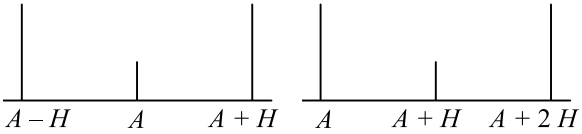


Рисунок 2.7

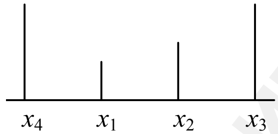


Рисунок 2.8

Обчислювальна процедура має такі кроки:

1 Обчислити $f(A)$ і $f(A + H)$.

2 Якщо $f(A) < f(A + H)$, то взяти за третю точку $A - H$ і обчислити $f(A - H)$. У протилежному разі за третю точку взяти $A + 2H$ і знайти $f(A + 2H)$ (рис. 2.5).

3 Використовуючи ці три точки, знайти δ з рівняння (2.13) і обчислити $f(\delta)$.

4 Якщо різниця між найменшим значенням функції і наступним найменшим значенням функції менша заданої точності, то процедура закінчується.

5 Якщо процедура не була завершена на кроці 4, то точка з найбільшим значенням звичайно відкидається, і ми повертаємося на крок 3. Але якщо, залишивши точку з найбільшим значенням функції, ми визначимо кінцеві межі інтервалу, у якому є мінімум, то варто дійсно залишити це значення і потім повернутися на крок 3. Наприклад, на рис. 2.8 залишені точки x_1, x_2 і x_4, а не точки x_1, x_2 і x_3.

Якщо точність ε задана занадто малою, то a, b, c, а також f_a, f_b, f_c будуть дуже близькі одне до одного і значення δ (див. рівняння (2.13)) може стати взагалі недосяжним. Щоб подолати ці труднощі, перепишемо рівняння (2.13) для другої і наступної інтерполяцій:

$$\delta = \frac{1}{2}(a+b) + \frac{0.5(f_a - f_b)(b-c)(c-a)}{(b-c)f_a + (c-a)f_b + (a-b)f_c}\,. \qquad (2.14)$$

Вищеописаний метод часто називають методом Пауела.

2.2.2 Інтерполяція вищих порядків

У загальному випадку функцію, що підлягає оптимізації, можна апроксимувати не тільки квадратичним поліномом, але й поліномом n-го ступеня. Якщо значення функції $f(x)$ у n різних точках $a, b, c, \ldots, n$, є відомими і дорівнюють відповідно f_a, f_b, f_c, …, f_n, то функція $f(x)$ може бути апроксимована поліномом такого виду

$$\varphi(x) = A\,x^n + B\,x^{n-1} + C\,x^{n-2} + \ldots + D\,x^2 + E\,x + F. \qquad (2.15)$$

Невідомі коефіцієнти A, B, C, …, D, E і F можна визначити з такої системи рівнянь:

$$\begin{aligned}
&A\,a^n + B\,a^{n-1} + C\,a^{n-2} + \ldots + D\,a^2 + E\,a + F = f_a,\\
&A\,b^n + B\,b^{n-1} + C\,b^{n-2} + \ldots + D\,b^2 + E\,b + F = f_b,\\
&A\,c^n + B\,c^{n-1} + C\,c^{n-2} + \ldots + D\,c^2 + E\,c + F = f_c,\\
&\ldots\\
&A\,f^n + B\,f^{n-1} + C\,f^{n-2} + \ldots + D\,f^2 + E\,f + F = f_f.
\end{aligned}$$

Цю систему лінійних алгебраїчних рівнянь (n рівнянь, n невідомих) можна розв’язати, наприклад, за допомогою відомого метода Гауса. Далі процедура пошуку мінімуму має відбуватися аналогічно методу Пауела, що описаний у попередньому підрозділі.

Відзначимо, однак, що інтерполяція вищих порядків є неекономічною через необхідність обчислень великої кількості значень функції. На практиці інтерполяція поліномом ступеня вище 3 використовується рідко. Перевага методів інтерполяції вищих порядків – можливість відшукувати мінімуми функції навіть у тому випадку, коли функція не є унімодальною.

2.3 Методи з використанням похідних. метод ньютона

Доцільно допустити, що ефективність пошукових процедур істотно підвищиться, якщо на додаток до умови безперервності ввести вимогу диференційованості цільової функції. Для функцій однієї змінної класичний підхід при пошуку значень x в точках перегину функції $f\ (x)$ полягає у розв’язанні рівняння

$$f'\,(x) = 0.$$

У тому випадку, якщо цільова функція містить члени, що містять, наприклад, x у третій і більш високих ступенях, то одержати аналітичне розв’язання рівняння $f'\ (x) = 0$ важко. У цих випадках доцільно використовувати чисельні методи знаходження коренів нелінійних рівнянь. У даному підрозділі ми розглянемо один з таких методів – метод Ньютона, відомий також як метод дотичних.

Цей метод орієнтований на знаходження кореня рівняння $f'\ (x) = 0$ в інтервалі $[a, b]$, такому, що знаки похідних $f'\ (a)$ та $f'\ (b)$ є протилежними. Тоді, в силу очевидних припущень про безперервність, буде існувати корінь x^* даного рівняння, причому $a < x^* < b$ (рис. 2.9).

Робота алгоритму починається з точки x_0, що являє початкове наближення кореня рівняння $f'\ (x) = 0$. Далі будується лінійна апроксимація функції $f'\ (x)$ у точці x_1, і точка, у якій апроксимуюча лінійна функція обертається в нуль, береться як наступне наближення (рис. 2.10). Якщо точка x_k прийнята як поточне наближення до оптимальної точки, то лінійна функція, що апроксимує функцію $f'(x)$ у точці x_k, записується у вигляді

$$f'(x - x_k) = f'(x_k) + f''(x_k)\,(x - x_k).$$

Порівнявши праву частину рівняння до нуля, отримаємо наступне наближення до шуканої точки.

Крок 1 Наступне наближення до стаціонарної точки x^* визначається за формулою

$$x_{k+1} = x_k - [f'(x_k) \,/\, f''(x_k)].$$

Крок 2 Обчислити $f'(x_{k+1})$, $f''(x_{k+1})$.

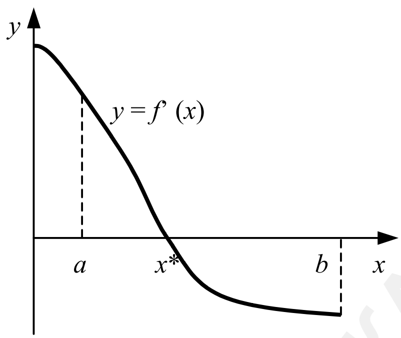


Рисунок 2.9

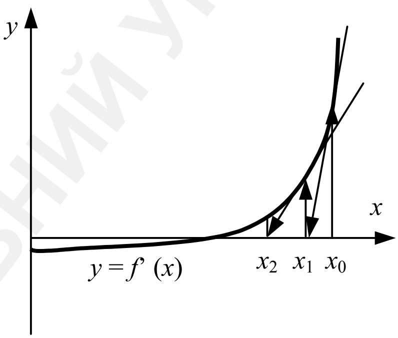


Рисунок 2.10

Крок 3 Якщо $|f'(x_{k+1})| < \varepsilon$, то закінчити пошук. У протилежному разі необхідно повернутися до кроку 1.

Як виявляється з алгоритму, цільова функція $f(x)$ повинна бути двічі диференційована.

Приклад Використовуючи метод Ньютона, відшукати мінімум функції $f(x) = 2x^2 + 16/x$. Початкова точка $x_1 = 1$. Потрібна точність $\varepsilon = 0.1$.

Знаходимо першу та другу похідні функції $f(x)$:

$$f'(x) = 4x - \frac{16}{x^2}; \quad f''(x) = 4 + \frac{32}{x^3}.$$

Ітерація 1. $x_1 = 1$; $f'(x_1) = -12$; $f''(x_1) = 36$.
Ітерація 2. $x_2 = x_1 - f'(x_1) \,/\, f''(x_1) = 1.33$; $f'(x_2) = -3.73$; $f''(x_2) = 17.6$.
Ітерація 3. $x_3 = x_2 - f'(x_2) \,/\, f''(x_2) = 1.54$; $f'(x_3) = -4.23$; $f''(x_3) = 12.7$.
Ітерація 4. $x_4 = x_3 - f'(x_3) \,/\, f''(x_3) = 1.87$; $f'(x_4) = -1.07$; $f''(x_4) = 8.89$.

Ітерація 5. $x_5 = x_4 - f'(x_4) / f''(x_4) = 1.99$; $f'(x_5) = -0.07$; $f''(x_5) = 8.05$.

Розрахунок можна зупинити, оскільки $|f'(x_5)| < \varepsilon$.

Завдання до підрозділу 2.3 Виконати пошук мінімуму або максимуму функції 1 додатка за методом Ньютона.

РОЗДІЛ 3. МЕТОДИ ПОШУКУ ДЛЯ ФУНКЦІЇ N ЗМІННИХ

3.1 Методи прямого пошуку

3.1.1 Попереднє обговорення. Метод покоординатного спуску

Одним із методів знаходження мінімуму функції *n* змінних є методи прямого пошуку. Методи прямого пошуку є методами, у яких використовуються тільки значення функції. Ми докладно розглянемо лише два таких методи. Практика показала, що ці два методи є ефективними та можуть бути застосовані для великої кількості прикладних задач [2].

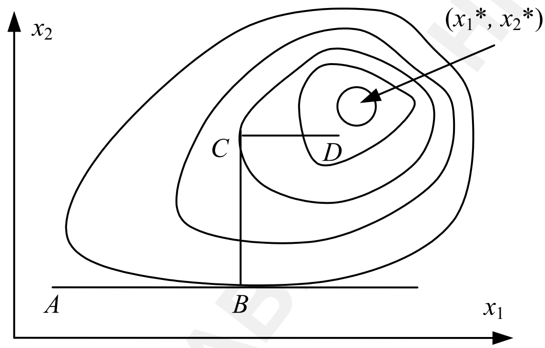


Рисунок 3.1 – Метод покоординатного спуску

Розглянемо функцію двох змінних. Її лінії рівня подані на рис. 3.1, а мінімум лежить у точці (x_1^*, x_2^*). Найпростішим методом пошуку є метод покоординатного спуску. З точки *A* робимо пошук мінімуму уздовж напрямку осі x_1 і, таким чином, знаходимо точку *B*, у котрої дотична до лінії постійного рівня паралельна осі x_1. Потім, роблячи пошук із точки *B* в напрямку осі x_2, отримуємо точку *C*, роблячи пошук паралельно осі x_2, одержуємо точку *D* і таке інше. Таким чином, ми приходимо до оптимальної точки. Цю ідею можна застосувати для функції *n* змінних.

Теоретично даний метод є ефективним у випадку єдиного мінімуму функції. Але на практиці він виявляється занадто повільним. Тому були розроблені більш складні методи, що використовують більше інформації на підставі вже отриманих значень функції.

3.1.2 Метод Хука-Дживса

Суть методів прямого пошуку полягає в переборі пробних точок. Зрозуміло, що основна мета побудови безлічі таких точок полягає у визначенні напрямку, у якому повинен проходити пошук. Найпростіший підхід полягає в тому, що пошук ведеться на основі рекурсивного перебору напрямків з довільно заданої безлічі. З іншого боку, можна побудувати стратегію пошуку, в рамках якої одне чи кілька напрямків пошуку уточнюється на кожній ітерації. До того ж необхідно гарантувати проведення пошуку по всій області, що розглядається.

Елементарним прикладом методу рекурсивного перебору на безлічі напрямків пошуку є метод циклічної зміни змінних, відповідно до якого кожного разу змінюється тільки одна змінна. При цьому підході безліч напрямків пошуку вибирається у вигляді безлічі координатних напрямків у просторі змінних задачі. Але така стратегія може виявитися не тільки не ефективною, але й привести до відсутності збіжності до точки локального оптимуму навіть при нескінченному числі ітерацій. Підвищення ефективності цього методу пов'язано з тією обставиною, що пошук, проведений у напрямку $(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i)$, де $f(\boldsymbol{x}_{i+1}) < f(\boldsymbol{x}_i)$, дозволяє істотно прискорити збіжність.

Ця умова була покладена в основу методу, розробленого Хуком і Дживсом у 1961 році. Дотепер цей метод є дуже ефективним і оригінальним. Метод Хука-Дживса характеризується нескладною стратегією пошуку, відносною простотою обчислень і невисоким рівнем вимог до пам'яті ЕОМ. До того ж це один з перших алгоритмів, у якому при визначенні нового напрямку пошуку враховується інформація, отримана на попередніх ітераціях. Процедура Хука-Дживса являє собою комбінацію досліджувального пошуку з циклічною зміною змінних та прискорювального пошуку за зразком.

Досліджувальний пошук орієнтований на виявлення характеру локального поводження цільової функції і визначення напрямку подальшого дослідження. Для проведення досліджувального пошуку необхідно знати величину кроку, яка може бути різною для різних координатних напрямків і змінюватися в процесі пошуку. Досліджувальний пошук починається в деякій вихідній точці. Якщо значення цільової функції в пробній точці не перевищує значення у вихідній точці, то крок пошуку розглядається як успішний. У протилежному разі треба повернутися в попередню точку і зробити крок у протилежному напрямку з подальшою перевіркою значення цільової функції. Після перебору всіх n координат досліджувальний пошук закінчується. Отриману в результаті точку називають базовою точкою.

Пошук за зразком. Пошук за зразком полягає в реалізації єдиного кроку з отриманої базової точки уздовж прямої, що з'єднує цю точку з попередньою базовою точкою. Нова точка зразка визначається відповідно до формули

$$P_i = b_i + 2\,(b_{i+1} - b_i).$$

Як тільки рух за зразком не приводить до зменшення цільової функції, точка P_i фіксується як тимчасова базова точка і знову проводиться досліджуючий пошук. Якщо в результаті виходить точка з меншим значенням цільової функції, чим у точці b_{i+1}, то вона розглядається як нова базова точка. З іншого боку, якщо досліджувальний пошук виявиться невдалим, необхідно повернутися в точку b_{i+1} і провести досліджувальний пошук для виявлення нового напрямку мінімізації. У результаті виникає ситуація, коли такий пошук не приводить до успіху. У цьому разі потрібно зменшити величину кроку шляхом введення деякого множника і відновити досліджувальний пошук. Пошук завершується, коли величина кроку стає досить малою.

Опис алгоритму методу Хука-Дживса [2]:

А Вибрати:

- початкову базову точку $\boldsymbol{b}_1$ з координатами $x_1, x_2, \ldots, x_n$;
- крок довжиною h_j для кожної змінної x_j, $j = 1, 2, ..., n$; крок для всіх змінних може бути прийнятий одним і тим самим;
- коефіцієнт зменшення кроку $a > 1$;
- параметр закінчення пошуку $\varepsilon > 0$.

Б 1 Обчислити $f(x)$ у базовій точці $\boldsymbol{b}_1$ для отримання відомостей про локальну поведінку функції $f(x)$ – отримуємо $f(\boldsymbol{b}_1)$.

2 Кожна змінна по черзі змінюється додаванням довжини кроку. Таким чином, ми обчислюємо значення функції $f(\boldsymbol{b}_1 + h_1\,\boldsymbol{e}_1)$, де $\boldsymbol{e}_1$ – одиничний вектор у напрямку осі x_1.

Якщо це призводить до зменшення значення функції, то $\boldsymbol{b}_1$ замінюється на $\boldsymbol{b}_1 + h_1\,\boldsymbol{e}_1$. У протилежному разі обчислюється значення функції $f(\boldsymbol{b}_1 - h_1\,\boldsymbol{e}_1)$, і якщо її значення зменшилося, то $\boldsymbol{b}_1$ замінюємо на $\boldsymbol{b}_1 - h_1\,\boldsymbol{e}_1$.

Якщо жоден із виконаних кроків не приводить до зменшення значення функції, то точка $\boldsymbol{b}_1$ залишається незмінною і розглядаються зміни в напрямку осі x_2, тобто обчислюємо значення функції $f(\boldsymbol{b}_1 + h_2\,\boldsymbol{e}_2)$ і т. д. Коли будуть розглянуті всі n змінних, ми будемо мати нову базову точку $\boldsymbol{b}_2$.

3 Якщо $\boldsymbol{b}_2 = \boldsymbol{b}_1$, тобто зменшення функції не було досягнуто, то дослідження повторюється навколо тієї самої базової точки $\boldsymbol{b}_2$, але зі зменшеною довжиною кроку: $h_j = h_j / a$.

4 Якщо $\boldsymbol{b}_2 \neq \boldsymbol{b}_1$, то виконується пошук за зразком.

В Під час пошуку за зразком використовується інформація, отримана в процесі дослідження, і мінімізація функції закінчується пошуком у напрямку, заданому зразком. Ця процедура проводиться таким чином:

1 Розумно рухатися з базової точки $\boldsymbol{b}_2$ у напрямку $\boldsymbol{b}_2 - \boldsymbol{b}_1$, оскільки пошук у цьому напрямку вже привів до зменшення значення функції. Тому обчислимо функцію в точці зразка

$$\boldsymbol{P}_1 = \boldsymbol{b}_1 + (\boldsymbol{b}_2 - \boldsymbol{b}_1). \tag{3.1}$$

У загальному разі

$$\boldsymbol{P}_i = \boldsymbol{b}_i + (\boldsymbol{b}_{i+1} - \boldsymbol{b}_i). \qquad (3.2)$$

2 Потім дослідження варто продовжувати навколо точки $\boldsymbol{P}_1$ ($\boldsymbol{P}_i$).

3 Якщо найменше значення на кроці В2 менше за значення в базовій точці $\boldsymbol{b}_2$ (у загальному випадку $\boldsymbol{b}_{i+1}$), то одержують нову базову точку $\boldsymbol{b}_2$ ($\boldsymbol{b}_{i+2}$), після чого варто повторити крок В1. У протилежному разі не робити пошук за зразком з точки $\boldsymbol{b}_2$ ($\boldsymbol{b}_{i+1}$), а продовжити дослідження в точці $\boldsymbol{b}_2$ ($\boldsymbol{b}_{i+1}$).

Г Завершити цей процес, коли довжини кроків h_j не будуть перевищувати заданого малого значення ε.

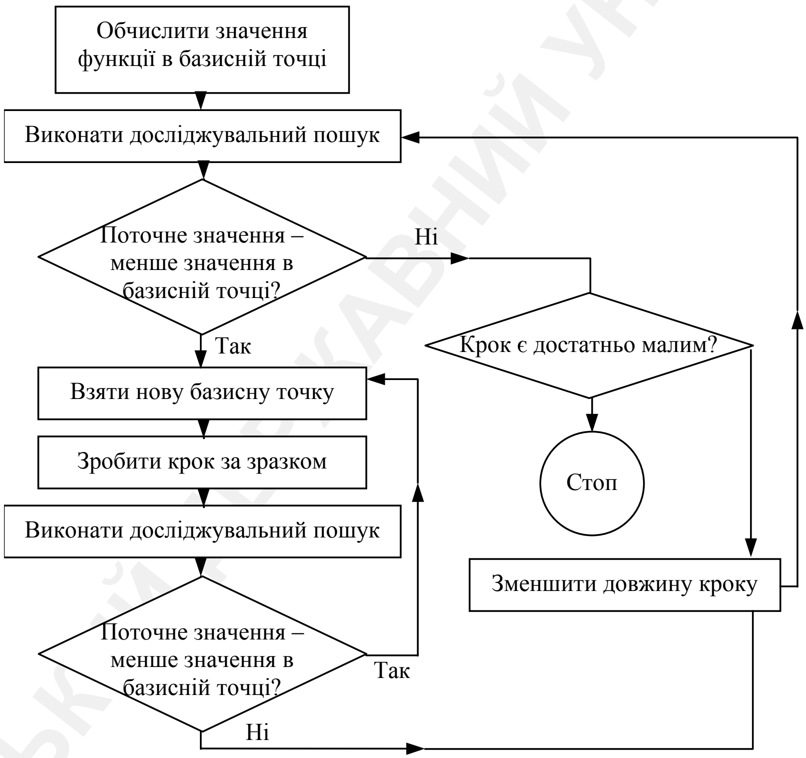


Рисунок 3.2 – Блок-схема методу Хука – Дживса

Рисунок 3.3 – Досліджувальний пошук у методі Хука – Дживса

Приклад Знайти мінімум функції $f(x) = (x_1 - 2)^2 + (x_2 - 5)^2 + (x_3 + 2)^4$ з точністю до $\varepsilon = 0.1$. Використати як початкову точку (4, –2, 3). Взяти початковий крок $h = 1$.

А Початкова точка: $\boldsymbol{x}_0$ (4, –2, 3).

Ітерація 1

Б 1 Обчислимо значення функції в початковій точці $\boldsymbol{b}_1$:

$$f(\boldsymbol{b}_1) = (4-2)^2 + ((-2)-5)^2 + (3+2)^4 = 678.$$

2 Виконуємо досліджувальний пошук, змінюючи по черзі кожну змінну додаванням довжини кроку. Обчислюємо відповідні значення функції.

$f(4+1, -2, 3) = f(5, -2, 3) = (5-2)^2 + ((-2)-5)^2 + (3+2)^4 = 683 > f(\boldsymbol{b}_1)$;
$f(4-1, -2, 3) = f(3, -2, 3) = (3-2)^2 + ((-2)-5)^2 + (3+2)^4 = 675 < f(\boldsymbol{b}_1)$.

Продовжуємо пошук від нової точки $\boldsymbol{b}_1{}^* = \boldsymbol{b}_1 - h_1\,\boldsymbol{e}_1$, тобто (3, –2, 3).

$f(3, -2+1, 3) = f(3, -1, 3) = (3-2)^2 + ((-1)-5)^2 + (3+2)^4 = 662 < f(\boldsymbol{b}_1{}^*)$;
$f(3, -2-1, 3) = f(3, -3, 3) = (3-2)^2 + ((-3)-5)^2 + (3+2)^4 = 690 > f(\boldsymbol{b}_1{}^*)$.

Продовжуємо пошук від нової точки $\boldsymbol{b}_1{}^* = \boldsymbol{b}_1{}^* + h_2\,\boldsymbol{e}_2$, тобто (3, –1, 3).

$f(3, -1, 3+1) = f(3, -1, 4) = (3-2)^2 + ((-1)-5)^2 + (4+2)^4 = 1333 > f(\boldsymbol{b}_1{}^*)$;
$f(3, -1, 3-1) = f(3, -1, 2) = (3-2)^2 + ((-1)-5)^2 + (2+2)^4 = 293 < f(\boldsymbol{b}_1{}^*)$.

Отримуємо точку $\boldsymbol{b}_2$ (3, –1, 2), в якій було досягнуто найменше значення функції на даному кроці.

В Виконуємо пошук за зразком.

За формулою (3.1) отримаємо координати точки $\boldsymbol{P}_1$ (4 + 2 (3 – 4); –2 + 2 (–1 – (–2)); 3 + 2 (2 – 3)). Маємо точку $\boldsymbol{P}_1$ (2, 0, 1).

Відповідне значення функції:

$$f(\boldsymbol{P}_1) = (2-2)^2 + (0-5)^2 + (1+2)^4 = 106 < f(\boldsymbol{b}_2).$$

Досягнуто зменшення значення функції порівняно з точкою $\boldsymbol{b}_2$.

Встановлюємо нові координати $\boldsymbol{b}_2 = \boldsymbol{P}_1$, і далі переходимо на наступну ітерацію.

Ітерація 2

Досліджувальний пошук навколо точки $\boldsymbol{b}_2$ (2, 0, 1).

$f(2+1, 0, 1) = f(3, 0, 1) = (3-2)^2 + (0-5)^2 + (1+2)^4 = 107 > f(\boldsymbol{b}_2)$;
$f(2-1, 0, 1) = f(1, 0, 1) = (1-2)^2 + (0-5)^2 + (1+2)^4 = 107 > f(\boldsymbol{b}_2)$.

Зменшення функції не досягнуто.

$f(2, 0+1, 1) = f(2, 1, 1) = (2-2)^2 + (1-5)^2 + (1+2)^4 = 97 < f(\boldsymbol{b}_2)$;
$f(2, 0-1, 1) = f(2, -1, 1) = (2-2)^2 + (-1-5)^2 + (1+2)^4 = 117 > f(\boldsymbol{b}_2)$.

Продовжуємо пошук від нової точки $\boldsymbol{b}_2{}^* = \boldsymbol{b}_2{}^* + h_2\,\boldsymbol{e}_2$, тобто (2, 1, 1).

$f(2, 1, 1+1) = f(2, 1, 2) = (2-2)^2 + (1-5)^2 + (2+2)^4 = 272 > f(\boldsymbol{b}_2{}^*)$;
$f(2, 1, 1-1) = f(2, 1, 0) = (2-2)^2 + (1-5)^2 + (0+2)^4 = 32 < f(\boldsymbol{b}_2{}^*)$.

Отримуємо точку $\boldsymbol{b}_3$ (2, 1, 0), в якій було досягнуто найменше значення функції на даному кроці.

В Виконуємо пошук за зразком.

За формулою (3.1) отримаємо координати точки $\boldsymbol{P}_2$ (3 + 2 (2 – 3); –1 + 2 (1 – (–1)); 2 + 2 (0 – 2)). Маємо точку $\boldsymbol{P}_2$ (1, 3, –2).

Відповідне значення функції

$$f(\boldsymbol{P}_2) = (1-2)^2 + (3-5)^2 + (-2+2)^4 = 5 < f(\boldsymbol{b}_3).$$

Досягнуто зменшення значення функції порівняно з точкою $\boldsymbol{b}_3$.

Встановлюємо нові координати $\boldsymbol{b}_3 = \boldsymbol{P}_2$, і далі переходимо на наступну ітерацію.

Для наступних ітерацій наведемо лише основні проміжні результати.

Ітерація 3

Досліджувальний пошук навколо точки $\boldsymbol{b}_3$ (1, 3, –2).

Знаходимо точку $\boldsymbol{b}_4$ (2, 4, –2), значення функції в якій складає $f(\boldsymbol{b}_4) = 1 < f(\boldsymbol{b}_3)$.

Пошук за зразком: отримуємо точку $\boldsymbol{P}_3$ (2, 7, –4), значення функції в якій складає $f(\boldsymbol{P}_3) = 20 > f(\boldsymbol{b}_4)$. Відкидаємо цю точку.

Ітерація 4

Досліджувальний пошук навколо точки $\boldsymbol{b}_4$ (2, 4, –2).

Знаходимо точку $\boldsymbol{b}_5$ (2, 5, –2), значення функції в якій складає $f(\boldsymbol{b}_5) = 0 < f(\boldsymbol{b}_4)$.

Пошук за зразком: отримуємо точку $\boldsymbol{P}_4$ (2, 6, –2), значення функції в якій складає $f(\boldsymbol{P}_4) = 1 > f(\boldsymbol{b}_5)$. Відкидаємо цю точку.

Ітерація 5

Досліджувальний пошук навколо точки $\boldsymbol{b}_5$ (2, 5, –2).

Використовуючи задану довжину кроку ($h = 1$), не вдається знайти точку, значення функції в якій менше, ніж в $\boldsymbol{b}_5$.

Переходимо на крок Б3.

Зменшуємо довжину кроку, встановлюємо $h = 0.5$.

Ітерація 6

Досліджувальний пошук навколо точки $\boldsymbol{b}_5$ (2, 5, –2).

Використовуючи задану довжину кроку ($h = 0.5$), не вдається знайти точку, значення функції в якій менше, ніж в $\boldsymbol{b}_5$.

Зменшуємо довжину кроку, встановлюємо $h = 0.25$.

Ітерація 7

Досліджувальний пошук навколо точки $\boldsymbol{b}_5$ (2, 5, –2).

Використовуючи задану довжину кроку ($h = 0.25$), не вдається знайти точку, значення функції в якій менше, ніж в $\boldsymbol{b}_5$.

Зменшуємо довжину кроку, встановлюємо $h = 0.1$.

Ітерація 8

Досліджувальний пошук навколо точки $\boldsymbol{b}_5$ (2, 5, –2).

Використовуючи задану довжину кроку (h = 0.1), не вдається знайти точку, значення функції в якій менше, ніж в $\boldsymbol{b}_5$.

Оскільки довжина кроку $h = \varepsilon$, закінчуємо розрахунок.

Отже, мінімум функції (з точністю до заданого значення ε) має місце в точці (2, 5, –2) та складає 0.

Завдання до підрозділу 3.1.2 Виконати пошук мінімуму або максимуму функції 2 додатка за методом Хука – Дживса.

3.1.3 Методи випадкового пошуку

У методах випадкового пошуку напрямок пошуку на кожному кроці вибирається випадковим чином. У даному підрозділі ми розглянемо один такий метод, що має назву адаптивний метод випадкового пошуку [3].

Задається початкова точка $\boldsymbol{x}_0$. Кожна наступна точка $\boldsymbol{x}_{k+1}$ обчислюється за формулою

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + h_k\, \boldsymbol{e}_k,$$

де $h_k > 0$ – величина кроку; $\boldsymbol{e}_k$ – випадковий вектор одиничної довжини, що визначає напрямок пошуку; k – номер ітерації. На поточній ітерації шляхом генерування випадкових векторів $\boldsymbol{e}_k$ отримуються точки, що розміщуються на гіперсфері радіуса h_k з центром в точці $\boldsymbol{x}_k$ (рис. 3.4). Якщо значення функції в отриманій точці не менше, ніж у центрі, крок вважається невдалим (точки $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ при пошуку з $\boldsymbol{x}_0$; $\boldsymbol{y}_1$, $\boldsymbol{y}_3$ при пошуку з $\boldsymbol{x}_1$). Якщо число невдалих кроків з поточної точки сягає певного числа M, подальший пошук продовжується з тієї самої точки, але з меншим кроком до тих пір, поки він не стане меншим заздалегідь заданої величини ε. Якщо ж значення функції в отриманій точці менше, ніж в центрі, крок вважається вдалим і в знайденому напрямку робиться збільшений крок, що відіграє роль прискорювального кроку (як при пошуку за методом Хука – Дживса). Якщо при цьому значення функції знов менше, ніж в центрі, напрямок вважається вдалим і подальший пошук продовжується з цієї точки (точки $\boldsymbol{z}_3 = \boldsymbol{x}_1$ при пошуку з $\boldsymbol{x}_0$, $\boldsymbol{z}_4 = \boldsymbol{x}_2$ при пошуку з $\boldsymbol{x}_1$). Якщо ж значення функції стало не менше, ніж в центрі, напрямок вважається невдалим і пошук продовжується зі старого центра (у точці $\boldsymbol{y}_2$ при пошуку з $\boldsymbol{x}_1$ функція менша, ніж в $\boldsymbol{x}_1$, а в точці $\boldsymbol{z}_2$ вже не менша, тому напрямок ($\boldsymbol{z}_2 - \boldsymbol{x}_1$) є невдалим).

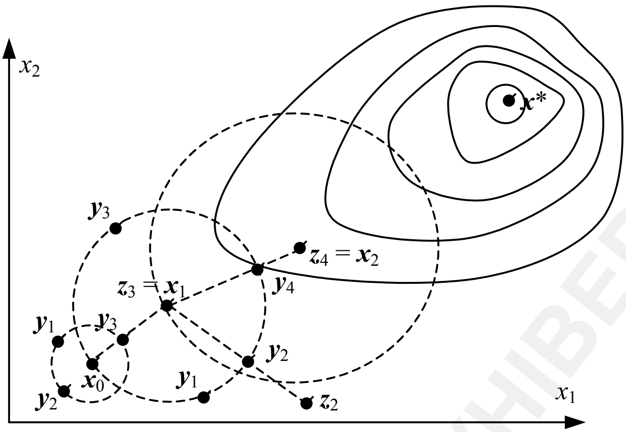


Рисунок 3.4 – Адаптивний метод випадкового пошуку

3.1.4 Метод Нелдера-Міда

Перші спроби розв’язання оптимізаційних задач без обмежень на основі прямого пошуку пов’язані з використанням одномірних методів оптимізації. Як правило, припустима область визначення цільової функції заміняється дискретною множиною (решіткою) точок простору змінних, а потім використовуються різні стратегії зменшення області, що містить розв’язок задачі.

Одна із стратегій пошуку, що викликає особливий інтерес, покладена в основу методу пошуку за симплексом, запропонованого Спендлі, Хекстом і Хімсвортом. Процедура симплексного пошуку базується на тому, що експериментальним зразком, що містить найменшу кількість точок, є регулярний симплекс. Регулярний симплекс у n-мірному просторі являє собою багатогранник, утворений $n + 1$ рівновіддаленими одна від одної точками-вершинами. Наприклад, у випадку двох змінних симплексом є трикутник; у тривимірному просторі симплекс являє собою тетраедр. Ідея методу полягає в порівнянні значень функції в $(n + 1)$ вершинах симплекса і переміщенні в напрямку оптимальної точки за допомогою ітераційної процедури.

Метод Нелдера-Міда (пошук за багатогранником, що деформується) є розвитком симплексного методу Спендлі, Хекста і Хімсворта. У симплексному методі, запропонованому спочатку, регулярний симплекс використовувався на кожному етапі. Нелдер і Мід запропонували кілька модифікацій цього методу, які допускають, щоб симплекси були неправильними. У результаті вийшов дуже надійний метод прямого пошуку, що є одним з найефективніших, якщо $n \le 6$.

Результати окремих чисельних випробувань показують, що метод Нелдера-Міда має достатню ефективність і високу надійність за умов наявності випадкових збурювань чи помилок під час визначення значень цільової функції.

У методі Спендлі, Хекста і Хімсворта симплекс пересувається за допомогою трьох основних операцій: *відбиття*, *розтягування* і *стискання*. Зміст цих операцій стане зрозумілим під час розгляду кроків процедури.

Алгоритм методу

А Знайдемо значення функції

$$f_1 = f(x_1), f_1 = f(x_2), ..., f_{n+1} = f(x_{n+1})$$

у вершинах симплекса.

Початковий симплекс (координати його вершин) можна вибрати довільно, на розсуд користувача. Формули, що наведені нижче, дозволяють побудувати регулярний симплекс навколо начальної (базової) точки $\boldsymbol{x}^{(0)}$ з масштабним множником h. Координати решти n вершин симплекса у n-вимірному просторі обчислюються так

$$\boldsymbol{x}^i = \begin{cases} x_j^{(0)} + \delta_1, & \textit{якщо } \; i \neq j, \\ x_j^{(0)} + \delta_2, & \textit{якщо } \; i = j, \end{cases}$$

де i та $j = 1, 2, 3, \ldots, n$; i – номер точки, j – індекс змінної.

Прирощення δ_1 та δ_2 залежать лише від n та вибраного масштабного множника h і визначаються за формулами:

$$\delta_1 = \left[\frac{\sqrt{n+1} + n - 1}{n\sqrt{2}}\right] h,$$

$$\delta_2 = \left[\frac{\sqrt{n+1} - 1}{n\sqrt{2}}\right] h.$$

Б Знайдемо найбільше серед f_i значення функції f_h, наступне за найбільшим значенням функції f_g, найменше значення функції f_l і відповідні їм точки $\boldsymbol{x}_h$, $\boldsymbol{x}_g$ і $\boldsymbol{x}_l$.

В Знайдемо центр тяжіння всіх точок, за винятком точки $\boldsymbol{x}_h$. Нехай центром тяжіння буде

$$\boldsymbol{x}_0 = \frac{1}{n}\sum_{i \neq h} \boldsymbol{x}_i \tag{3.3}$$

і обчислимо $f(\boldsymbol{x}_0) = f_0$.

Г Зручніше за все почати переміщення від точки $\boldsymbol{x}_h$. Відбивши точку $\boldsymbol{x}_h$ відносно точки $\boldsymbol{x}_0$, отримаємо нову точку $\boldsymbol{x}_r$ і знайдемо $f_r = f(\boldsymbol{x}_r)$.

Операція відбиття ілюструється на рис. 3.5. Якщо $\alpha > 0$ – це коефіцієнт відбиття, то положення $\boldsymbol{x}_r$ визначається таким способом:

$$\boldsymbol{x}_r - \boldsymbol{x}_0 = \alpha(\boldsymbol{x}_0 - \boldsymbol{x}_h),$$

тобто

$$\boldsymbol{x}_r = (1+\alpha)\,\boldsymbol{x}_0 - \alpha\,\boldsymbol{x}_h. \tag{3.4}$$

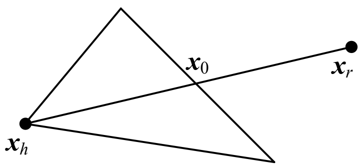


Рисунок 3.5 – Операція відбиття

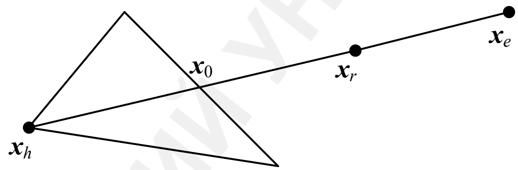


Рисунок 3.6 – Операція розтягування

Д Порівняємо значення функцій f_r і f_l.

1 Якщо $f_r < f_l$, то ми одержали найменше значення функції. Напрямок із точки $\boldsymbol{x}_0$ у точку $\boldsymbol{x}_r$ найбільш зручний для переміщення. Таким чином, ми робимо розтягування у цьому напрямку і знаходимо точку $\boldsymbol{x}_e$, і значення функції $f_e = f(\boldsymbol{x}_e)$. Рис. 3.5 ілюструє операцію розтягування симплекса. Коефіцієнт розтягування $\gamma > 1$ можна знайти з таких співвідношень:

$$\boldsymbol{x}_e - \boldsymbol{x}_0 = \gamma(\boldsymbol{x}_r - \boldsymbol{x}_0),$$

тобто

$$\boldsymbol{x}_e = \gamma\,\boldsymbol{x}_r + (1-\gamma)\,\boldsymbol{x}_0. \tag{3.5}$$

Зауваження $\gamma = |\boldsymbol{x}_e - \boldsymbol{x}_0| \,/\, |\boldsymbol{x}_r - \boldsymbol{x}_0|$.

а) якщо $f_e < f_l$, то замінюємо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_e$ (відповідно f_h на f_e) і перевіряємо точки симплекса на збіжність до мінімуму (див. крок І). Якщо збіжність досягнута, то процес зупиняється; у протилежному разі повертаємося на крок Б (обчислюємо нові значання f_h, f_l, f_g і відповідні їм точки $\boldsymbol{x}_h$, $\boldsymbol{x}_l$, $\boldsymbol{x}_g$).

б) якщо $f_e \geq f_l$, то відкидаємо точку $\boldsymbol{x}_e$. Очевидно, ми перемістилися занадто далеко від точки $\boldsymbol{x}_0$ до точки $\boldsymbol{x}_r$. Тому варто замінити точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_r$, у якій було отримане поліпшення (крок Д.1), перевірити збіжність і, якщо вона не досягнута, повернутися на крок Б (визначити нові значення f_h, f_l, f_g і відповідні їм значення $\boldsymbol{x}_h$, $\boldsymbol{x}_l$, $\boldsymbol{x}_g$).

2 Якщо $f_r > f_l$, але $f_r \le f_g$, то $\boldsymbol{x}_r$ є кращою точкою порівняно з іншими двома точками симплекса і ми замінюємо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_r$ і, якщо збіжність не досягнута, повертаємося на крок Б, тобто виконуємо описаний вище пункт 1 Б.

3 Якщо $f_r > f_l$ та $f_r > f_g$, то перейдемо на крок Е.

Е Порівняємо значення функцій f_r і f_h.

1 Якщо $f_r > f_h$, то переходимо безпосередньо до кроку стискання Е.2.

Якщо $f_r < f_h$, то замінюємо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_r$ і значення функції f_h на значення функції f_r. Запам'ятовуємо значення $f_r > f_g$ із кроку Д.2, наведеного вище. Потім переходимо на крок Е.2.

2 У цьому разі $f_r > f_h$, зрозуміло, що ми перемістилися надто далеко від точки $\boldsymbol{x}_h$ до точки $\boldsymbol{x}_0$. Намагаємося виправити це, знайшовши точку $\boldsymbol{x}_c$ (а потім і f_c) за допомогою кроку стискання, показаного на рис. 3.7.

Якщо $f_r > f_h$, то відразу переходимо до кроку стискання і знаходимо точку $\boldsymbol{x}_c$ із співвідношення

$$\boldsymbol{x}_c - \boldsymbol{x}_0 = \beta\,(\boldsymbol{x}_h - \boldsymbol{x}_0),$$

де β $(0 < \beta < 1)$ – коефіцієнт стискання. Тоді

$$\boldsymbol{x}_c = \beta\,\boldsymbol{x}_h + (1 - \beta)\,\boldsymbol{x}_0. \tag{3.6}$$

Якщо $f_r < f_h$, то спочатку замінимо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_r$, а потім проводимо стискання (рис. 3.8). Тоді точку $\boldsymbol{x}_c$ знайдемо із співвідношення:

$$\boldsymbol{x}_c - \boldsymbol{x}_0 = \beta\,(\boldsymbol{x}_r - \boldsymbol{x}_0),$$

тобто

$$\boldsymbol{x}_c = \beta\,\boldsymbol{x}_r + (1 - \beta)\,\boldsymbol{x}_0. \tag{3.7}$$

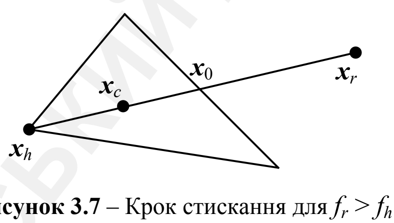


Рисунок 3.7 – Крок стискання для $f_r > f_h$

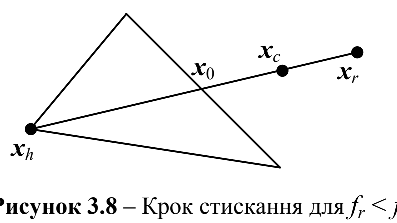


Рисунок 3.8 – Крок стискання для $f_r < f_h$

Ж Порівнюємо значення функцій f_c і f_h.

1 Якщо $f_c < f_h$, то замінюємо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_c$, і якщо збіжність не досягнута, то повертаємося на крок Б (виконуємо перевизначення значень f_h, f_l, f_g і відповідних їм значень $\boldsymbol{x}_h$, $\boldsymbol{x}_l$, $\boldsymbol{x}_g$).

2 Якщо $f_c > f_h$, то очевидно, що всі наші спроби знайти значення менше f_h закінчилися поразкою, тому ми переходимо на крок З.

З На цьому кроці ми зменшуємо розмірність симплекса поділом навпіл відстані від кожної точки симплекса до $\boldsymbol{x}_l$ – точки, в якій отримано найменше значення функції.

Таким чином, точка $\boldsymbol{x}_i$ замінюється на точку $\boldsymbol{x}_i + (\boldsymbol{x}_l - \boldsymbol{x}_i) / 2$, тобто заміняємо точку $\boldsymbol{x}_i$ точкою

$$\boldsymbol{x}_i = (\boldsymbol{x}_i + \boldsymbol{x}_l) / 2. \tag{3.8}$$

Потім обчислюємо f_i для $i = 1, 2, ..., n + 1$, перевіряємо збіжність і, якщо вона не досягнута, повертаємося на крок Б.

І Перевірка збіжності базується на тому, щоб стандартне відхилення σ $(n + 1)$-го значення функції було менше деякого заданого малого значення ε. У цьому випадку обчислюється

$$\sigma^2 = \sum_{i=1}^{n+1} (f_i - \bar{f})^2 \Big/ (n+1)\ , \tag{3.9}$$

де $\bar{f} = \sum\limits_{i=1}^{n} f_i \Big/ (n+1)$.

Якщо $\sigma < \varepsilon$, то всі значення функції дуже близькі одне до одного, і тому вони, можливо, знаходяться поблизу точки мінімуму функції x^*. Виходячи з цього, такий критерій збіжності є розумним.

Коефіцієнти α, β, γ у вищенаведеному алгоритмі є відповідно коефіцієнтами відбиття, стискання і розтягування. Нелдер і Мід рекомендують брати $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$. Рекомендація заснована на результатах експериментів з різними комбінаціями значень. Ці значення параметрів дозволяють методу бути ефективним.

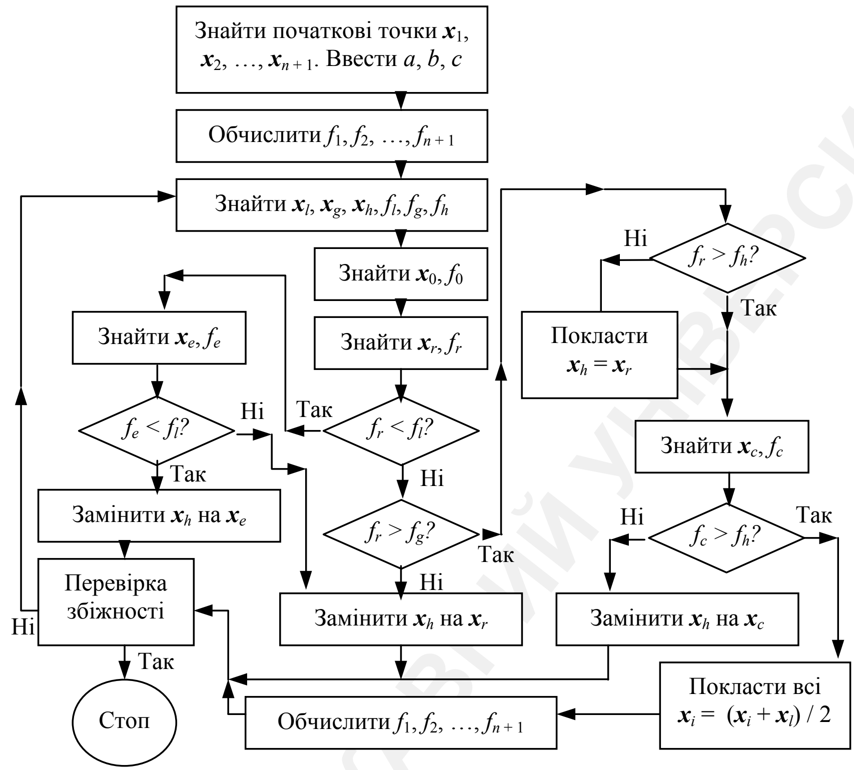


Рисунок 3.9 – Блок-схема методу Нелдера – Міда

Приклад Знайти мінімум функції $f(x) = (1 - x_1)^2 + (2 - x_2)^2$ з точністю до $\varepsilon = 0.01$. Використати як початкову точку (0, 0). Взяти початковий крок $h = 2$. Використати значення $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$.

А Побудуємо початковий симплекс.

$$\delta_1 = \left[\frac{\sqrt{2+1}+2-1}{2\sqrt{2}}\right] 2 = 1.9318\,,$$

$$\delta_2 = \left[\frac{\sqrt{2+1}-1}{2\sqrt{2}}\right] 2 = 0.5176\,.$$

Використовуючи ці два параметри, отримаємо дві інші точки початкового симплексу.

$$\boldsymbol{x}^{(1)} = (0 + 0.5176;\ 0 + 1.9318) = (0.5176;\ 1.9318);$$
$$\boldsymbol{x}^{(2)} = (0 + 1.9318;\ 0 + 0.5176) = (1.9318;\ 0.5176).$$

Ітерація 1

Б Обчислюємо значення функції в точках початкового симплексу.

$$f(\boldsymbol{x}^{(0)}) = 5; f(\boldsymbol{x}^{(1)}) = 0.2374; f(\boldsymbol{x}^{(2)}) = 3.0658.$$

Отже, $\boldsymbol{x}^{(0)} = \boldsymbol{x}_h$ – найбільше значення функції, $\boldsymbol{x}^{(1)} = \boldsymbol{x}_g$ – наступне за найбільшим значення функції, $\boldsymbol{x}^{(2)} = \boldsymbol{x}_l$ – найменше значення функції.

В Знайдемо за формулою (3.3) центр тяжіння $\boldsymbol{x}_0$ всіх точок, за винятком точки $\boldsymbol{x}_h$

$$\boldsymbol{x}_0 = \frac{1}{n}\sum_{i \neq h} \boldsymbol{x}_i = 1/2\ (\boldsymbol{x}_g + \boldsymbol{x}_l) = ((0.5176 + 1.9318) / 2; (1.9318 + 0.5176) / 2) = (1.2247; 1.2247).$$

Обчислимо $f(\boldsymbol{x}_0) = f_0 = (1 - 1.2247)^2 + (1 - 1.2247)^2 = 0.6516$.

Г Зручніше за все почати переміщення від точки $\boldsymbol{x}_h$. Відбивши точку $\boldsymbol{x}_h$ відносно точки $\boldsymbol{x}_0$, одержимо нову точку $\boldsymbol{x}_r$ і знайдемо $f_r = f(\boldsymbol{x}_r)$ (формула 3.4).

$$\boldsymbol{x}_r = (1 + \alpha)\ \boldsymbol{x}_0 - \alpha\ \boldsymbol{x}_h = (1 + 1)\ \boldsymbol{x}_0 - \boldsymbol{x}_h = (2 * 1.2247 - 0; 2 * 1.2247 - 0) = (2.4494; 2.4494).$$

Обчислимо $f(\boldsymbol{x}_r) = f_r = (1 - 2.4494)^2 + (1 - 2.4494)^2 = 2.3027$.

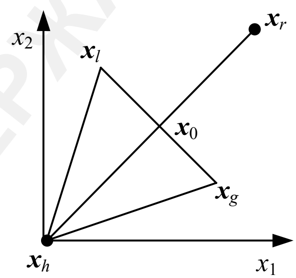


Рисунок 3.10 – Ітерація 1, операція розтягування

Маємо $f_r > f_l$, проте $f_r < f_g$, тобто $\boldsymbol{x}_r$ є кращою точкою порівняно з двома іншими точками, і ми замінюємо точку $\boldsymbol{x}_h$ на точку $\boldsymbol{x}_r$.

І Перевіряємо збіжність за формулою (3.9)

$$\bar{f} = \sum_{i=1}^{n} f_i / (n+1) = (0.2374 + 3.0658 + 2.3027) / (2 + 1) = 1.8686;$$

$$\sigma^2 = \sum_{i=1}^{n+1} (f_i - \bar{f})^2 \Big/ (n+1) = ((0.2374 - 1.8686)^2 + (3.0658 - 1.8686)^2 +$$
$$+ (2.3027 - 1.8686)^2) / (2 + 1) = 1.4275; \quad \sigma = 1.1948 > \varepsilon.$$

Оскільки збіжність не досягнуто, починаємо другу ітерацію і переходимо на крок Б.

Для стислості наведемо на другій і наступних ітераціях лише основні результати.

Ітерація 2

$\boldsymbol{x}_l$ (0.5176; 1.9318); $f_l = 0.2374$;

$\boldsymbol{x}_g$ (2.4494; 2.4494); $f_g = 2.3027$;

$\boldsymbol{x}_h$ (1.9318; 0.5176); $f_h = 3.0658$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (1.0352; 3.8636); $f_r = 3.4742$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок Е.

Виконавши операцію стискання, отримуємо точку $\boldsymbol{x}_c$ (1.7077; 1.3541); $f_c = 0.9180$.

Перевірка збіжності $\sigma = 0.859 > \varepsilon$, збіжність не досягнуто.

Ітерація 3

$\boldsymbol{x}_l$ (0.5176; 1.9318); $f_l = 0.2374$;

$\boldsymbol{x}_g$ (1.7077; 1.3541); $f_g = 0.9180$;

$\boldsymbol{x}_h$ (2.4494; 2.4494); $f_h = 2.3027$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (–0.224; 0.8366); $f_r = 2.8517$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок Е.

Виконавши операцію стискання, отримуємо точку $\boldsymbol{x}_c$ (1.7810; 2.0462); $f_c = 0.6121$.

Перевірка збіжності $\sigma = 0.278 > \varepsilon$, збіжність не досягнуто.

Ітерація 4

$\boldsymbol{x}_l$ (0.5176; 1.9318); $f_l = 0.2374$;

$\boldsymbol{x}_g$ (1.7810; 2.0462); $f_g = 0.6121$;

$\boldsymbol{x}_h$ (1.7077; 1.3541); $f_h = 0.9180$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (0.5909; 2.6239); $f_r = 0.5566$. Маємо $f_r > f_l$, $f_r < f_g$, тобто точка $\boldsymbol{x}_r$ є кращою порівняно з двома іншими точками симплекса, і ми замінюємо точку $\boldsymbol{x}_h$ на $\boldsymbol{x}_r$.

Перевірка збіжності $\sigma = 0.278 > \varepsilon$, збіжність не досягнуто.

Ітерація 5

$\boldsymbol{x}_l$ (0.5176; 1.9318); $f_l = 0.2374$;

$\boldsymbol{x}_g$ (0.5909; 2.6239); $f_g = 0.5566$;

$\boldsymbol{x}_h$ (1.7810; 2.0462); $f_h = 0.6121$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (–0.6725; 2.5095); $f_r = 3.0569$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок Е.

Виконавши операцію стискання, отримуємо точку $\boldsymbol{x}_c$ (1.1676; 2.1620); $f_c = 0.0543$.

Перевірка збіжності $\sigma = 0.2075 > \varepsilon$, збіжність не досягнуто.

Ітерація 6

$\boldsymbol{x}_l$ (1.1676; 2.1620); $f_l = 0.0543$;

$\boldsymbol{x}_g$ (0.5176; 1.9318); $f_g = 0.2374$;

$\boldsymbol{x}_h$ (0.5909; 2.6239); $f_h = 0.5566$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (1.0943; 1.4699); $f_r = 0.29$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок Е.

Виконавши операцію стискання, отримуємо точку $\boldsymbol{x}_c$ (0.9685; 1.7584); $f_c = 0.0594$.

Перевірка збіжності $\sigma = 0.085 > \varepsilon$, збіжність не досягнуто.

Ітерація 7

$\boldsymbol{x}_l$ (1.1676; 2.1620); $f_l = 0.0543$;

$\boldsymbol{x}_g$ (0.9685; 1.7584); $f_g = 0.0594$;

$\boldsymbol{x}_h$ (0.5176; 1.9318); $f_h = 0.2374$.

Виконавши операцію відбиття, отримуємо точку $\boldsymbol{x}_r$ (1.6185; 1.9886); $f_r = 0.3827$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок Е.

Виконавши операцію стискання, отримуємо точку $\boldsymbol{x}_c$ (0.7928; 1.946); $f_c = 0.0458$.

Перевірка збіжності $\sigma = 0.0097 < \varepsilon$, збіжність досягнуто.

Отже, мінімум функції (з точністю до заданого значення ε) має місце в точці (0.7928; 1.946) та складає 0.0458.

Завдання до підрозділу 3.1.4 Виконати пошук мінімуму або максимуму функції 2 додатка за методом Нелдера – Міда.

3.2 градієнтні методи. метод найшвидшого спуску

За допомогою розглянутого у попередньому підрозділі методу покоординатного спуску здійснюється пошук із заданої точки у напрямку, паралельному одній із осей, до точки мінімуму у даному напрямку. Далі пошук проводиться у напрямку, паралельному другій осі і т.п. Напрямки, звичайно, фіксовані. Здається доцільним модифікувати цей метод таким чином, щоб на кожному етапі пошук точки мінімуму здійснювався вдовж "найкращого" напрямку. Не зрозуміло, який напрямок є "найкращим", але відомо, що напрямок градієнта $\operatorname{grad} f(\boldsymbol{x})$ є напрямком найшвидшого зростання функції f ($\boldsymbol{x}$). Отже, протилежний напрямок є напрямком найшвидшого убування функції.

Метод оптимізації, в якому довжина кроку λ вибирається з умови мінімізації функції вдовж напрямку антиградієнта, отримав назву методу найшвидшого спуску. Цей варіант градієнтного методу заснований на використанні лінійної частини розкладення функції, що мінімізується, в околі точки в ряд Тейлора.

Алгоритм методу найшвидшого спуску описано нижче.

А Пошук мінімуму функції $f(x_1, x_2, \ldots, x_n)$ починається із заданої точки $\boldsymbol{x}_0$ з координатами $(x_1, x_2, \ldots, x_n)_0$. Визначається початкове значення функції $f(\boldsymbol{x}_0)$ (вважаючи $\lambda = 0$).

Б Визначити часткові похідні функції (проекції градієнта на координатні напрямки) $\partial f / \partial x_1, \partial f / \partial x_2, \ldots, \partial f / \partial x_n$.

Прийняти за напрямок пошуку напрямок

$$\boldsymbol{d} = -\operatorname{grad} f(\boldsymbol{x}) = -\left[\frac{\partial f}{\partial x_1}\bar{i} + \frac{\partial f}{\partial x_2}\bar{j} + \ldots + \frac{\partial f}{\partial x_n}\bar{n}\right].$$

Обчислити $\boldsymbol{d}$ у початковій точці $(x_1, x_2, \ldots, x_n)_0$, тобто отримати $\boldsymbol{d}_0$.

В Знайти значення λ_i, що мінімізує функцію

$$\varphi(\lambda_i) = f(\boldsymbol{x}_i + \lambda_i \boldsymbol{d}_i),$$

де i – номер ітерації.

Для пошуку мінімуму функції $\varphi(\lambda_i)$ в напрямі $\boldsymbol{d}_i$ з точки $\boldsymbol{x}_i$ можна використовувати методи одновимірного пошуку. Як показує досвід, гарні результати дає застосування методу квадратичної інтерполяції. Довжина кроку λ_i вибирається таким чином, щоб крок “перекрив” мінімум функції $\varphi(\lambda)$. Умова “перекриття” мінімуму виконується, якщо $\varphi(\lambda) \geq \varphi(0)$.

Якщо мінімум не потрапив у відрізок $(0, \lambda)$, то λ подвоюється, і це повторюється стільки разів, скільки необхідно для виконання умови “перекриття”.

Переконавшись, що відрізок $(0, \lambda)$ містить мінімум, як третю точку беруть точку $\lambda / 2$. Мінімальну точку згладжувального квадратичного полінома знаходять за формулою (2.13). Не обов’язково проводити одновимірний пошук з великою точністю, як правило, достатньо досягти убування функції $\varphi(\lambda)$. Для цього необхідно виконати 2 – 3 квадратичних ітерації.

Г Присвоїти $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i$, і якщо $\left|\operatorname{grad} f(\boldsymbol{x}_{i+1})\right|$ є достатньо малим, то процес закінчується. У протилежному разі слід повернутися на крок Б. Враховуючи, що в процесі пошуку відбувається наближення до екстремуму, для підвищення ефективності процедури доцільно після кожної ітерації зменшувати довжину кроку λ.

Слід відзначити, що метод найшвидшого спуску не рекомендується у якості серйозної оптимізаційної процедури. На перший погляд він є привабливим, проте для практичного застосування “працює” надто повільно. Справа в тому, що властивість найшвидшого спуску є лише локальною властивістю, і тому у ряді випадків доводиться часто змінювати напрям пошуку, що й призводить у підсумку до неефективної обчислювальної процедури. Метод найшвидшого спуску не використовує другі похідні цільової функції. В роботах [2, 3] описані також більш досконалі градієнтні методи, зокрема, метод Ньютона – Рафсона, Давидона – Флетчера – Пауелла та Флетчера – Рівса, але через складність алгоритмів цих методів їх розгляд не внесено до даного курсу.

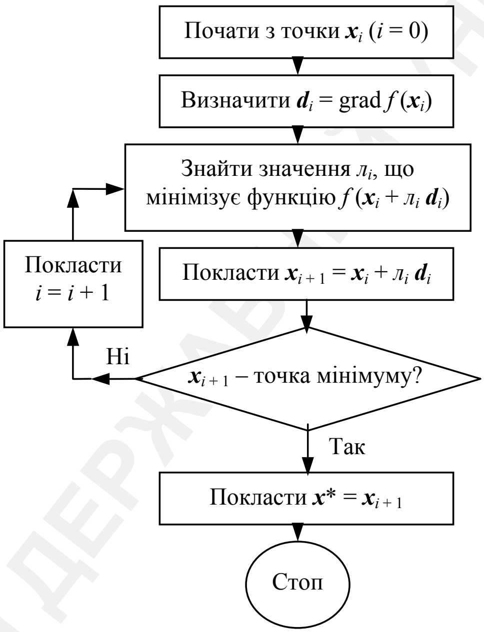


Рисунок 3.11 – Блок-схема методу найшвидшого спуску

Приклад Знайти мінімум функції $f(x) = (x_1 - 1)^2 + (x_2 - 3)^2 + 4\,(x_3 + 5)^2$ з точністю до $\varepsilon = 0.1$. Використати як початкову точку (4, –1, 2). Взяти початковий крок $\lambda = 4$.

Ітерація 1

А Обчислимо значення функції в початковій точці $\boldsymbol{x}_0$

$$f(\boldsymbol{x}_0) = (4 - 1)^2 + ((-1) - 3)^2 + 4\,(2 + 5)^2 = 221.$$

Б Похідні функції знайдемо аналітично.

$$\partial f / \partial x_1 = 2x_1 - 2; \quad \partial f / \partial x_2 = 2x_2 - 6; \quad \partial f / \partial x_3 = 8x_3 + 40.$$

Отримаємо вираз для градієнта функції.

$$\boldsymbol{d} = -\operatorname{grad} f(\boldsymbol{x}) = (-2x_1 + 2)\,\boldsymbol{i} + (-2x_2 + 6)\,\boldsymbol{j} + (-8x_3 - 40)\,\boldsymbol{k}.$$

У початковій точці $\boldsymbol{d}_0 = (-2 * 4 + 2)\,\boldsymbol{i} + (-2 * (-1) + 6)\,\boldsymbol{j} + (-8 * 2 - 40)\,\boldsymbol{k} =$
$= -6\,\boldsymbol{i} + 8\,\boldsymbol{j} - 56\,\boldsymbol{k}.$

$$|\boldsymbol{d}_0| = |-\operatorname{grad} f(\boldsymbol{x}_0)| = \sqrt{6^2 + 8^2 + 56^2} = 56.89.$$

В. Отримаємо вираз для $\varphi(\lambda) = f(\boldsymbol{x} + \lambda\,\boldsymbol{d})$.

$$\varphi(\lambda) = (x_1 + \lambda(-2x_1 + 2) - 1)^2 + (x_2 + \lambda(-2x_2 + 6) - 3)^2 + $$
$$+ 4(x_3 + \lambda(-8x_3 - 40) + 5)^2.$$

У точці $\boldsymbol{x}_0$ маємо

$$\varphi(\lambda) = (4 + \lambda(-2 * 4 + 2) - 1)^2 + (-1 + \lambda((-2) * (-1) + 6) - 3)^2 + 4(2 + \lambda((-$$
$$-8) * 2 - 40) + 5)^2 = (3 - 6\lambda)^2 + (-4 + 8\lambda)^2 + 4(7 - 56\lambda)^2.$$

Знаходимо значення $\varphi(\lambda)$ при $\lambda = 0$, $\lambda / 2$ та λ (де $\lambda = 4$).

$$\varphi(4) = (3 - 6 * 4)^2 + (-4 + 8 * 4)^2 + 4(7 - 56 * 4)^2 = 48314;$$
$$\varphi(2) = (3 - 6 * 2)^2 + (-4 + 8 * 2)^2 + 4(7 - 56 * 2)^2 = 11250;$$
$$\varphi(0) = (3 - 6 * 0)^2 + (-4 + 8 * 0)^2 + 4(7 - 56 * 0)^2 = 221.$$

Знайдемо мінімум функції $\varphi(\lambda)$, скориставшись методом квадратичної інтерполяції.

Внутрішня ітерація 1

Вважаючи $a = 0$, $b = 2$, $c = 4$, $f_a = 221$, $f_b = 11250$, $f_c = 48314$, скористаємось формулою (2.13).

$$\delta = \frac{1}{2}\left[\frac{(2^2 - 4^2)\cdot 221 + (4^2 - 0^2)\cdot 11250 + (0^2 - 2^2)\cdot 48314}{(2-4)\cdot 221 + (4-0)\cdot 11250 + (0-2)\cdot 48314}\right] = 0.1528.$$

Відповідне значення функції $\varphi(0.1528) = 21.72$.

Відкидаючи точку $\lambda = 4$, переходимо до наступної ітерації.

Внутрішня ітерація 2

Вважаючи $a = 0$, $b = 0.1528$, $c = 2$, $f_a = 221$, $f_b = 21.72$, $f_c = 11250$, аналогічно знаходимо нове значення $\delta = 0.2531$ та відповідне значення $\varphi\,(0.2531) = 211.87$. Відкидаючи точку $\lambda = 2$, переходимо до наступної ітерації.

Внутрішня ітерація 3

Вважаючи $a = 0$, $b = 0.1528$, $c = 0.2531$, $f_a = 221$, $f_b = 21.72$, $f_c = 211.87$, аналогічно знаходимо нове значення $\delta = 0.1280$ та відповідне значення $\varphi\,(0.1280) = 13.95$. На цьому закінчуємо метод квадратичної інтерполяції, оскільки 3 ітерацій звичайно достатньо. Встановлюємо $\lambda_{min} = 0.1280$

Г Маємо координати нової точки $\boldsymbol{x}_1\,(x_1, x_2, x_3)$:

$x_1 = x_1 + \lambda_{min}\,(-2x_1 + 2) = 4 + 0.128 * (-2 * 4 + 2) = 3.232;$

$x_2 = x_2 + \lambda_{min}\,(-2x_2 + 6) = -1 + 0.128 * (-2 * (-1) + 6) = 0.024;$

$x_3 = x_3 + \lambda_{min}\,(-8x_3 - 40) = 2 + 0.128 * (-8 * 2 - 40) = -5.166.$

Обчислимо значення функції в новій точці $\boldsymbol{x}_1$ та значення градієнта в цій точці

$$f\,(\boldsymbol{x}_1) = (3.232 - 1)^2 + (0.024 - 3)^2 + 4\,(-5.166 + 5)^2 = 13.96.$$

$$|\boldsymbol{d}_1| = |-\operatorname{grad} f\,(\boldsymbol{x}_1)| =$$
$$\sqrt{(-2\cdot 3.232 + 2)^2 + (-2\cdot 0.024 + 6)^2 + (-8\cdot(-5.166) - 40)^2} = 7.56.$$

Відмітимо, що значення і $f\,(\boldsymbol{x}_1)$, і $|\boldsymbol{d}_1|$ стали значно менші, ніж вони були в початковій точці. Все ж вони ще є досить великими і пошук необхідно продовжувати. Для наступних ітерацій наведемо лише основні проміжні результати.

Ітерація 2

Зменшуємо початковий крок удвічі, встановлюємо $\lambda = 2$.

Після застосування методу квадратичної інтерполяції (одна внутрішня ітерація) отримуємо значення $\lambda_{min} = 0.4576$ та відповідне $\varphi\,(\lambda_{min}) = 0.88$.

Маємо координати нової точки $\boldsymbol{x}_2$ (1.189, 2.747, –4.558).

Отримуємо значення функції в новій точці $\boldsymbol{x}_2$ та значення градієнта в цій точці:

$$f\,(\boldsymbol{x}_2) = (1.189 - 1)^2 + (2.747 - 3)^2 + 4\,(-4.558 + 5)^2 = 0.88.$$

$$|\boldsymbol{d}_2| = |-\operatorname{grad} f\,(\boldsymbol{x}_2)| =$$
$$= \sqrt{(-2\cdot 1.189 + 2)^2 + (-2\cdot 2.747 + 6)^2 + (-8\cdot(-4.558) - 40)^2} = 3.59.$$

Ітерація 3

Зменшуємо початковий крок удвічі, встановлюємо $\lambda = 1$.

Після застосування методу квадратичної інтерполяції (1 внутрішня ітерація) отримуємо значення $\lambda_{min} = 0.1280$ та відповідне $\varphi\,(\lambda_{min}) = 0.056$.

Маємо координати нової точки $\boldsymbol{x}_3$ (1.141, 2.812, –5.011).

Отримуємо значення функції в новій точці $\boldsymbol{x}_3$ та значення градієнта в цій точці:

$$f\,(\boldsymbol{x}_2) = (1.141 - 1)^2 + (2.812 - 3)^2 + 4\,(-5.011 + 5)^2 = 0.06.$$

$$|\boldsymbol{d}_2| = |-\operatorname{grad} f\,(\boldsymbol{x}_2)| =$$
$$\sqrt{(-2 \cdot 1.141 + 2)^2 + (-2 \cdot 2.812 + 6)^2 + (-8 \cdot (-5.011) - 40)^2} = 0.48.$$

І так далі. Відмітимо, що з кожною ітерацією значення як функції, так і її градієнта стрімко зменшується.

Вже після трьох ітерацій маємо точку (1.141, 2.812, –5.011), значення функції в якій складає 0.06, що майже збігається зі справжнім мінімумом функції. Справжній мінімум даної функції $f\,(\boldsymbol{x})$ має місце в точці (1, 3, –5), у чому можна переконатися, прирівнявши до нуля перші похідні функції $(\boldsymbol{x})$. Значення функції $(\boldsymbol{x})$ у цій точці дорівнює нулю.

Завдання до підрозділу 3.2 Виконати пошук мінімуму або максимуму функції 2 додатка за методом найшвидшого спуску.

РОЗДІЛ 4 ОПТИМІЗАЦІЯ ПРИ НАЯВНОСТІ ОБМЕЖЕНЬ

4.1 загальна теорія

4.1.1 Обмеження у вигляді рівностей

В багатьох задачах на пошук найбільших та найменших значень функції, питання зводиться до пошуку екстремумів функції від декількох змінних, котрі не є незалежними, а пов'язані одна з одною деякими додатковими умовами (наприклад, рівнянням). Такий екстремум називається умовним.

За допомогою методу множників Лагранжа, по суті, встановлюються необхідні умови, що дозволяють ідентифікувати точки оптимуму в задачах оптимізації з обмеженнями у вигляді рівностей. При цьому задача з обмеженнями перетворюється в еквівалентну задачу безумовної оптимізації, в якій фігурують певні невідомі параметри, що називаються *множниками Лагранжа*.

Розглянемо задачу мінімізації функції двох змінних [2]

$$z = f(x, y),$$

де на x та y накладене обмеження, що задане рівнянням

$$g(x, y) = 0. \qquad (4.1)$$

Взагалі, рівняння $g(x, y) = 0$ можна розв'язати відносно y як функцію від x, тобто $y = h(x)$. Звичайно, на практиці може виявитись складним або навіть неможливим знайти явний вигляд функції $h(x)$. У разі виконання певних умов диференційованості похідна функції $h(x)$

$$\frac{dy}{dx} = \frac{d}{dx}h(x) = -\frac{\partial g}{\partial x}\bigg/\frac{\partial g}{\partial y}. \qquad (4.2)$$

Тоді функцію

$$z = f(x, h(x)) \qquad (4.3)$$

можна записати як функцію однієї незалежної змінної x. Необхідною умовою мінімуму функції z буде співвідношення

$$\frac{dz}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0,$$

тобто

$$\frac{\partial f}{\partial x}+\left(\frac{-\dfrac{\partial f}{\partial y}}{\dfrac{\partial g}{\partial y}}\right)\frac{\partial g}{\partial x}=0\,. \tag{4.4}$$

Співвідношення (4.1) та (4.2) можуть бути розв'язані для отримання значень x^*, y^* в точці мінімуму.

Цей результат може бути поданий в такій формі. Якщо покласти

$$\lambda=\frac{-\partial f}{\partial y}(x,y)\Big/\frac{\partial g}{\partial y}(x,y) \tag{4.5}$$

при $x = x^*$, $y = y^*$, то в точці мінімуму виконуються співвідношення

$$g\,(x,y)=0,$$
$$\frac{\partial f}{\partial x}(x,y)+\lambda\frac{\partial g}{\partial x}(x,y)=0\,,$$
$$\frac{\partial f}{\partial y}(x,y)+\lambda\frac{\partial g}{\partial y}(x,y)=0\,,$$

причому останнє витікає безпосередньо із співвідношення (4.5).

Отримати ці три необхідні умови можна, використовуючи *функцію Лагранжа*, записану у такому вигляді

$$F\,(x,y,\lambda)=f\,(x,y)+\lambda\,g\,(x,y), \tag{4.6}$$

що являє собою суму цільової функції та добутку множника Лагранжа λ на функцію обмеження. Тоді необхідні умови мінімуму функції $f\ (x,\ y)$ при наявності обмежень можуть бути записані у такому вигляді:

$$\frac{\partial F}{\partial x}(x,y,\lambda)=\frac{\partial f}{\partial x}(x,y)+\lambda\frac{\partial g}{\partial x}(x,y)=0\,,$$
$$\frac{\partial F}{\partial y}(x,y,\lambda)=\frac{\partial f}{\partial y}(x,y)+\lambda\frac{\partial g}{\partial y}(x,y)=0\,,$$
$$\frac{\partial F}{\partial \lambda}(x,y,\lambda)=g(x,y)=0\,. \tag{4.7}$$

Це система трьох рівнянь, розв'язком якої є значення x^*, y^* та λ^* – в точці мінімуму.

Необхідні умови мінімуму (4.7) можуть бути узагальнені для функції n змінних за наявності m обмежень у вигляді рівностей.

Розглянемо задачу мінімізації функції

$$z = f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n),$$

де на змінну $\boldsymbol{x}$ накладені обмеження

$$g_1(\boldsymbol{x}) = 0, \quad g_2(\boldsymbol{x}) = 0, \quad \ldots, g_m(\boldsymbol{x}) = 0. \qquad (4.8)$$

Обмеження можна використати для того, щоб виразити m змінних (без обмеження загальності їх можна позначити $x_1, x_2, \ldots, x_m$) через решту $(n - m)$ змінних, які можна розглядати як незалежні змінні. У точці мінімуму за наявності обмежень $f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) \geq 0$ для всіх $\boldsymbol{h}$, що задовольняють умову $g_i(\boldsymbol{x} + \boldsymbol{h}) = g_i(\boldsymbol{x}) = 0$ при $i = 1, 2, \ldots, m$.

Тоді з точністю до першого порядку h_j будемо мати

$$\sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j} = 0,$$

де

$$\sum_{j=1}^{n} h_j \frac{\partial g_i}{\partial x_j} = 0 \quad \text{при } i = 1, 2, \ldots, m.$$

Цю умову можна записати інакше:

$$\sum_{j=1}^{n} h_j \left(\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} \right) = 0, \qquad (4.9)$$

де $\lambda_1, \lambda_2, \ldots, \lambda_m$ – множники Лагранжа.

Звідси витікає

$$\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \quad \text{при } j = 1, 2, \ldots, n.$$

Отже, якщо визначити функцію Лагранжа у вигляді

$$F(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}), \tag{4.10}$$

то необхідні умови мінімуму функції $f(\boldsymbol{x})$ за наявності обмежень можна записати таким чином:

$$\frac{\partial F}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \quad \text{при } j = 1, 2, \ldots, n, \tag{4.11}$$

$$\frac{\partial F}{\partial \lambda_i} = g_i(\boldsymbol{x}) = 0 \quad \text{при } i = 1, 2, \ldots, m. \tag{4.12}$$

Розв’язання цієї розширеної системи, що складається з $n + m$ рівнянь, що містять $n + m$ невідомих, визначає стаціонарну точку функції $F(\boldsymbol{x}, \boldsymbol{\lambda})$. Далі реалізується процедура перевірки на мінімум або максимум, що проводиться на основі обчислення елементів матриці Гессе функції $F(\boldsymbol{x}, \boldsymbol{\lambda})$.

Приклад Знайти мінімум функції $f(x, y) = x^2 + y^2$ при обмеженні $x + y = 4$.
Функція Лагранжа набуде вигляду

$$F(x, y, \lambda) = x^2 + y^2 + \lambda(4 - x - y).$$

Відповідні умови мінімума можна записати таким чином:

$$\frac{\partial F}{\partial x} = 2x - \lambda = 0,$$

$$\frac{\partial F}{\partial y} = 2y - \lambda = 0,$$

$$\frac{\partial F}{\partial \lambda} = 4 - x - y = 0.$$

Розв’язанням цієї системи рівнянь є $x = y = 2$, $\lambda = 4$. Матриця Гессе функції F має вигляд $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ і отже, є додатно визначеною, а це доводить, що точка (2, 2) є точкою мінімуму.

4.1.2 Обмеження у вигляді нерівностей

У цьому підрозділі метод множників Лагранжа буде поширений на обмеження у вигляді нерівностей. Розглянемо загальну задачу математичного програмування: мінімізувати функцію $f(\boldsymbol{x})$ за наявності m обмежень $g_i(\boldsymbol{x}) \le b_i$ $(i = 1, 2, \ldots, m)$.

Обмеження у вигляді нерівностей можуть бути перетворені в обмеження у вигляді рівностей доданням до кожного з них невід'ємної послаблювальної змінної u_i^2

$$g_i(\boldsymbol{x}) + u_i^2 = b_i,$$

або

$$g_i(\boldsymbol{x}) + u_i^2 - b_i, = 0. \tag{4.13}$$

Таким чином, задача зводиться до мінімізації функції $f(\boldsymbol{x})$ за наявності m обмежень у вигляді рівності $g_i(\boldsymbol{x}) + u_i^2 - b_i, = 0$. Відповідно до викладеного у попередньому підрозділі методу, сформуємо функцію Лагранжа

$$F(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{u}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \, [g_i(\boldsymbol{x}) + u_i^2 - b_i]. \tag{4.14}$$

У стаціонарній точці мають виконуватися такі умови:

$$\frac{\partial F}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \quad \text{при } j = 1, 2, \ldots, n, \tag{4.15}$$

$$\frac{\partial F}{\partial \lambda_i} = g_i(\boldsymbol{x}) + u_i^2 - b_i = 0 \quad \text{при } i = 1, 2, \ldots, m, \tag{4.16}$$

$$\frac{\partial F}{\partial u_i} = 2\,\lambda_i\, u_i = 0 \quad \text{при } i = 1, 2, \ldots, m. \tag{4.17}$$

Помноживши останнє рівняння на $u_i / 2$, отримуємо $\lambda_i\, u_i^2 = 0$, тобто

$$\lambda_i\, [b_i - g_i(\boldsymbol{x})] = 0 \quad \text{при } i = 1, 2, \ldots, m. \tag{4.18}$$

Іще одна додаткова умова, яка має бути виконана в точці мінімуму за наявності обмежень: $\lambda_i \ge 0$.

Отже, необхідні умови мінімуму функції $f(\boldsymbol{x})$ за наявності обмежень $g_i(\boldsymbol{x}) \le b_i$ $(i = 1, 2, \ldots, m)$ мають вигляд

$$\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \quad \text{при } j = 1, 2, \ldots, n,$$

$$g_i(\boldsymbol{x}) \le b_i \quad \text{при } i = 1, 2, \ldots, m,$$

$$\lambda_i\,[b_i - g_i\,(\boldsymbol{x})] = 0 \quad \text{при } i = 1, 2, \ldots, m,$$

$$\lambda_i \geq 0 \quad \text{при } i = 1, 2, \ldots, m. \qquad (4.19)$$

Знак λ_i змінюється на протилежний, якщо розглядається максимум. Ці умови відомі як умови Куна – Такера.

Загальна задача математичного програмування, що сформульована на початку попереднього розділу, є дуже складною і досі не має повного розв'язання. Деякі труднощі виникають у задачах, що графічно проілюстровані на рис. 4.1. На рисунку зображені лінії постійного рівня функції. Можливо, що наявність обмежень буде призводити до появи локального мінімуму. Це може відбутись навіть у випадку, коли функція має лише одну точку мінімуму за відсутності обмежень. Така ситуація ілюструється рис. 4.2.

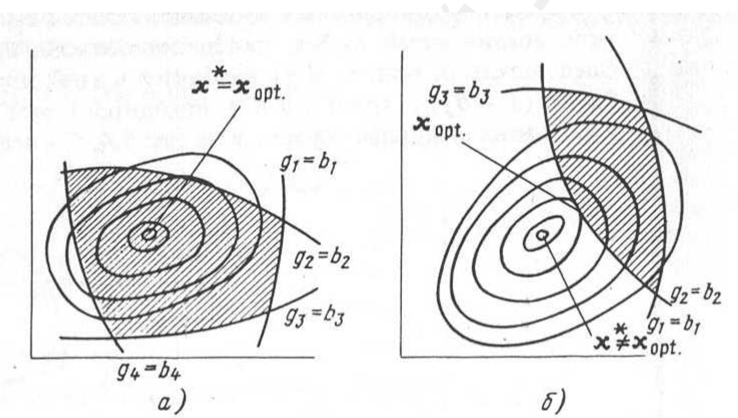


Рисунок 4.1 – Функція і області обмеження

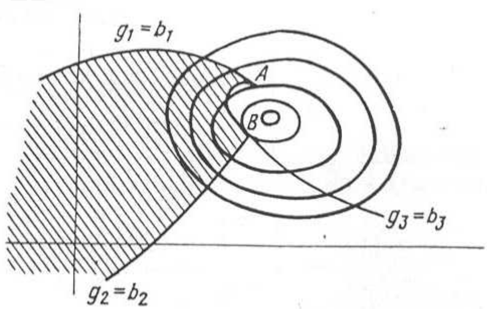


Рисунок 4.2 – Функція і області обмеження. Локальний мінімум в т. А

Приклад Записати умови Куна – Такера для мінімума функції $f(\boldsymbol{x}) = 3x_1^2 + 4x_1x_2 + 5x_2^2$ при обмеженнях $x_1 \geq 0$, $x_2 \geq 0$ та $x_1 + x_2 \geq 4$.

Цю задачу можна подати таким чином:
мінімізувати функцію $f(\boldsymbol{x}) = 3x_1^2 + 4x_1x_2 + 5x_2^2$
при обмеженнях $-x_1 \leq 0$, $-x_2 \leq 0$, $-x_1 - x_2 \leq 4$. Функція Лагранжа $F(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{u})$ матиме вигляд

$$F = 3x_1^2 + 4x_1x_2 + 5x_2^2 + \lambda_1(u_1^2 - x_1) + \lambda_2(u_2^2 - x_2) + \lambda_3(u_3^2 - x_1 - x_2 + 4).$$

Необхідними умовами мінімуму є

$$\begin{gathered} 6x_1 + 4x_2 - \lambda_1 - \lambda_3 = 0, \\ 4x_1 + 10x_2 - \lambda_2 - \lambda_3 = 0, \\ -x_1 \leq 0, -x_2 \leq 0, -x_1 - x_2 \leq 4, \\ \lambda_1 x_1 = 0, \lambda_2 x_2 = 0, \lambda_3(4 - x_1 - x_2) = 0, \\ \lambda_1, \lambda_2, \lambda_3 \geq 0. \end{gathered}$$

Нескладно перевірити, що ці умови дотримуються при $x_1 = 3$, $x_2 = 1$, $\lambda_1 = 0$, $\lambda_2 = 0$ та $\lambda_3 = 22$, і функція має мінімум, що дорівнює 44, в точці A з координатами (3, 1) (рис. 4.3).

Лініями постійного рівня функції $f(\boldsymbol{x})$ є еліпси $3x_1^2 + 4x_1x_2 + 5x_2^2 = c$.

Мінімум функції $f(\boldsymbol{x})$ за відсутності обмежень дорівнює нулю та знаходиться на початку координат. Область обмежень на рис. 4.3 заштрихована.

Рисунок 4.3 – Приклад функції і області обмеження

4.2 методи пошуку

4.2.1 Проста модифікація методів прямого пошуку

Методи прямого пошуку раніше були розглянуті для розв'язання задач оптимізації без обмежень. Ці методи нескладно модифікувати і для урахування обмежень. Було висунуто припущення, що для цього буде цілком достатньо при розв'язанні задачі мінімізації присвоїти цільовій функції дуже велике значення там, де ці обмеження порушуються. До того ж, таку ідею просто реалізувати за допомогою програмування.

Треба перевірити, чи кожна точка, що отримана в процесі пошуку, належить області визначення. Якщо кожна, то цільова функція обчислюється звичайним шляхом. Якщо ні, то цільовій функції привласнюється дуже велике значення. Таким чином, пошук буде здійснюватись знов у допустимій області в напрямку до мінімальної точки всередині цієї області.

На жаль, в загальному випадку такий підхід є недосконалим. За допомогою такого підходу неможливо просуватись вдовж границі області обмежень, оскільки збіжність досягається в першій точці границі. Загальна задача оптимізації при наявності обмежень є дуже складною, і для отримання практичного методу розв'язання потрібно використовувати більш складні процедури.

4.2.2 Послідовна оптимізація без обмежень. Штрафні функції

Основна ідея методу штрафної функції полягає в перетворенні задачі мінімізації функції $z = f(\boldsymbol{x})$ з відповідними обмеженнями, накладеними на $\boldsymbol{x}$, в задачу пошуку мінімуму *без обмежень* функції

$$Z = f(\boldsymbol{x}) + P(\boldsymbol{x}).$$

Функція $P(\boldsymbol{x})$ є штрафною. Необхідно, щоб при порушенні обмежень вона штрафувала функцію Z, тобто збільшувала її значення. У цьому разі мінімум Z буде знаходитись всередині області обмежень. Фунція $P(\boldsymbol{x})$, що задовольняє цю умову, може бути не єдиною.

Задачу мінімізації можна сформулювати таким чином: мінімізувати функцію $z = f(\boldsymbol{x})$ при обмеженнях $c_j(\boldsymbol{x}) > 0$, $j = 1, 2, \ldots, m$ [2].

Зауваження Обмеження вигляду "менше або дорівнює", $h(\boldsymbol{x}) \le 0$, завжди може бути записане як "більше або дорівнює", $-h(\boldsymbol{x}) \ge 0$, тому в наведеному формулюванні немає втрати загальності.

Функцію $P(\boldsymbol{x})$ зручно записати таким чином

$$P(\boldsymbol{x}) = r \sum_{j=1}^{m} (1 / c_j(\boldsymbol{x})), \tag{4.20}$$

де r – додатна величина. Тоді функція $Z = \varphi\,(\boldsymbol{x}, r)$ набуде вигляду

$$Z = \varphi\,(\boldsymbol{x}, r) = f\,(\boldsymbol{x}) + r \sum_{j=1}^{m} (1\,/\,c_j\,(\boldsymbol{x})). \tag{4.21}$$

Якщо $\boldsymbol{x}$ приймає допустимі значення, тобто значення, для яких $c_j\,(\boldsymbol{x}) \geq 0$, то Z приймає значення, більші відповідних значень $f\,(\boldsymbol{x})$ (справжньої цільової функції даної задачі), і різницю можна зменшити за рахунок того, що r може бути дуже малою величиною. Але якщо $\boldsymbol{x}$ приймає значення, що є хоча і допустимими, проте близькими до границі області обмеження, і щонайменше одна з функцій $c_j\,(\boldsymbol{x})$ близька до нуля, то значення функції $P\,(\boldsymbol{x})$ і, отже, значення функції Z стають дуже великими. Таким чином, вплив функції $P\,(\boldsymbol{x})$ полягає у створенні "гребню з крутими краями" вздовж кожної границі області обмежень. Отже, якщо пошук починається з допустимої точки і здійснюється пошук мінімуму функції $\varphi\,(\boldsymbol{x}, r)$ без обмежень, то мінімум, звичайно, буде досягатись всередині допустимої області для задачі з обмеженнями. Вважаючи r достатньо малою величиною для того, щоб вплив $P\,(\boldsymbol{x})$ був малим в точці мінімуму, ми можемо зробити точку мінімуму функції $\varphi\,(\boldsymbol{x}, r)$ без обмежень такою, що збігається з точкою мінімуму функції $f\,(\boldsymbol{x})$ з обмеженнями.

Приклад Використовуючи штрафну функцію, що задана рівнянням (4.21), мінімізувати функцію $f\,(x) = x$ при обмеженні $x \geq 2$, тобто $x - 2 \geq 0$. Мінімальним значенням функції є 2 при $x = 2$. Як за допомогою штрафної функції можна знайти рішення? Розглянемо функцію

$$\varphi\,(x, r) = x + \frac{r}{x-2}\,.$$

На рис. 4.4 зображений графік функції $\varphi\,(x, r)$ та показано положення точок її мінімуму для різних значень r (1; 0.25 та 0.01).

Область обмежень лежить справа від вертикальної прямої $x = 2$. Нескладно побачити, що послідовність точок Q_1, Q_2, Q_3 прагне до точки Q – мінімуму функції при наявності обмежень. Дійсно, знайдемо першу похідну функції φ (x, r)

$$\frac{d\varphi}{dx} = 1 - \frac{r}{(x-2)^2}\,.$$

Отже, якщо $d\varphi\,/\,dx = 0$, $(x-2)^2 = r$, то $x = 2 \pm \sqrt{r}$.

Тоді

$$\frac{d^2\varphi}{dx^2} = \frac{2r}{(x-2)^3},$$

і мінімум досягається при $x = 2 + \sqrt{r}$ всередині області обмежень.

Отже, функція φ $(x,\ r)$ має мінімум, що дорівнює $2 + 2\sqrt{r}$ при $x = 2 + \sqrt{r}$. Тоді Q_1 є точка з координатами (3, 4), Q_2 – точка з координатами (2.5, 3), Q_3 – точка з координатами (2.1, 2.2). Ясно, що при $r \to 0$ мінімум без обмежень функції φ $(x,\ r)$ наближається до значення 2 і мінімальною точкою є точка $x = 2$.

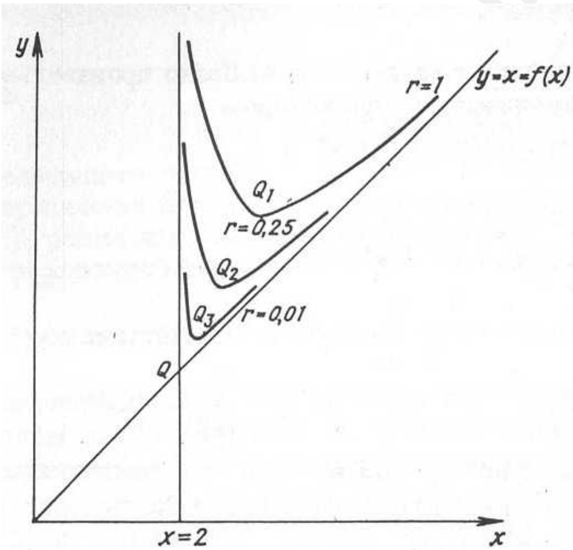


Рисунок 4.4 – Приклад застосування штрафної функції

РОЗДІЛ 5 ПРИКЛАДИ ЗАСТОСУВАННЯ МЕТОДІВ ОПТИМІЗАЦІЇ ДЛЯ РОЗВ'ЯЗАННЯ КОНКРЕТНИХ ТЕХНІЧНИХ ЗАДАЧ

Розглянемо два приклади, що взяті з монографії [1]. Оскільки ці приклади носять ілюстративний характер, з метою запобігання втрати наочності, в них збережено систему мір, що була використана автором.

5.1 оптимізація системи сонячного опалення для житлового будинку

У зв'язку з наростаючим дефіцитом дешевого палива сонячна енергія, запаси якої величезні, можливо знайде нарешті широкомасштабне практичне застосування в такому напрямку, як обігрів житла. Однак оскільки капітальні витрати, що пов'язані зі створенням систем сонячного опалення, які забезпечують потрібний обігрів житлових приміщень, вище, ніж витрати на традиційні опалювальні системи, що працюють на газі або мазуті, необхідним є ретельний аналіз розподілу витрат за окремими елементами цього нового типу опалювальних систем.

У наш час є понад 10 варіантів систем сонячного опалення житлового будинку. Одна з таких схем описана нижче. У поставленні розглянутої нижче задачі використано спрощений опис конструкції будинку, умов оточуючого середовища та процесу використання енергії сонячного випромінювання.

Характерна складність практичної реалізації таких систем полягає у тому, що енергію сонячного випромінювання можна отримувати лише в денний час, що може не відповідати бажаному графіку її споживання. Отже, слід створити системи приймання та акумулювання сонячної енергії, а також відповідних пристроїв перетворення та регулювання.

У запропонованому прикладі як приймач сонячної енергії використовується шаруватий колектор, який складається з майже абсолютно чорних листів металу, між якими прокачується вода. Після нагріву у колекторі вода, що є акумулятором тепла і теплоносієм одночасно, подається до бака великої ємності. З бака вода надходить до радіаторів з регульованою поверхнею. Там вона передає тепло приміщенням шляхом конвекції та випромінювання (рис. 5.1).

Характеристики колектора та акумулюючої системи повинні бути узгоджені між собою та відповідати потрібному тепловому навантаженню. Витрата тепла є функцією конструкції будинку, товщини ізоляції, площини вікон, довжини периметричних щілин у віконних та дверних прорізах і, звичайно, погодних умов.

У данному прикладі ми беремо деякі розумні розміри будинку, стандартну конструкцію (щитові стіни із заповнювачем – див. рис. 5.2) та встановимо "розрахункові погодні умови", залишаючи змінними товщину ізоляції стін і даху, площу поверхні колектора (вважаємо, що його ККД дорівнює 50%, що є достатньо близьким до реального значення), довжину ущільнюючих прокладок, які використовуються для запобігання витоків тепла з приміщення, та ємність бака. Продуктивність насоса, пропускну здатність трубопровода та параметри регулюючої системи будемо розглядати як функції інших параметрів.

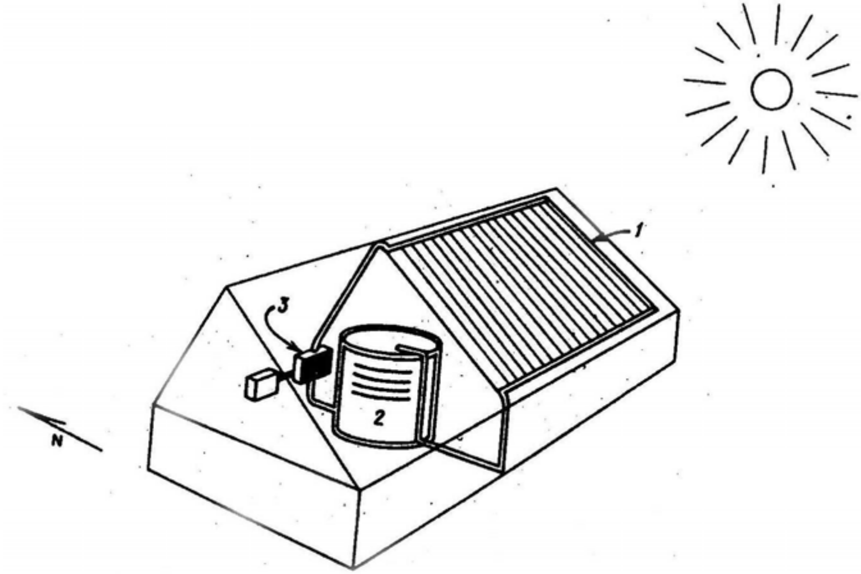


Рисунок 5.1 – Схема будинку з сонячним опаленням
1 – сонячний колектор; 2 – водяний бак; 3 – регульований радіатор

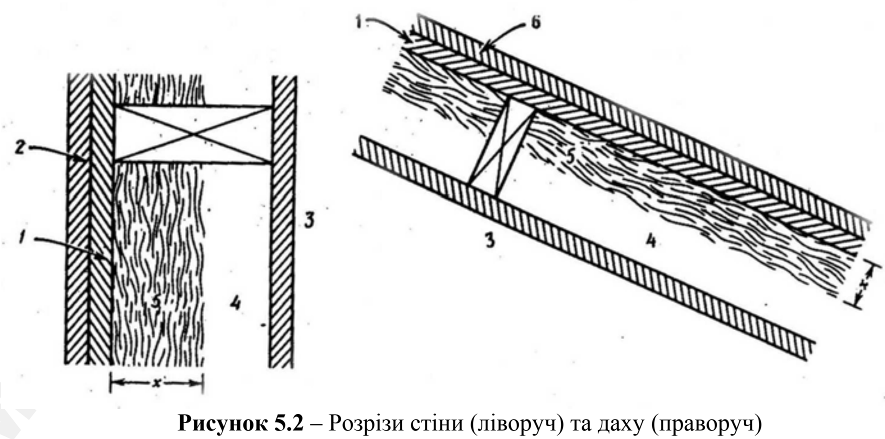


Рисунок 5.2 – Розрізи стіни (ліворуч) та даху (праворуч)
1 – фанера, 1/2 дюйма; 2 – зовнішній щит, 1/2 дюйма; 3 – внутрішній щит, 1/2 дюйма; 4 – повітря; 5 – скловолокно; 6 – покрівельна обшивка

Цільова функція та врахованві обмеження наведені нижче.

Розміри будинку (відповідно до ескізу)

Довжина	60 фут
Висота стін	8 фут
Ширина	30 фут
Ширина південного схилу даху	25 фут
Ширина північного схилу даху	15 фут
3 дверей	7 × 3.5 фут
6 вікон	6 × 3 фут

Розрахункові погодні умови

Середня температура приміщення 68°F	(градієнт у вертикальному напрямку 2%)
Температура зовнішнього повітря 35°F	
Швидкість вітру 10 миль/год	
Коеффіцієнт тепловіддачі стін із зовнішньої сторони будинку	h_0 = 6 БТЕ/год × фут^2 × °F.
Коеффіцієнт тепловіддачі стін із внутрішньої сторони будинку	h_i = 1 БТЕ/год × фут^2 × °F

Теплоізоляція

Внутрішню поверхню стіни утворює оздоблювальна панель товщиною 1/2 дюйма.

Ширина простору для скловолокнистої ізоляції:

між зовнішніми та внутрішніми щитами стін	3.5 дюйм
між покрівельною обшивкою та внутрішнім щитом даху (товщина балки)	5.5 дюйм
Втрати тепла внаслідок тепловіддачі від стін	$= (T_{вн} - T_{зовн}) / RA = (33/4.5x)$ (БТЕ/год × фут^2)
Втрати тепла внаслідок тепловіддачі від даху	$= (T_{вн} - T_{зовн}) / RA = (39/4.5x)$ (БТЕ/год × фут^2)

Тут x – товщина скловолокнистої ізоляції в дюймах.

Дані для програми оптимізації

Оптимізується вартість системи сонячного опалення будинку з такими параметрами

Площа стін	2730 фут^2
Площа даху	2400 фут^2 (площа, яку можна використовувати для розміщення колектора, дорівнює 1500 фут^2)
Площа підлоги	1800 фут^2
Площа вікон	108 фут^2
Довжина щілин віконних отворів	108 фут
Довжина щілин дверних отворів	63 фут
Втрати тепла через стіни	= (33/4.5x) (БТЕ/год × фут^2)
Втрати тепла через дах	= (39/4.5x) (БТЕ/год × фут^2)

Змінні

x_1 – площа поверхні колектора, фут^2
x_2 – радіус бака, фут
x_3 – висота бака, фут
x_4 – товщина ізоляції стін, дюйм
x_5 – товщина ізоляції даху, дюйм
x_6 – довжина ущільнюючих прокладок, фут

Вартість в умовних одиницях

$3.5x_1 + 150$ (1) Колектор	$+ 15.7x_2 x_3 + 100$ (5) Матеріали бокової поверхні бака
$+ 0.0045x_1^{3/2} + 100$ (2) Система циркуляції води	$+ 31.4x_2^{1/2}x_3^{1/2} +100$ (6) Система регулювання бака
$+ x_1^{1/2}$ (3) Система регулювання колектора	$+ 137x_4$ (7) Ізоляція стін
$+ 6.28x_2^2$ (4) Матеріали днищ бака	$+ 72x_5$ (8) Ізоляція даху
	$+ 0.2x_6$ (9) Ущільнення щілин

Загалом вартість складає (колектор, системи акумулювання, циркуляції, регулювання та ізоляції):
P_0 = 450 умовних одиниць (незмінні величини із функції вартості P_0 виключаються) =

$$= \underset{(1)}{3.5x_1} + \underset{(2)}{0.0045x_1^{3/2}} + \underset{(3)}{x_1^{1/2}} + \underset{(4)}{6.28x_2^2} + \underset{(5)}{15.7x_2 x_3} + \underset{(6)}{31.4x_2^{1/2}x_3^{1/2}} + \underset{(7)}{137x_4} +$$

$$+ \underset{(8)}{72x_5} + \underset{(9)}{0.2x_6}$$

Обмеження	*Фізичне значення*
(1) $x_1 \leq 1500$.	Розміри колектора обмежуються площею південного схилу даху.
(2) $x_4 \leq 3.5$.	Товщина ізоляції стін обмежується шириною простору між зовнішнім та внутрішнім щитами стіни.
(3) $x_5 \leq 5.5$.	Товщина ізоляції даху обмежується товщиною балки.
(4) $2x_2 x_3 \geq 150$.	Площа поверхні бака обмежується умовою природної конвекції.
(5) $770x_1 \geq 48\ [2730 \times (33\ /\ 4.5x_4) + 2400\ (39\ /\ 4.5x_5) + 1800\ (1.5) + 108\ (3.3\ /\ 1.19) - (10^2 + 63)\ (0.075)\ (20) \times (0.24)\ (33)\ x_6]$.	Кількість тепла, яке необхідно віддати воді при ККД колектора 50%, для підігріву будинку протягом двох діб.
(6) $770x_1 \leq 1{,}2\ (140 - 68)\ (0.988)\ (\pi)\ x_2^2 x_3$.	Підведення тепла за добу = 1.2 × (теплоємність маси води всередині бака); напрямок знака нерівності визначається вимогою, щоб теплоємність бака була не меншою, ніж добова теплопродуктивність колектора.
(7) $x_6 \leq 171$.	Довжина ущільнюючих прокладок обмежується периметром віконних та дверних отворів.

5.2 оптимізація лінії електропередачі з проводом кільцевого перерізу

Цей приклад стосується оптимізації високовольтної лінії електропередачі. Зрозуміло, що проектувальник повинен прагнути до зменшення кількості опор на шляху від електростанції до центра споживання електроенергії. Один із способів збільшення відстані між опорами ЛЕП – використання більш товстих проводів. Але висока вартість електропровідних матеріалів вимагає зменшення їхньої витрати до можливого мінімуму. Крім того, при змінному струмі найбільша щільність струму припадає на периферійну зону перерізу. Тому центральна зона перерізу провідника може бути видалена без помітного зниження електропровідності. Таким чином, виявляються два аспекти розв’язання задачі: вибір розмірів поперечного перерізу провідника та визначення оптимальної відстані між опорами.

При заданому значенні струму I та при умові лінійної залежності щільності струму від радіуса, $J = k\ r$, розміри кільцевого перерізу провідника визначаються із співвідношення $I = 2\ \pi\ k\ (r_0^3 - r_i^3)\ /\ 3$. Береться, що теоретична функція, що описує криву провисання проводів між опорами, може бути апроксимована параболою: $d = wS^2/8t$, $T = t + wd$, де w – погонна вага проводу, а інші параметри зазначені на рис. 5.3.

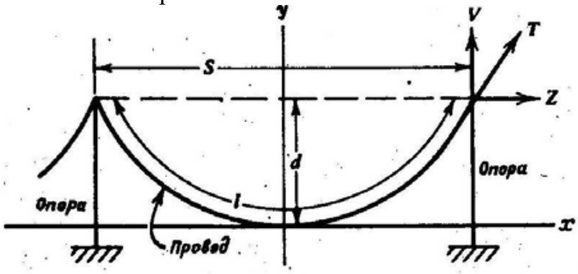


Рисунок 5.3 – Опори та провід лінії електропередачі

Введемо позначення: n – кількість опор (ціле число); D – загальна довжина лінії; S – відстань між опорами; m – кількість проводів, що підвішена на опорах; C_1 – вартість проводу у перерахунку на фунт матеріалу провідника; C_2 – вартість підпори.

Загальна вартість лінії електропередачі, яка являє собою мінімізовану цільову функцію, записується у вигляді $C_1\, m\, w\, D\, l\, /\, S + C_2\, (n + 1)$.

Задані такі сім обмежень:

(1) Величина прогину проводу d не повинна перевищувати 10 фут.

(2) Максимальне напруження розтяжіння при використанні алюмінієвого сплаву не повинно перевищувати 11000 фут/дюйм^2.

(3) Максимально припустима щільність струму біля поверхні провідника дорівнює 2400 А/дюйм^2.

(4) Максимально припустима сила струму у провіднику дорівнює 1000 А.

(5) Максимальне значення зовнішнього діаметра проводу дорівнює 1.5 дюйм.

(6) Максимально припустима товщина стінки кільцевого проводу складає 0.2 дюйм.

(7) Границя міцності матеріалу ізолятора дорівнює 15000 фут/дюйм^2.

Завдання до розділу 5

Придумати власний приклад поставлення задачі оптимізації в будь-якій технічній системі за своєю спеціальністю. Навести опис поставлення своєї задачі оптимізації згідно з прикладами, що наведені вище. Зокрема, визначити критерій оптимальності, параметри оптимізації, цільову функцію та обмеження.

РОЗДІЛ 6 ПАКЕТИ ПРИКЛАДНИХ ПРОГРАМ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ ОПТИМІЗАЦІЇ

6.1 загальні відомості

У даний час розроблені програмні продукти, що дозволяють розв'язувати оптимізаційні задачі за невеликий відрізок часу і з необов'язковим розумінням методу, який застосовується для розв'язання. Це істотно полегшує їх використання для розв'язання прикладних задач. Однак необхідно разібратися у самому пакеті перш ніж братися за розв'язання конкретної задачі.

Microsoft Excel – один із найпотужніших програмних продуктів для створення електронних таблиць і роботи з ними. Сильний бік Excel не тільки в його здатності виконувати різні обчислення, це можна зробити і за допомогою калькулятора, а й те, що Excel дозволяє проводити глибокий аналіз даних і отримувати в результаті нову корисну інформацію. У кожній новій версії компанія Microsoft пропонує додаткові можливості й удосконалює старі, щоб полегшити роботу користувачів із Excel. Версія Excel 97, зокрема, істотно удосконалена порівняно з Excel 95.

Maple – пакет програм, призначений для математичних обчислень. Дозволяє знаходити аналітичне рішення, чисельні рішення, дозволяє будувати графіки як на площині, так і в просторі. У Maple є бібліотеки. Використання бібліотеки дозволяє більш оптимально вирішувати поставлену задачу. За допомогою бібліотек можна знаходити екстремуми функції, її найбільше і найменше значення, максимум або мінімум функції, значення координат максимуму або мінімуму.

Програмний комплекс "Met_Opt" – пакет програм, що розроблений на кафедрі прикладної математики СумДУ. За його допомогою можна розв'язати будь-яку задачу оптимізації (звичайно за умови, що поставлення в математичному вигляді правильне). Комплекс являє собою сукупність програмних продуктів, кожен з яких може виступати як автономна програма.

Більш докладні відомості про проведення оптимізації за допомогою цих пакетів програм, а також огляд інших розробок, призначених для оптимізації, посилання на які можна знайти в Інтернет, наведені на сайті лабораторії дистанційоного навчання СумДУ в дистанційному курсі "Методи оптимізації" за адресою http://dl.sumdu.edu.ua/mo/ukr/ukr.html, [5].

6.2 розв'язання задач оптимізації за допомогою microsoft excel

По-перше, пакет Microsoft Excel є зручним засобом для проведення оптимізаційних розрахунків у "ручному" режимі, оскільки він дозволяє легко складати та копіювати формули, бачити одночасно всі результати проміжних розрахунків та відображати результати розрахунків графічно.

По-друге, пакет Microsoft Excel, як і інші продукти Microsoft Office, містить потужний засіб програмування – редактор Visual Basic для написання макросів (меню “Сервіс” –> “Макрос” –> “Редактор Visual Basic”). Достатньо досвідчений користувач за допомогою цього редактора може створити власну програму для виконання розрахунків за будь-яким з розглянутих методів оптимізації у середовищі Microsoft Excel.

По-третє, пакет Microsoft Excel містить вже готовий інструмент для виконання оптимізаційних розрахунків. Його можна визвати через меню “Сервіс” –> “Пошук розв’язання”. Для виконання пошуку оптимуму слід задати в Excel аналітичний вираз функції, що підлягає оптимізації, та початкову точку пошуку. Вікно “Пошук розв’язання” дозволяє також вибрати один із запропонованих методів оптимізації та задати потрібну точність. Пошук оптимуму виконується ітераційно. Після кожної ітерації користувач має можливість побачити проміжні результати пошуку. Після знаходження оптимального рішення Excel може видати звіт за результатами розрахунку. Приклад використання інструмента “Пошук розв’язання” також наведений в дистанційному курсі “Методи оптимізації” за адресою http://dl.sumdu.edu.ua/mo/ukr/ukr.html, [5].

Завдання до розділу 6

Виконати пошук мінімуму функції 2 додатка за допомогою інструмента “Пошук розв’язання” пакета Microsoft Excel.

ДОДАТОК А

(обов'язковий)

ФУНКЦІЯ 1

Побудувати функцію $f(x)$ та, залежно від її вигляду, знайти точку мінімуму або максимуму цієї функції на відрізку $[a, b]$. Почати пошук з точки $x = a$. Початковий крок взяти таким, що дорівнює $(x – a) / 4$. Варіант зазначає викладач.

1 $f(x) = \ln(x^2 – 2x + 2)$, $[0; 3]$.
2 $f(x) = 3x / (x^2 + 1)$, $[0; 5]$.
3 $f(x) = (2x – 1) / (x – 1)^2$, $[–1/2; 0]$.
4 $f(x) = (x + 2)\, e^{1 – x}$, $[–2; 2]$.
5 $f(x) = \ln(x^2 – 2x + 4)$, $[–1; 3/2]$.
6 $f(x) = x^3 / (x^2 – x + 1)$, $[–1; 1]$.
7 $f(x) = ((x + 1) / x)^3$, $[1; 2]$.
8 $f(x) = \sqrt{x - x^3}$, $[–2; 2]$.
9 $f(x) = 4 - e^{-x^2}$, $[0; 1]$.
10 $f(x) = (x^3 + 4) / x^2$, $[1; 2]$.
11 $f(x) = x\, e^x$, $[–2; 0]$.
12 $f(x) = (x – 2)\, e^x$, $[–2; 1]$.
13 $f(x) = (x – 1)\, e^{–x}$, $[0; 3]$.
14 $f(x) = x / (9 – x^2)$, $[–2; 2]$.
15 $f(x) = (1 + \ln x) / x$, $[1/e; e]$.
16 $f(x) = e^{4x-x^2}$, $[1; 3]$.
17 $f(x) = (x^5 – 8) / x^4$, $[–3; –1]$.
18 $f(x) = \dfrac{e^{2x} + 1}{e^x}$, $[–1; 2]$.
19 $f(x) = x \ln x$, $[1/e^2; 1]$.
20 $f(x) = x^3\, e^{x + 1}$, $[–4; 0]$.
21 $f(x) = x^2 – 2x + 2 / (x – 1)$, $[–1; 3]$.
22 $f(x) = (x + 1)\sqrt[3]{x^2}$, $[–4/5; 3]$.
23 $f(x) = e^{6x-x^2}$, $[–3; 3]$.
24 $f(x) = (\ln x) / x$, $[1; 4]$.
25 $f(x) = 3x^4 – 16x^3 + 2$, $[–3; 1]$.
26 $f(x) = x^5 – 5x^4 + 5x^3 + 1$, $[–1; 2]$.
27 $f(x) = (3 – x)\, e^{– x}$, $[0; 5]$.
28 $f(x) = \sqrt{3}/2 + \cos x$, $[0; \pi/2]$.
29 $f(x) = 108x – x^4$, $[–1; 4]$.
30 $f(x) = x^4 / 4 – 6x^3 + 7$, $[16; 20]$.

Продовження додатка А

ФУНКЦІЯ 2

Побудувати функцію $f\,(x,\,y)$ та знайти точку мінімуму (f_{min}) або максимуму (f_{max}) цієї функції. Почати пошук з точки (3, 3).

1 $f\,(x, y) = y\sqrt{x} - 2y^2 - x + 14y$, f_{max}.
2 $f\,(x, y) = x^3 + 8y^3 - 6xy + 5$, f_{min}.
3 $f\,(x, y) = 1 + 15x - 2x^2 - xy - 2y^2$, f_{max}.
4 $f\,(x, y) = 1 + 6x - x^2 - xy - y^2$, f_{max}.
5 $f\,(x, y) = x^3 + y^2 - 6xy - 39x + 18y + 20$, f_{min}.
6 $f\,(x, y) = 2x^3 + 2y^3 - 6xy + 5$, f_{min}.
7 $f\,(x, y) = 3x^3 + 3y^3 - 9xy + 10$, f_{min}.
8 $f\,(x, y) = x^2 + xy + y^2 + x - y + 1$, f_{min}.
9 $f\,(x, y) = 4\,(x - y) - x^2 - y^2$, f_{max}.
10 $f\,(x, y) = 6\,(x - y) - 3x^2 - 3y^2$, f_{max}.
11 $f\,(x, y) = x^2 + xy + y^2 - 6x - 9y$, f_{min}.
12 $f\,(x, y) = (x - 2)^2 + 2y^2 - 10$, f_{min}.
13 $f\,(x, y) = (x - 5)^2 + y^2 + 1$, f_{min}.
14 $f\,(x, y) = x^3 + y^3 - 3xy$, f_{min}.
15 $f\,(x, y) = 2xy - 2x^2 - 4y^2$, f_{max}.
16 $f\,(x, y) = x\sqrt{y} - x^2 - y + 6x + 3$, f_{max}.
17 $f\,(x, y) = 2xy - 5x^2 - 3y^2 + 2$, f_{max}.
18 $f\,(x, y) = xy\,(12 - x - y)$, f_{max}.
19 $f\,(x, y) = xy - x^2 - y^2 + 9$, f_{max}.
20 $f\,(x, y) = 2xy - 3x^2 - 2y^2 + 10$, f_{max}.
21 $f\,(x, y) = x^3 + 8y^3 - 6xy + 1$, f_{min}.
22 $f\,(x, y) = y\sqrt{x} - y^2 - x + 6y$, f_{max}.
23 $f\,(x, y) = x^2 - xy + y^2 + 9x - 6y + 20$, f_{min}.
24 $f\,(x, y) = xy\,(6 - x - y)$, f_{max}.
25 $f\,(x, y) = x^2 + y^2 - xy + x + y$, f_{min}.
26 $f\,(x, y) = x^2 + xy + y^2 - 2x - y$, f_{min}.
27 $f\,(x, y) = (x - 1)^2 + 2y^2$, f_{min}.
28 $f\,(x, y) = xy - 3x^2 - 2y^2$, f_{max}.
29 $f\,(x, y) = x^2 + 3\,(y + 2)^2$, f_{min}.
30 $f\,(x, y) = 2\,(x + y) - x^2 - y^2$, f_{max}.

Продовження додатка А

ФУНКЦІЯ 3

Побудувати функцію $f\,(x,\ y)$ та знайти найбільше та найменше значення функції в області D, обмеженій заданими лініями. Почати пошук з точки з найменшою абсцисою та ординатою.

1 $f(x, y) = 3x + y - xy$, D: $y = x$, $y = 4$, $x = 0$.
2 $f(x, y) = xy - x - 2y$, D: $x = 3$, $y = x$, $y = 0$.
3 $f(x, y) = x^2 + 2xy - 4x + 8y$, D: $x = 0$, $x = 1$, $y = 0$, $y = 2$.
4 $f(x, y) = 5x^2 - 3xy + y^2$, D: $x = 0$, $x = 1$, $y = 0$, $y = 1$.
5 $f(x, y) = x^2 + 2xy - y^2 - 4x$, D: $x - y + 1 = 0$, $x = 3$, $y = 0$.
6 $f(x, y) = x^2 + y^2 - 2x - 2y + 8$, D: $x = 0$, $y = 0$, $x + y - 1 = 0$.
7 $f(x, y) = 2x^3 - xy^2 + y^2$, D: $x = 0$, $x = 1$, $y = 0$, $y = 6$.
8 $f(x, y) = 3x + 6y - x^2 - xy - y^2$, D: $x = 0$, $x = 1$, $y = 0$, $y = 1$.
9 $f(x, y) = x^2 - 2y^2 + 4xy - 6x - 1$, D: $x = 0$, $y = 0$, $x + y - 3 = 0$.
10 $f(x, y) = x^2 + 2xy - 10$, D: $y = 0$, $y = x^2 - 4$.
11 $f(x, y) = xy - 2x - y$, D: $x = 0$, $x = 3$, $y = 0$, $y = 4$.
12 $f(x, y) = 0.5x^2 - xy$, D: $y = 8$, $y = 2x^2$.
13 $f(x, y) = 3x^2 + 3y^2 - 2x - 2y + 2$, D: $x = 0$, $y = 0$, $x + y - 1 = 0$.
14 $f(x, y) = 2x^2 + 3y^2 + 1$, D: $y = \sqrt{9 - \frac{9}{4}x^2}$, $y = 0$.
15 $f(x, y) = x^2 - 2xy - y^2 + 4x + 1$, D: $x = -3$, $y = 0$, $x + y + 1 = 0$.
16 $f(x, y) = 3x^2 + 3y^2 - x - y + 1$, D: $x = 5$, $y = 0$, $x - y - 1 = 0$.
17 $f(x, y) = 2x^2 + 2xy - 0.5y^2 - 4x$, D: $y = 2x$, $y = 2$, $x = 0$.
18 $f(x, y) = x^2 - 2xy + 2.5y^2 - 2x$, D: $x = 0$, $x = 2$, $y = 0$, $y = 2$.
19 $f(x, y) = xy - 3x - 2y$, D: $x = 0$, $x = 4$, $y = 0$, $y = 4$.
20 $f(x, y) = x^2 + xy - 2$, D: $y = 4x^2 - 4$, $y = 0$.
21 $f(x, y) = x^2y\,(4 - x - y)$, D: $x = 0$, $y = 0$, $y = 6 - x$.
22 $f(x, y) = x^3 + y^3 - 3xy$, D: $x = 0$, $x = 2$, $y = -1$, $y = 2$.
23 $f(x, y) = 4\,(x - y) - x^2 - y^2$, D: $x + 2y = 4$, $x - 2y = 4$, $x = 0$.
24 $f(x, y) = x^2 + 2xy - y^2 - 4x$, D: $x = 3$, $y = 0$, $y = x + 1$.
25 $f(x, y) = 6xy - 9x^2 - 9y^2 + 4x + 4y$, D: $x = 0$, $x = 1$, $y = 0$, $y = 2$.
26 $f(x, y) = x^2 + 2xy - y^2 - 2x + 2y$, D: $y = x + 2$, $y = 0$, $x = 2$.
27 $f(x, y) = 4 - 2x^2 - y^2$, D: $y = 0$, $y = \sqrt{1 - x^2}$.
28 $f(x, y) = 5x^2 - 3xy + y^2 + 4$, D: $x = -1$, $x = 1$, $y = -1$, $y = 1$.
29 $f(x, y) = x^2 + 2xy + 4x - y^2$, D: $x + y + 2 = 0$, $x = 0$, $y = 0$.
30 $f(x, y) = 2x^2y - x^3y - x^2y^2$, D: $x = 0$, $y = 0$, $x + y = 6$.

СПИСОК ЛІТЕРАТУРИ

1. Шевченко В.С. Введение в оптимальное проектирование машин. – Минск: Наука и техника, 1974. – 112 с.

2. Банди Б. Методы оптимизации. Вводный курс / Пер. с англ. – М.: Радио и связь, 1988. – 128с.

3. Пантелеев А.В., Летова Т.А. Методы оптимизации в примерах и задачах. – М.: Высшая школа, 2002. – 544с.

4. Табунщиков Ю.А., Бродач М.М. Математическое моделирование и оптимизация тепловой эффективности зданий. – М.: АВОК-ПРЕСС, 2002. – 194 с.

5. Любчак В.О., Острівна Л.Г. Комп'ютерна реалізація методів оптимізації: Навч. посіб. – Суми: Вид-во СумДУ. – 2002. – 160 с.

Додаткові посилання наведені також у дистанційному курсі "Методи оптимізації" за адресою http://dl.sumdu.edu.ua/mo/ukr/ukr.html.