Міністерство освіти і науки України
Сумський державний університет

Гусак О. Г., Панченко В. О., Хацко К. О.

МОДЕЛІ І МЕТОДИ ОПТИМІЗАЦІЇ

Конспект лекцій

Суми
Сумський державний університет
2022

Міністерство освіти і науки України
Сумський державний університет

МОДЕЛІ І МЕТОДИ ОПТИМІЗАЦІЇ

Конспект лекцій
для студентів спеціальності *144* «*Теплоенергетика*»
(освітня програма **«Енергетичний менеджмент»**)
усіх форм навчання

Затверджено
на засіданні кафедри
прикладної гідроаеромеханіки
як конспект лекцій
із дисципліни «Моделі
і методи оптимізації».
Протокол № 6 від 27.12.2022.

Суми
Сумський державний університет
2022

Конспект лекцій із дисципліни «Моделі і методи оптимізації» / укладачі: О. Г. Гусак, В. О. Панченко, К. О. Хацко. – Суми : Сумський державний університет, 2022. – 109 с.

Кафедра прикладної гідроаеромеханіки

ЗМІСТ

ВСТУП

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

У практичній діяльності часто з багатьох можливих розв'язків завдань необхідно вибрати оптимальний. Наприклад:

– з декількох варіантів перевезення сировини споживачам необхідно вибрати найбільш дешевий, але такий, що враховує обмеження на допустимі терміни поставок;

– з можливих планів розкроювання матеріалу вибрати такий, який би дозволив виконати план за якнайменшої кількості відходів тощо.

Зокрема вибір оптимальної схеми теплових та енергетичних мереж, а також найбільш придатних установок для використання у цих мережах та інші проблеми такого характеру є, по суті, головним завданням 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 комплексного порівняльного оцінювання якості схем, що розглядаються, та необхідного математичного апарату 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 способів підвищення ефективності роботи теплових та енергетичних мереж наполегливо потребує вивчення курсу «Моделі і методи оптимізації», змістовна основа якого здебільшого ґрунтується на спеціальному розділі математики – теорії оптимізації, а також на відповідних чисельних методах.

Курс «Моделі і методи оптимізації» ставить за мету:

– ознайомити студентів із 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.

Унаслідок вивчення дисципліни студент повинен знати основні, найбільш ефективні чисельні методи розв'язання задач оптимізації та їхні особливості.

Студент повинен уміти:

– правильно формулювати і класифікувати завдання оптимізації;

– вибирати або розробляти методи для їхнього розв'язання та реалізовувати ці методи у вигляді комп'ютерних програм;

– застосовувати сучасні моделі та методи оптимізації для виконання практичних завдань вибору та вдосконалення енергетичного обладнання.

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 сприяли створенню спеціальних математичних методів, і вже у XVIII столітті були закладені математичні основи оптимізації (варіаційне числення, чисельні методи та ін.). Однак до другої половини XX століття методи оптимізації в багатьох галузях науки і техніки застосовували дуже рідко, оскільки практичне використання математичних методів оптимізації вимагало величезної обчислювальної роботи, що без ЕОМ реалізувати було досить складно, а в низці випадків – неможливо. Особливо великі труднощі виникали під час розв'язання задач оптимізації процесів у хімічній технології через велику кількість параметрів і їхній складний взаємозв'язок між собою. За наявності ЕОМ розв'язання задач помітно спрощуються.

Як відомо, інженерні розрахунки, що проводять на етапі конструювання, завжди багатоваріантні, тобто задача створення будь-якої технічної системи завжди має безліч розв'язань. У міру ускладнення конструкції проєктувальнику все важче вирішувати, якою саме потрібно створити технічну систему, щоб вона виявилася найбільш досконалою. Зазвичай кількість можливих варіантів проєкту під час введення в розгляд кожного нового параметра зростає в експоненціальній залежності.

Наведемо кілька прикладів поставлення конструкторських завдань [1].

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

2. Перед авіаконструктором постійно постають завдання створення максимально легких і компактних конструкцій за умови забезпечення необхідної міцності та надійності.

3. Вибір конструктивних параметрів усіх типів гідравлічних об'ємних насосів визначено прагненням отримати максимальну продуктивність за мінімальних розмірів, ваги, вартості. Під час проєктування шестеренних, шиберних, гвинтових і поршневих насосів необхідно враховувати цілу низку аналогічних обмежувальних умов: в усіх типах насосів колова швидкість ротора обмежується вимогою відсутності кавітації; для кожного розмірного типу насоса є оптимальна кількість рухомих елементів, їхні оптимальні розміри тощо.

4. Під час конструювання гідравлічних і пневматичних циліндрів-підйомників ставлять завдання забезпечення максимальної вантажопідйомності за мінімальних габаритів, ваги та вартості. Водночас необхідно отримати оптимальні значення діаметра циліндра та довжини ходу за заданої потужності насоса або компресора і забезпечення надійності та стійкості. У телескопічних циліндрах до цього переліку шуканих оптимальних величин додано ще й оптимальну кількість ланок.

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

– економічність пристрою – потужність пристрою (за фіксованих габаритних розмірів);

– економічність пристрою – вартість пристрою (за фіксованих номінальних параметрів роботи).

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

Поставлення завдання 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 проєктного розв’язання.

Одночасно об’єкту оптимізації не потрібно приписувати два та більше критеріїв оптимальності, тому що практично завжди екстремум одного критерію не відповідає екстремуму іншого.

Типовий приклад неправильного поставлення задавдання оптимізації: «Одержати максимальну продуктивність за мінімальної собівартості».

Помилка полягає в тому, що поставлено завдання пошуку оптимуму двох показників, що суперечать один одному по своїй суті.

Правильне поставлення задавдання могло бути таким:

а) одержати максимальну продуктивність за заданої собівартості;

б) одержати мінімальну собівартість за заданої продуктивності.

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

Поставлення завдання оптимізації вимагає також наявність можливості варіювати значеннями деяких параметрів, від яких залежить значення критерію оптимальності об’єкта. Інакше кажучи, об’єкт повинен мати ступені свободи – *параметри оптимізації*. У завданнях оптимізації інженерних пристроїв і систем параметрами оптимізації можуть бути насамперед геометричні розміри та кількість елементів конструкції. Звичайно в поставлення завдання оптимізації вводять лише ті параметри, що справляють найбільш значущий вплив на критерій оптимальності. Відзначимо, що діапазон варіювання цих параметрів зазвичай обмежений із міркувань міцності або економічності.

Вплив, який здійснюють ці параметри на значення критерію оптимальності, описує цільова функція. *Цільова функція* – функція, що пов’язує критерій оптимальності з параметрами оптимізації. Як цільову функцію, так і параметри оптимізації потрібно вибирати так, щоб їх можна було оцінити кількісно, оскільки лише в цьому разі можна порівнювати ефекти від вибору тих або інших значень параметрів оптимізації.

1.1.2 Математична постановка задач оптимізації. Види обмежень

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

Надалі функцію $f(x)$ будемо називати *цільовою функцією*, рівняння $h_k(x) = 0$ – обмеженнями-рівняннями, а нерівності $g_i(x) \ge 0$ – обмеженнями-нерівностями. Водночас передбачено, що всі функції, що фігурують у задачі, є дійсно значними, а числом обмежень є скінченна величина.

Завдання загального вигляду «мінімізувати функцію $f(x)$ за обмежень $h_k(x) = 0$, $k = 1, ..., K$, $g_i(x) \ge 0$, $j = 1, ..., J$, $x_i(u) \ge x_i \ge x_i(l)$, $i = 1, ..., N$» 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 $f(x)$, $h_k(x)$, $g_i(x)$ і розмірності вектора x. Задачі без обмежень, у яких x являє собою одномірний вектор, називаються задачами з однією змінною і становлять найпростіший, але водночас дуже важливий підклас оптимізаційних задач. Завдання умовної оптимізації, у яких функції $h_k(x)$ і $g_i(x)$ є лінійними, називаються задачами з лінійними обмеженнями.

1.1.3 Про критерій оптимальності та його визначення

Зазвичай оптимізують величину, яка пов'язана з економічністю роботи об'єкта, що розглядають (апарат, цех, завод). Варіант роботи, що оптимізує об'єкт, потрібно оцінювати деякою кількісною мірою – критерієм оптимальності.

Критерієм оптимальності називають кількісне оцінювання якості об'єкта, що оптимізують.

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 визначає конкретне завдання оптимізації.

Отже, завдання оптимізації зводиться до знаходження екстремуму цільової функції.

Найбільш загальним поставленням оптимального завдання є вираження критерію оптимальності у вигляді економічної оцінки (продуктивність, собівартість продукції, прибуток, рентабельність). Однак $f(x)=(x_1-2)^2+(x_2-5)^2$ в окремих завданнях 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 цьому разі критерієм оптимальності може бути технологічна характеристика, що побічно оцінює економічність роботи агрегата (час контакту, вихід продукту, ступінь перетворення, температура). Наприклад, установлюється оптимальний температурний профіль, тривалість циклу «реакція – регенерація». Але в будь-якому разі будь-який критерій оптимальності має економічну природу.

Розглянемо більш докладно вимоги, що потрібно висувати 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.

Під час поставлення конкретних завдань 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 завданнях конструювання, вид цільової функції є невідомим або відомим лише наближено. Тоді отримання цільової функції передбачає проведення тривалого експериментального дослідження (або розрахункового – за допомогою застосування сучасних чисельних методів моделювання фізичних процесів). У цьому разі для планування дослідження використовують теорію планування експерименту, що також ґрунтується на методах оптимізації, які ми будемо розглядати далі. Метою застосування теорії планування експерименту є відшукання таких значень параметрів оптимізації, за яких критерій оптимальності сягає максимуму, із забезпеченням найбільшої точності та найменшої можливої кількості дослідів.

Отже, для виконання завдання оптимізації необхідно:

а) вибрати критерій оптимальності та параметри оптимізації;

б) установити 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 за кожного набору значень параметрів оптимізації доведеться отримувати за допомогою трудомісткості дослідження;

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

1.2 Функції однієї змінної. Необхідні та 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 ніж $f(x_0)$. Функція $f(x)$ має глобальний мінімум у точці x^*, якщо для всіх x справедлива нерівність $f(x) \geq f(x^*)$ [2].

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

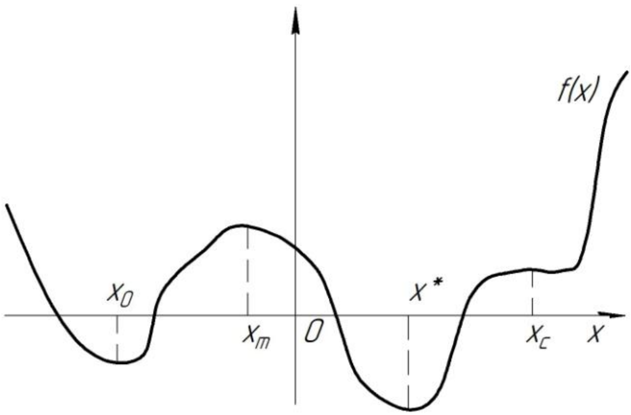

Рисунок 1.1 – Графічне зображення функції

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

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

Точка x_m, у якій досягається локальний мінімум, і точка x_c, яка є точкою 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 із від'ємного на додатний. У точці x_m знак змінюється з додатного на від'ємний, тоді 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

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

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. (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

$$f''(x_0) > 0, \tag{1.3}$$

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 $x=1$.

За умови $x=1/3$ похідна $f'(x)$ змінює знак на від'ємний, а за умови $x=1$ – з від'ємного на додатний. Отже, у точці $x=1/3$ досягається мінімум, а у точці $x=1$ – максимум.

1.3 Функції *n* змінних. Необхідні та достатні умови

Розглянемо функцію *n* дійсних змінних

$$f(x_1, x_2, x_3, ..., x_n)=f(x).$$

Точку у *n*-мірному евклідовому просторі з координатами $x_1, x_2, x_3, ..., x_n$ позначають вектором-стовпцем *x*. Градієнт функції, тобто вектор із компонентами $\partial f/\partial x_1, \partial f/\partial x_2, ..., \partial f/\partial x_n$, позначають $\mathrm{grad}\, f(x)$ або, іноді, $g(x)$. Матрицю Гессе (гессіан) функції $f(x)$ позначають як $G(x)$, і вона є симетричною матрицею $n \times n$ елементів виду

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

Тобто матриця $G(x)$ набуде вигляду

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

Функція $f(x)$ має локальний мінімум у точці x_0, якщо існує окіл точки x_0 такий, що $f(x) > f(x_0)$ у всіх точках цього околу, тобто існує додатна величина δ, така, що для $|x - x_0| < 0$ справедлива нерівність $f(x) \geq f(x_0)$.

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

За таких визначень і очевидних припущень щодо диференційованості можна узагальнити рівняння (1.2) і отримати

$$\begin{gathered} f(x_0 + h) - f(x_0) = \\ = \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n) + \ldots = \qquad (1.4) \\ = h^T \operatorname{grad} f(x) + 0{,}5 h^T G(x_0) h + \ldots \end{gathered}$$

Тоді, якщо x_0 є точкою мінімуму функції $f(x)$, то кожна перша часткова похідна $\partial f / \partial x_i$ $(i = 1, \ldots, n)$ повинна обертатися в нуль у точці x_0. Якщо це не так, то відповідним вибором h_i

можна домогтися того, що різниця $f\left(x_0+h\right)-f\left(x_0\right)$ буде від'ємною.

Отже, необхідною умовою мінімуму в точці x_0 є рівняння

$$\operatorname{grad} f\left(x_0\right)=0, \tag{1.5}$$

тобто

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

Тоді знак різниці $f\left(x_0+h\right)-f\left(x_0\right)$ визначає член

$$0{,}5h^T G\left(x_0\right)h. \tag{1.7}$$

Якщо матриця $G\left(x_0\right)$ додатно визначена, то цей член є додатним для всіх h. Отже, необхідними і достатніми умовами мінімуму є такі:

$$\operatorname{grad} f\left(x_0\right)=0,\ G\left(x_0\right) \text{ додатно визначена.} \tag{1.8}$$

Матриця називається додатно визначеною $\left(G\left(x_0\right)>0\right)$, якщо для будь-якого ненульового h виконується нерівність $h^T G\left(x_0\right)h>0$.

Необхідними і достатніми умовами максимуму є такі:

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

Приклад

Дослідити екстремальну точку (точки) функції $f\left(x\right)=x_1^2+x_2^2+x_3^2-4x_1-8x_2-12x_3+100$.

$$\operatorname{grad} f\left(x\right)=\begin{pmatrix} 2x_1-4 \\ 2x_2-8 \\ 2x_3-12 \end{pmatrix}=0 \text{ за умови } x_1=2,\ x_2=4,\ x_3=6.$$

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

$$h^T G\left(x_0\right) h = \left[h_1 \ h_2 \ h_3\right] \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 2h_1^2 + 2h_2^2 + 2h_3^2 > 0.$$

Тобто матриця $G\left(x_0\right)$ дійсно є додатно визначеною. Усі власні значення додатні і дорівнюють 2.

Отже, у точці (2; 4; 6) функція $f\left(x\right)$ сягає мінімуму.

РОЗДІЛ 2
МЕТОДИ ПОШУКУ ДЛЯ ФУНКЦІЇ ОДНІЄЇ ЗМІННОЇ

Незважаючи на те, що безумовна оптимізація функції однієї змінної – найбільш простий тип оптимізаційних задач, вона посідає центральне місце в теорії оптимізації як із теоретичного, так і з практичного погляду. Це пов'язано з тим, що задачі однопараметричної оптимізації досить поширені в інженерній практиці і, крім того, застосовують під час 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 за допомогою зменшення інтервалу пошуку, називаються методами виключення інтервалів.

Усі методи одновимірної оптимізації ґрунтуються 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 мінімуму, 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 так, 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 за допомогою послідовного виключення частин вихідного обмеженого інтервалу. Пошук завершується тоді, коли підінтервал, що залишився, зменшується до досить малих розмірів.

Головна перевага таких пошукових методів – ґрунтуються на обчисленні тільки значень функції і, отже, не вимагають виконання умови диференційованості та запису в аналітичному вигляді. Остання властивість особливо цінна під час імітаційного моделювання.

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

1) *етап установлення 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 установленої величини.

Етап установлення границь інтервалу

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

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

де x_0 – довільно обрана початкова точка;

Δ – величина кроку, яку підбирають певним способом.

Знак Δ визначають за допомогою порівняння значень $f(x_0)$, $f(x_0 + |\Delta|)$ та $f(x_0 - |\Delta|)$:

– якщо $f\left(x_0-|\Delta|\right)>f\left(x_0\right)>f\left(x_0+|\Delta|\right)$, то згідно з припущенням про унімодальність точка мінімуму має розміщуватися правіше точки x_0, і величина Δ має мати додатне значення;

– якщо $f\left(x_0-|\Delta|\right)<f\left(x_0\right)<f\left(x_0+|\Delta|\right)$, то Δ має від’ємне значення;

– якщо ж $f\left(x_0-|\Delta|\right)>f\left(x_0\right)<f\left(x_0+|\Delta|\right)$, то точка мінімуму лежить між $x_0-|\Delta|$ і $x_0+|\Delta|$, і пошук граничних точок завершено;

– якщо $f\left(x_0-|\Delta|\right)<f\left(x_0\right)>f\left(x_0+|\Delta|\right)$, то маємо суперечність припущенню про унімодальність.

Етап зменшення інтервалу

Після того, як установлено 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 в 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 $f(x)=(100-x)^2$ за заданої початкової точки $x_0=30$ та величини кроку $|\Delta|=5$.

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

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

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

Застосовуємо формулу (2.1).

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

В 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 точно половину інтервалу на кожній ітерації. Алгоритм розрахунку наведено нижче.

Завдання. Знайти мінімальне значення функції $f\left(x\right)$ на відрізку [*a*; *b*].

Крок 1

Прийняти: $x_m=\left(a+b\right)/2$; довжина інтервалу $L=b-a$. Обчислити значення $f\left(x_m\right)$.

Крок 2

Прийняти: $x_1=a+L/4$; $x_2=b-L/4$. Обчислити значення $f\left(x_1\right)$ та $f\left(x_2\right)$.

Крок 3

Порівняти $f\left(x_1\right)$ та $f\left(x_m\right)$:

– якщо $f\left(x_1\right) < f\left(x_m\right)$, 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 $f\left(x_1\right) \geq f\left(x_m\right)$, то перейти до кроку 4.

Крок 4

Порівняти $f\left(x_2\right)$ та $f\left(x_m\right)$:

– якщо $f\left(x_2\right) < f\left(x_m\right)$, то виключити інтервал [a; x_m), далі покласти $a = x_m$, $x_m = x_2$. Перейти до кроку 5;

– якщо $f\left(x_2\right) \geq f\left(x_m\right)$, 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 $\left|L\right|$ є меншою від заздалегідь установленої точності ε, тобто достатньо мала, то закінчити пошук. В іншому разі 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 ніж два обчислення значень функції.

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

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

Приклад

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

Ітерація 1

Крок 1

Приймаємо $a=60$; $b=150$; $L=150-60=90$; $x_m=\left(60+90\right)/2=105$; $f\left(x_m\right)=25$.

Крок 2

Обчислюємо

$x_1=60+90/4=82,5$; $x_2=150-90/4=127,5$.

Крок 3

Маємо $f\left(x_1\right)=306,25<f\left(x_m\right)$; $f\left(x_2\right)=756,25>f\left(x_m\right)$.

Крок 4–5

Виключаємо з розгляду інтервали (60; 82,5) та (127,5; 150). Довжина інтервалу пошуку зменшується з 90 до 45.

Ітерація 2

Приймаємо $a=82,5$; $b=127,5$; $x_m=105$.

Обчислюємо

$x_1=82,5+45/4=93,75$; $x_2=127,5-45/4=116,25$.

Маємо $f\left(x_1\right)=39,06>f\left(x_m\right)$; $f\left(x_2\right)=264,06>f\left(x_m\right)$.

Новий інтервал невизначеності становить від 93,75 до 116,25. Довжина інтервалу становить 22,5.

Ітерація 3

Приймаємо $a = 93,75$; $b = 116,25$; $x_m = 105$.

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 становить від 99,38 до 110,63. Довжина інтервалу становить 11,25, тобто 1/8 від початкового інтервалу невизначеності.

2.1.3 Метод Фібоначчі

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

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

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?

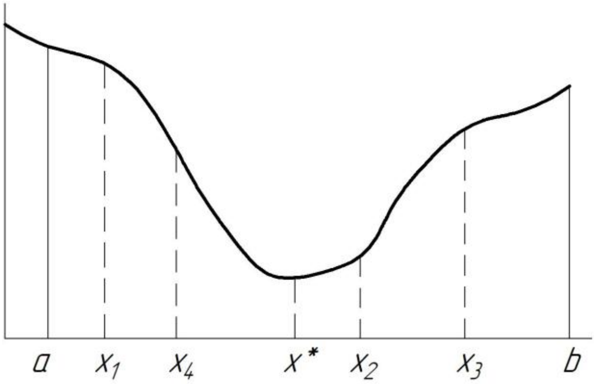

Рисунок 2.1 – Графічне зображення функції

Будемо вважати $x_2 - x_1 = L$ і $x_3 - x_2 = R,$ причому $L > R,$ як подано на рисунку 2.2, і ці значення будуть фіксовані, якщо відомі $x_1,$ x_2 і x_3. Якщо x_4 є 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, виберемо x_4 так, 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 щодо точки x_2, що вже є всередині інтервалу. Будь-яке положення точки x_4 може призвести до того, що отриманий інтервал буде більшим за L. Поміщаючи x_4 симетрично щодо x_2, ми нічим не ризикуємо в будь-якому разі.

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 всередині інтервалу невизначеності симетрично щодо точки, що вже є там. Парадоксально, але щоб зрозуміти, як варто починати обчислення, необхідно розібратися в тому, як його варто закінчити.

На *n*-му обчисленні *n*-ну точку варто помістити симетрично щодо $(n-1)$-ї точки. Положення цієї останньої точки в принципі залежить від нас. Для того щоб отримати найбільше зменшення інтервалу на цьому етапі, варто розділити навпіл попередній інтервал. Тоді точка x_n буде збігатися з точкою x_{n-1}, водночас ми не одержуємо жодної 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, є інтервал невизначеності. Вони розміщуються на відстані $\varepsilon / 2$ 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 матиме довжину L_n, отже, $L_{n-1} = 2L_n - \varepsilon$ (рис. 2.2, нижня частина).

На попередньому етапі точки x_{n-1} і x_{n-2} повинні бути розміщені симетрично всередині 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).

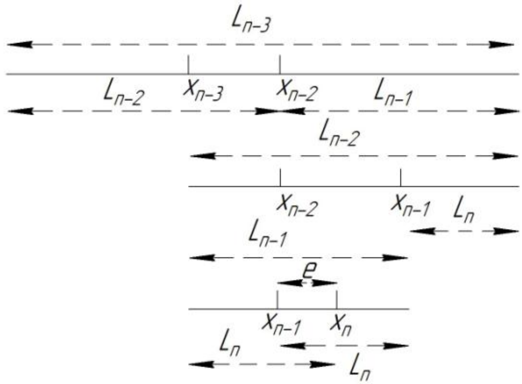

Рисунок 2.2 – Ілюстрація методу Фібоначчі

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).

Загалом

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

Отже

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

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

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

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

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 $L_{n-j} = F_{j+1}L_n - F_{j-1}\varepsilon$ за умови $j = 1, 2, ..., n-1.$ (2.3)

Якщо початковий інтервал $(a,\ b)$ має довжину $L = \left(b - a\right)$, 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 $1/F_n$ разів порівняно з його початковою довжиною (нехтуючи 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

$$L_2 = F_{n-1}L_n - \varepsilon F_{n-3} = F_{n-1}\frac{L_1}{F_n} + \varepsilon\frac{F_{n-1}F_{n-2} - F_nF_{n-3}}{F_n} = \\ = \frac{F_{n-1}}{F_n}L_1 + \frac{(-1)^n\varepsilon}{F_n}. \tag{2.5}$$

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

Отже, пошук методом Фібоначчі, 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, наступну точка x_4 завжди вибирають такою, що $x_3 - x_4 = x_2 - x_1$ або $x_4 - x_1 = x_3 - x_2$, тобто

$$x_4 = x_1 - x_2 + x_3. \tag{2.6}$$

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

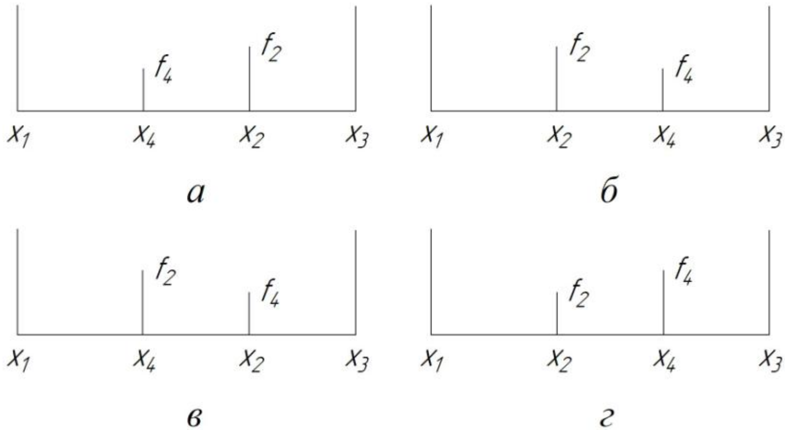

Рисунок 2.3 – Ілюстрація до методу Фібоначчі

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

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

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

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

Приклад

Знайти мінімум функції $f\left(x\right) = \left(100 - x\right)^2$ методом Фібоначчі. Інтервал пошуку $60 \le x \le 150$. Дозволено 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 $f(\omega) = (40-90\omega)^2$ за обмеження $0 \le \omega \le 1.$

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

Оскільки дозволено 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, у 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 рисунка 2.3 положення нової точки ω_4 вибираємо так, щоб вона розміщувалася симетрично щодо точки ω_2 (щоб нові інтервали невизначеності були однаковими) (див. рис. 2.4)

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

Тоді

$$\omega_4 = \omega_3 - \omega_2 + \omega_1 = 1 - 0,615 + 0 = 0,385.$$

Відповідне значення функції $f(0,385) = 29,0.$

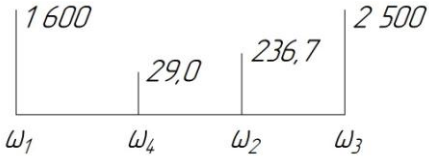

Рисунок 2.4 – Ілюстрація до прикладу методу Фібоначчі

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 щодо точки ω_4, щоб нові інтервали невизначеності були однаковими

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

Тоді

$$\omega_5 = \omega_2 - \omega_4 + \omega_1 = 0,615 - 0,385 + 0 = 0,23.$$

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 щодо точки ω_4, щоб нові інтервали невизначеності були однаковими

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

Тоді

$$\omega_6 = \omega_2 - \omega_4 + \omega_5 = 0,615 - 0,385 + 0,23 = 0,46.$$

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 становить $(150-60)/F_6 = = 90/13 = 6,92.$ Отримана точка мінімального значення функції $x = 90 \cdot 0,46 + 60 = 101,4.$

Завдання до підрозділу 2.1.3

Виконати пошук мінімуму або максимуму функції (додаток А) 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. формулу (2.5)) [2].

Метод «золотого перетину» майже настільки ж ефективний, як і метод Фібоначчі, однак водночас не потрібно знати *n* – кількість обчислень функції, що визначається спочатку. Після того як виконано *j* обчислень, на підставі тих самих міркувань, що і раніше (див. формулу (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 $L_{n-1} = 2L_n - \varepsilon.$ Якщо відношення наступних інтервалів буде постійним, тобто

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

то

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

тобто $\tau = 1 + 1/\tau.$

Отже, $\tau^2 - \tau - 1 = 0,$ звідки $\tau = \left(1 + 5^{1/2}\right)/2 \approx 1{,}618033989.$ Тоді

$$\frac{L_{j-1}}{L_{j+1}} = \tau^2, \quad \frac{L_{j-2}}{L_{j+1}} = \tau^3 \text{ тощо.}$$

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

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

Унаслідок аналізу двох розглянутих значень функції буде визначений той інтервал, що потрібно дослідити надалі. Цей інтервал буде містити одну з попередніх точок і наступну точку, що розміщуються симетрично їй. Перша точка розташована 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). Можна бачити, 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».

Отже, якщо відшукується інтервал $\left(x_0; x_3\right)$ 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 $\left(x_0; x_2\right)$;

б) $f_1 > f_2$. Новий інтервал $\left(x_1; x_3\right)$.

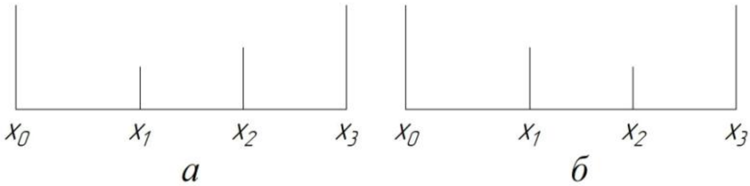

Рисунок 2.5 – Ілюстрація до методу «золотого перетину»

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

Крок 1

Обчислюємо коефіцієнт дроблення відрізка $\tau \approx 1,618$.

Крок 2

$x_1 + (1 - 1/\tau)(b - a)$, 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 із 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 кроків 2 і 4.

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

Так, 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 у двох пробних точках.

Приклад

Знайти мінімум функції $f\left(x\right)=\left(100-x\right)^2$ 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 $\omega=\left(x-60\right)/90$. Тоді задача полягає у відшуканні мінімуму функції $f\left(\omega\right)=\left(40-90\omega\right)^2$ за умови обмеження $0\le\omega\le 1$.

Спочатку отримуємо значення функції на краях інтервалу

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

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)

$$\omega_1 = a + \left(1 - 1/\tau\right)\left(b - a\right) = 0 + \left(1 - 0,618\right) \cdot 1 = 0,382.$$

Відповідне значення функції $f\left(0,382\right) = 31,6.$

$$\omega_2 = a + 1/\tau\left(b - a\right) = 0 + 0,618 \cdot 1 = 0,618.$$

Відповідне значення функції $f\left(0,618\right) = 244,0.$

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

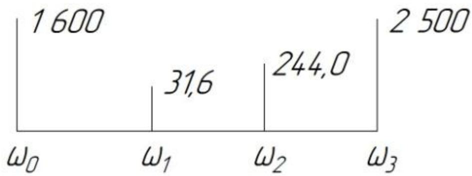

Рисунок 2.6 – Ілюстрація до прикладу методу «золотого перетину»

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)

$$\omega_4 = a + \left(1 - 1/\tau\right)\left(b - a\right) = 0 + \left(1 - 0,618\right) \cdot 0,618 = 0,236.$$

Відповідне значення функції $f\left(0,236\right) = 352.$

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)

$$\omega_5 = a + 1/\tau\left(b - a\right) = 0,236 + 0,618 \cdot 0,382 = 0,472.$$

Відповідне значення функції $f\left(0,472\right) = 6,15.$

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 $f\left(x\right)$ після шести обчислень становить $\left(150 - 60\right)/\tau^{6-1} = 90/11,1 = 8,12.$ Отримана точка мінімального значення функції $x = 90 \cdot 0,472 + 60 = 102,5.$

Завдання до підрозділу 2.1.4

Виконати пошук мінімуму або максимуму функції (див. додаток А) за методом «золотого перетину».

2.1.5 Порівняння методів виключення інтервалів

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

Нагадаємо, що за умови використання методу ділення інтервалу навпіл довжина отриманого інтервалу становить $L_1\left(0,5\right)^{n/2}$, у методі «золотого перетину» – $L_1/\left(1,618\right)^{n-1}$, у методі Фібоначчі – L_1 / F_n.

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

$Fr\left(n\right) = \left(0,5\right)^{n/2}$ – 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 в таблиці 2.1.

Таблиця 2.1 – Величини відносного зменшення інтервалу невизначеності

Методи 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 випливає, що пошук за допомогою методу Фібоначчі забезпечує найбільше відносне зменшення вихідного інтервалу за однієї і тієї самої кількості обчислень значень функції.

2.2 Методи поліноміальної інтерполяції

У двох попередніх розділах була зроблена спроба знайти малий інтервал, у якому є мінімум функції. У наступних двох розділах застосовано інший підхід. Використано 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

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

де коефіцієнти A, B, і C визначають із рівнянь

$$\begin{aligned} Aa^2 + Ba + C &= f_a, \\ Ab^2 + Bb + C &= f_b, \\ Ac^2 + Bc + C &= f_c. \end{aligned} \tag{2.11}$$

Після перетворень цих рівнянь маємо

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

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

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

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

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 $f(A-H)$. В іншому разі за третю точку взяти $A + 2H$ і знайти $f(A+2H)$ (рис. 2.7).

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

4. Якщо різниця між найменшим значенням функції і наступним найменшим значенням функції менша за задану точність, 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 рисунку 2.8 залишені точки x_1, x_2 і x_4, а не точки x_1, x_2 і x_3.

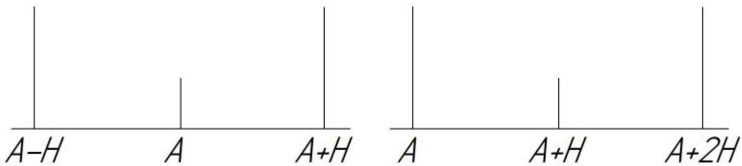

Рисунок 2.7

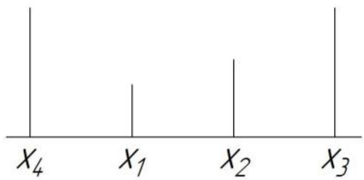

Рисунок 2.8

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. (2.13)) може стати взагалі недосяжним. Щоб подолати ці труднощі, перепишемо рівняння (2.13) для другої і наступної інтерполяції

$$\delta = \frac{1}{2}\left(a+b\right) + \frac{0{,}5\left(f_a - f_b\right)\left(b-c\right)\left(c-a\right)}{\left(b-c\right) f_a + \left(c-a\right) f_b + \left(a-b\right) f_c}. \qquad (2.14)$$

Вищеописаний метод часто називають методом Пауела.

2.2.2 Інтерполяція вищих порядків

Загалом 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*, …, *n* є відомими і дорівнюють відповідно f_a, f_b, f_c, …, f_n, то функція $f(x)$ може бути апроксимована поліномом такого виду:

$$\varphi(x) = Ax^n + Bx_{n-1} + Cx^{n-2} + ... + Dx^2 + Ex + F. \qquad (2.15)$$

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

$$Aa^n + Ba^{n-1} + Ca^{n-2} + ... + Da^2 + Ea + F = f_a,$$

$$Ab^n + Bb^{n-1} + Cb^{n-2} + ... + Db^2 + Eb + F = f_b,$$

$$Ac^n + Bc^{n-1} + Cc^{n-2} + ... + Dc^2 + Ec + F = f_c,$$

$$...$$

$$Af^n + Bf^{n-1} + Cf^{n-2} + ... + Df^2 + Ef + F = f_f.$$

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

Відзначимо, однак, що інтерполяція вищих порядків є неекономічною через необхідність обчислень великої кількості значень функції. На практиці інтерполяцію поліномом ступеня вище ніж три використовують рідко. Перевага методів інтерполяції вищих порядків – можливість відшукувати мінімуми функції навіть у тому разі, 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 у разі пошуку значень x у точках перегину функції $f(x)$ полягає в розв'язанні рівняння

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

У тому разі, якщо цільова функція містить члени, що містять, наприклад, x у третій і більш високих ступенях, то одержати аналітичне розв'язання рівняння $f'(x) = 0$ складно. У цих випадках доцільно використовувати чисельні методи знаходження коренів нелінійних рівнянь. У цьому підрозділі ми розглянемо один із 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 завдяки очевидним припущенням про безперервність буде існувати корінь x^* цього рівняння, причому $a < x^* < b$ (рис. 2.9).

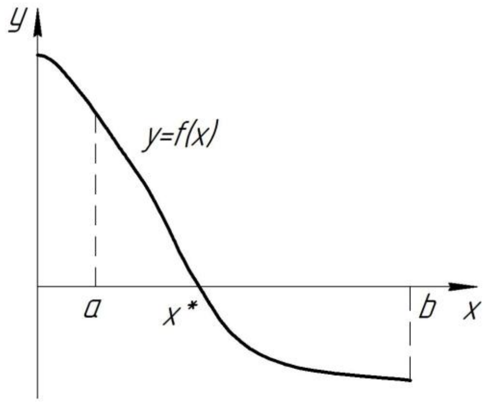

Рисунок 2.9

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 взята як поточне наближення до оптимальної точки, то лінійна функція, що апроксимує функцію $f'(x)$ у точці x_k, матиме вигляд

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

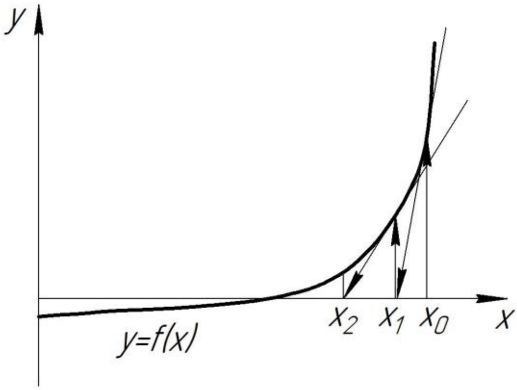

Рисунок 2.10

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

Крок 1

Наступне наближення до стаціонарної точки x^* визначається за формулою

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

Крок 2

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

Крок 3

Якщо $\left|f'(x_{k+1})\right|<\varepsilon$, то закінчити пошук. В іншому 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 $x_1 = 1$. Необхідна 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

$$x_1 = 1; \quad f'(x_1) = -12; \quad f''(x_1) = 36.$$

Ітерація 2

$$x_2 = x_1 - f'(x_1)/f''(x_1) = 1{,}33; \quad f'(x_2) = -3{,}73; \quad f''(x_2) = 17{,}6.$$

Ітерація 3

$$x_3 = x_2 - f'(x_2)/f''(x_2) = 1{,}54; \quad f'(x_3) = -4{,}23; \quad f''(x_3) = 12{,}7.$$

Ітерація 4

$$x_4 = x_3 - f'(x_3)/f''(x_3) = 1{,}87; \quad f'(x_4) = -1{,}07; \quad f''(x_4) = 8{,}89.$$

Ітерація 5

$$x_5 = x_4 - f'(x_4)/f''(x_4) = 1{,}99; \quad f'(x_5) = -0{,}07; \quad f''(x_5) = 8{,}05.$$

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

Завдання до підрозділу 2.3

Виконати пошук мінімуму або максимуму функції (див. додаток А) 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 довела, що ці два методи є ефективними та можуть бути застосовані для великої кількості прикладних задач [2].

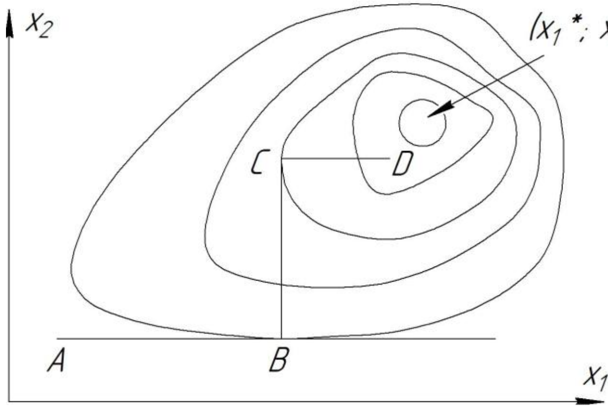

Рисунок 3.1 – Метод покоординатного спуску

Розглянемо функцію двох змінних. Її лінії рівня подані на рисунку 3.1, а мінімум лежить у точці $\left(x_1^*; x_2^*\right)$. Найпростішим методом пошуку є метод покоординатного спуску. З точки *A* робимо пошук мінімуму вздовж напрямку осі x_1 і, отже, знаходимо точку *B*, у котрої дотична до лінії постійного рівня паралельна осі x_1. Потім, роблячи пошук із точки *B* у напрямку осі x_2, отримуємо точку *C*, роблячи пошук паралельно осі x_2, одержуємо точку D тощо. Отже, ми приходимо до оптимальної точки. Цю ідею можна застосувати для функції n змінних.

Теоретично цей метод є ефективним у разі 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 із довільно заданої безлічі. З іншого боку, можна побудувати стратегію пошуку, у межах якої один 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. За цього підходу безліч напрямків пошуку вибирається у вигляді безлічі координатних напрямків у просторі змінних задачі. Але така стратегія може виявитися не тільки не ефективною, але й призвести до відсутності збіжності до точки локального оптимуму навіть за нескінченного числа ітерацій. Підвищення ефективності цього методу пов'язано з тією обставиною, що пошук, проведений у напрямку $\left(x_{i+1}-x_i\right)$, де $f\left(x_{i+1}\right)<f\left(x_i\right)$, 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 із перших алгоритмів, у якому під час визначення нового напрямку пошуку враховується інформація, отримана на попередніх ітераціях. Процедура Хука – Дживса являє собою комбінацію досліджувального пошуку з циклічною зміною змінних і 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 розглядають як успішний. В іншому 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 внаслідок цього 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. Нову точку зразка визначають відповідно до формули

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

Щойно рух за зразком не призводить до зменшення цільової функції, точку P_i фіксують як тимчасову базову точку і знову проводять досліджувальний пошук. Якщо внаслідок цього виходить точка з меншим значенням цільової функції, ніж у точці b_{i+1}, то її розглядають як нову базову точку. З іншого боку, якщо досліджувальний пошук виявиться невдалим, необхідно повернутися в точку b_{i+1} і провести досліджувальний пошук для виявлення нового напрямку мінімізації. Унаслідок цього виникає ситуація, коли такий пошук не сприяє успіху. У цьому разі потрібно зменшити величину кроку за допомогою 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 b_1 із 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 взятий одним і тим самим;

– коефіцієнт зменшення кроку $a > 1$;

– параметр закінчення пошуку $\varepsilon > 0$.

Б

1. Обчислити $f(x)$ у базовій точці b_1 для отримання відомостей про локальну поведінку функції $f(x)$ – отримуємо $f(b_1)$.

2. Кожна змінна по черзі змінюється додаванням довжини кроку. Отже, ми обчислюємо значення функції $f(b_1 + h_1 e_1)$, де e_1 – одиничний вектор у напрямку осі x_1.

Якщо це призводить до зменшення значення функції, то b_1 замінюється на $b_1 + h_1 e_1$. В іншому разі обчислюється значення функції $f(b_1 - h_1 e_1)$, і якщо її значення зменшилося, то b_1 замінюємо на $b_1 - h_1 e_1$.

Якщо жоден із виконаних кроків не призводить до зменшення значення функції, то точка b_1 залишається незмінною і розглядаються зміни в напрямку осі x_2, тобто обчислюємо значення функції $f(b_1 + h_2 e_2)$ тощо. Коли будуть розглянути всі n змінних, ми матимемо нову базову точку b_2.

3. Якщо $b_2 = b_1$, тобто зменшення функції не було досягнуто, то дослідження повторюється навколо тієї самої базової точки b_2, але зі зменшеною довжиною кроку: $h_j = h_j / a$.

Якщо $b_2 \neq b_1$, 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. Цю процедуру проводять так:

1. Розумно рухатися з базової точки b_2 у напрямку $b_2 - b_1$, оскільки пошук у цьому напрямку вже призвів до зменшення значення функції. Тому обчислимо функцію в точці зразка

$$P_1 = b_1 + \left(b_2 - b_1\right). \tag{3.1}$$

Загалом

$$P_i = b_i + \left(b_{i+1} - b_i\right). \tag{3.2}$$

2. Потім дослідження варто продовжувати навколо точки $P_1\left(P_i\right)$.

3. Якщо найменше значення на кроці ***В2*** менше за значення в базовій точці b_2 (загалом b_{i+1}), то одержують нову базову точку $b_2\left(b_{i+2}\right)$, після чого варто повторити крок ***В1***. В іншому разі не потрібно робити пошук за зразком із точки $b_2\left(b_{i+1}\right)$, а продовжити дослідження в точці $b_2\left(b_{i+1}\right)$.

Г

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

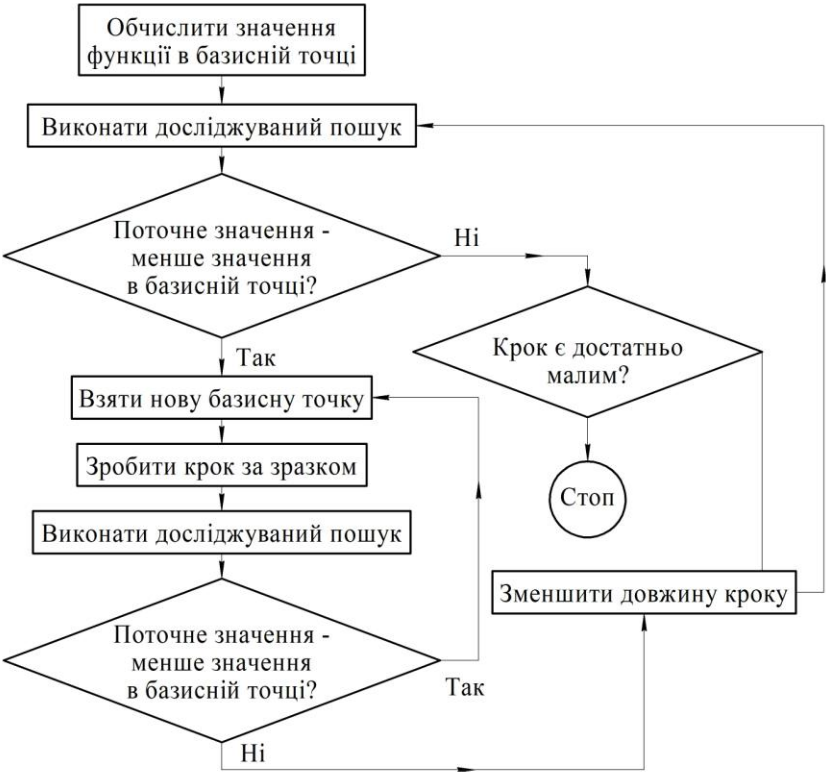

Рисунок 3.2 – Блок-схема методу Хука – Дживса

Рисунок 3.3 – Досліджуваний пошук у методі Хука – Дживса

Ітерація 1

Б

1. Обчислимо значення функції в початковій точці b_1

$$f\left(b_1\right)=\left(4-2\right)^2+\left(\left(-2\right)-5\right)^2+\left(3+2\right)^4=678.$$

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

$$f\left(4+1,-2,3\right)=f\left(5,-2,3\right)=\left(5-2\right)^2+\left(\left(-2\right)-5\right)^2+\left(3+2\right)^4=$$
$$=683>f\left(b_1\right);$$
$$f\left(4-1,-2,3\right)=f\left(3,-2,3\right)=\left(3-2\right)^2+\left(\left(-2\right)-5\right)^2+\left(3+2\right)^4=$$
$$=675>f\left(b_1\right).$$

Продовжимо пошук від нової точки $b_1^*=b_1-h_1e_1$, тобто $\left(3,-2,3\right)$.

$$f\left(3,-2+1,3\right)=f\left(3,-1,3\right)=\left(3-2\right)^2+\left(\left(-1\right)-5\right)^2+\left(3+2\right)^4=$$
$$=662>f\left(b_1^*\right);$$
$$f\left(3,-2-1,3\right)=f\left(3,-3,3\right)=\left(3-2\right)^2+\left(\left(-3\right)-5\right)^2+\left(3+2\right)^4=$$
$$=690>f\left(b_1^*\right).$$

Продовжимо пошук від нової точки $b_1^*=b_1^*+h_2e_2$, тобто $\left(3,-1,3\right)$.

$$f\left(3,-1,3+1\right)=f\left(3,-1,4\right)=\left(3-2\right)^2+\left(\left(-1\right)-5\right)^2+\left(4+2\right)^4=$$
$$=1333>f\left(b_1^*\right);$$
$$f\left(3,-1,3-1\right)=f\left(3,-1,2\right)=\left(3-2\right)^2+\left(\left(-1\right)-5\right)^2+\left(2+2\right)^4=$$
$$=293>f\left(b_1^*\right).$$

Отримаємо точку $b_2\left(3;-1;2\right)$, у якій було досягнуто найменше значення функції на цьому кроці.

B

Виконуємо пошук за зразком.

За формулою (3.1) отримаємо координати точки $P_1\left(4+2(3-4);-2+2(-1-(-2));3+2(2-3)\right)$. Маємо точку $P_1(2,0,1)$.

Відповідне значення функції

$$f(P_1)=(2-2)^2+(0-5)^2+(1+2)^4=106<f(b_2).$$

Досягнуто зменшення значення функції порівняно з точкою b_2.

Установлюємо нові координати $b_2=P_1$ і далі переходимо на наступну ітерацію.

Ітерація 2

Досліджувальний пошук навколо точки $b_2(2;0;1)$.

$$f(2+1;0;1)=f(3;0;1)=(3-2)^2+(0-5)^2+(1+2)^4=$$
$$=107>f(b_2),$$

$$f(2-1;0;1)=f(1;0;1)=(1-2)^2+(0-5)^2+(1+2)^4=$$
$$=107>f(b_2).$$

Зменшення функції не досягнуто.

$$f(2;0+1;1)=f(2;1;1)=(2-2)^2+(1-5)^2+(1+2)^4=$$
$$=97>f(b_2);$$

$$f(2;0-1;1)=f(2;-1;1)=(2-2)^2+(-1-5)^2+(1+2)^4=$$
$$=117>f(b_2).$$

Продовжуємо пошук від нової точки $b_2^*=b_2^*+h_2e_2$, тобто (2; 1; 1).

$$f(2;1;1+1)=f(2;1;2)=(2-2)^2+(1-5)^2+(2+2)^4=$$
$$=272>f(b_2^*);$$

$$f(2;1;1-1) = f(2;1;0) = (2-2)^2 + (1-5)^2 + (0+2)^4 =$$
$$= 32 > f(b_2^*).$$

Отримуємо точку $b_3(2;1;0)$, у якій було досягнуто найменше значення функції на цьому кроці.

B

Виконуємо пошук за зразком.

За формулою (3.1) отримаємо координати точки $P_2\big(3+2(2-3); -1+2(1-(-1)); 2+2(0-2)\big)$. Маємо точку $P_2(1;3;-2)$.

Відповідне значення функції

$$f(P_2) = (1-2)^2 + (3-5)^2 + (-2+2)^4 = 5 < f(b_3).$$

Досягнуто зменшення значення функції порівняно з точкою b_3.

Установлюємо нові координати $b_3 = P_2$ і далі переходимо на наступну ітерацію.

Для наступних ітерацій наведемо лише основні проміжні результати.

Ітерація 3

Досліджувальний пошук навколо точки $b_3(1;3;-2)$.

Знаходимо точку $b_4(2;4;-2)$, значення функції в якій становить $f(b_4) = 1 < f(b_3)$.

Пошук за зразком: отримуємо точку $P_3(2;7;-4)$, значення функції в якій становить $f(P_3) = 20 > f(b_4)$. Відкидаємо цю точку.

Ітерація 4

Досліджувальний пошук навколо точки $b_4(2;4;-2)$.

Знаходимо точку $b_5(2;5;-2)$, значення функції в якій становить $f(b_5)=0<f(b_4)$.

Пошук за зразком: отримуємо точку $P_4(2;6;-2)$, значення функції в якій становить $f(P_4)=1>f(b_5)$. Відкидаємо цю точку.

Ітерація 5

Досліджувальний пошук навколо точки $b_5(2;5;-2)$.

Використовуючи задану довжину кроку $(h=1)$, не вдається знайти точку, значення функції в якій менше, ніж у b_5.

Переходимо на крок ***Б3***.

Зменшуємо довжину кроку, установлюємо $h=0,5$.

Ітерація 6

Досліджувальний пошук навколо точки $b_5(2;5;-2)$.

Використовуючи задану довжину кроку $(h=0,5)$, не вдається знайти точку, значення функції в якій менше, ніж у b_5.

Зменшуємо довжину кроку, установлюємо $h=0,25$.

Ітерація 7

Досліджуваний пошук навколо точки $b_5(2;5;-2)$.

Використовуючи задану довжину кроку $(h=0,25)$, не вдається знайти точку, значення функції в якій менше, ніж у b_5.

Зменшуємо довжину кроку, установлюємо $h=0,1$.

Ітерація 8

Досліджувальний пошук навколо точки $b_5(2;5;-2)$.

Використовуючи задану довжину кроку $(h=0,1)$, не вдається знайти точку, значення функції в якій менше, ніж у b_5.

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

Отже, мінімум функції (з точністю до заданого значення ε) є в точці (2; 5; –2) та становить 0.

Завдання до підрозділу 3.1.2

Виконати пошук мінімуму або максимуму функції (додаток Б) за методом Хука – Дживса.

3.1.3 Методи випадкового пошуку

У методах випадкового пошуку напрямок пошуку на кожному кроці вибирають випадково. У цьому підрозділі ми розглянемо один такий метод, що має назву адаптивного методу випадкового пошуку [3].

Задають початкову точку x_0. Кожну наступну точку x_{k+1} обчислюють за формулою

$$x_{k+1} = x_k + h_k e_k,$$

де $h_k > 0$ – величина кроку;

e_k – випадковий вектор одиничної довжини, що визначає напрямок пошуку;

k – номер ітерації.

На поточній ітерації за допомогою генерування випадкових векторів e_k отримують точки, що розміщуються на гіперсфері радіуса h_k із центром у точці x_k (рис. 3.4). Якщо значення функції в отриманій точці не менше, ніж у центрі, крок вважають невдалим (точки y_1, y_2 за пошуку з x_0; y_1, y_3 за пошуку з x_1). Якщо число невдалих кроків із поточної точки сягає певного числа *M*, подальший пошук триває з тієї самої точки, але з меншим кроком доти, поки він не стане меншим заздалегідь заданої величини ε. Якщо ж значення функції в отриманій точці менше, ніж у центрі, крок вважають вдалим і в знайденому напрямку роблять збільшений крок, що відіграє роль прискорювального кроку (як за пошуку за методом Хука – Дживса). Якщо водночас значення функції знов менше, ніж у центрі, напрямок вважають вдалим, і подальший пошук триває з цієї точки (точки $z_3 = x_1$ за пошуку з x_0, $z_4 = x_2$ за пошуку з x_1). Якщо ж значення функції стало не менше, ніж у центрі, напрямок вважають невдалим, і пошук триває зі старого центра (у точці y_2 за пошуку з x_1 функція менша, ніж у x_1, а в точці z_2 уже не менша, тому напрямок $\left(z_2 - x_1\right)$ є невдалим).

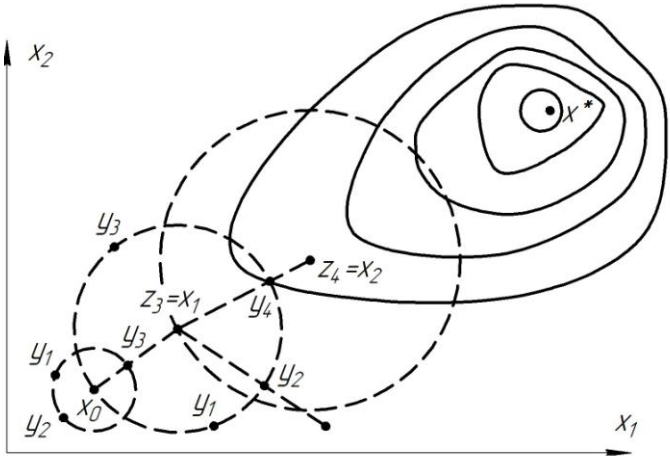

Рисунок 3.4 – Адаптивний метод випадкового пошуку

3.1.4 Метод Недлера – Міда

Перші спроби розв’язання оптимізаційних задач без обмежень на основі прямого пошуку пов’язані з використанням одномірних методів оптимізації. Зазвичай припустиму область визначення цільової функції замінюють дискретною множиною (решіткою) точок простору змінних, а потім використовують 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 ґрунтується 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 разі 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 використовували на кожному етапі. Нелдер і Мід запропонували кілька модифікацій цього методу, які допускають, щоб симплекси були неправильними. Унаслідок цього вийшов дуже надійний метод прямого пошуку, що є одним із найефективніших, якщо $n \le 6$.

Результати окремих чисельних випробувань доводять, 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 $x^{(0)}$ із масштабним множником h. Координати решти n вершин симплекса в n-вимірному просторі обчислюють так:

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

де i та $j = 1, 2, 3, ..., n$;

i – номер точки;

j – індекс змінної.

Прирощення δ_1 та δ_2 залежать лише від n і вибраного масштабного множника h, їх визначають за такими 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 x_h, x_g і x_l.

В

Знайдемо центр тяжіння всіх точок, за винятком точки x_h. Нехай центром тяжіння буде

$$x_0 = \frac{1}{n} \sum_{i \neq h} x_i, \tag{3.3}$$

і обчислимо $f(x_0) = f_0$.

Г

Зручніше за все почати переміщення від точки x_h. Відбивши точку x_h щодо точки x_0, отримаємо нову точку x_r і знайдемо $f_r = f(x_r)$.

Операцію відбиття проілюстровано на рисунку 3.5. Якщо $\alpha > 0$ – це коефіцієнт відбиття, то положення x_r визначено так:

$$x_r - x_0 = \alpha\left(x_0 - x_h\right),$$

тобто

$$x_r = \left(1+\alpha\right)x_0 - \alpha x_h. \qquad (3.4)$$

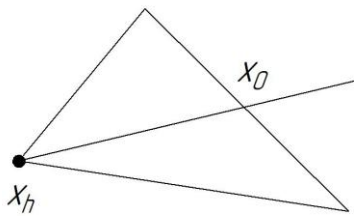

Рисунок 3.5 – Операція відбиття

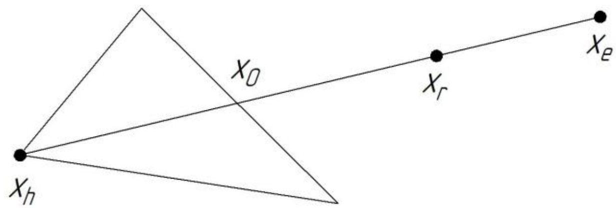

Рисунок 3.6 – Операція розтягування

Д

Порівняємо значення функцій f_r і f_l.

1. Якщо $f_r < f_l$, то ми одержали найменше значення функції. Напрямок із точки x_0 у точку x_r найбільш зручний для переміщення. Отже, ми робимо розтягування в цьому напрямку і знаходимо точку x_e та значення функції $f_e = f\left(x_e\right)$. Рисунок 3.5 ілюструє операцію розтягування симплекса. Коефіцієнт розтягування $\gamma > 1$ можна знайти з таких співвідношень:

$$x_e - x_0 = \gamma\left(x_r - x_0\right),$$

тобто

$$x_e = \gamma x_r + (1-\gamma) x_0. \tag{3.5}$$

Зауваження $\gamma = |x_e - x_0| / |x_r - x_0|$:

а) якщо $f_e < f_l$, то замінюємо точку x_h на точку x_e (відповідно f_h на f_e) і перевіряємо точки симплекса на збіжність до мінімуму (див. крок 1). Якщо збіжність досягнута, то процес зупиняється; в іншому разі повертаємося на крок ***Б*** (обчислюємо нові значення f_h, f_l, f_g і відповідні їм точки x_h, x_l, x_g);

б) якщо $f_e \geq f_l$, то відкидаємо точку x_e. Очевидно, ми перемістилися занадто далеко від точки x_0 до точки x_r. Тому варто замінити точку x_h на точку 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 x_h, x_l, x_g).

2. Якщо $f_r > f_l$, але $f_r \leq f_g$, то x_r є кращою точкою порівняно з іншими двома точками симплекса, і ми замінюємо точку x_h на точку 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 ***Е2***.

Якщо $f_r < f_h$, то замінюємо точку x_h на точку x_r і значення функції f_h на значення функції f_r. Запам’ятовуємо значення $f_r > f_g$ із кроку ***Д2***, наведеного вище. Потім переходимо на крок ***Е2***.

2. У цьому разі $f_r > f_h$ зрозуміло, що ми перемістилися надто далеко від точки x_h до точки x_0. Намагаємося виправити це, знайшовши точку x_c (а потім і f_c), за допомогою кроку стискання, поданого на рисунку 3.7.

Якщо $f_r > f_h$, то відразу переходимо до кроку стискання і знаходимо точку x_c із співвідношення

$$x_c - x_0 = \beta\left(x_h - x_0\right),$$

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

$$x_c = \beta x_h + \left(1 - \beta\right) x_0. \qquad (3.6)$$

Якщо $f_r < f_h$, то спочатку замінимо точку x_h на точку x_r, а потім проведемо стискання (рис. 3.8). Тоді точку x_c знайдемо із співвідношення

$$x_c - x_0 = \beta\left(x_r - x_0\right),$$

тобто

$$x_c = \beta x_r + \left(1 - \beta\right) x_0. \qquad (3.7)$$

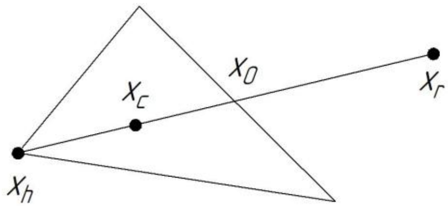

Рисунок 3.7 – Крок стискання для $f_r > f_h$

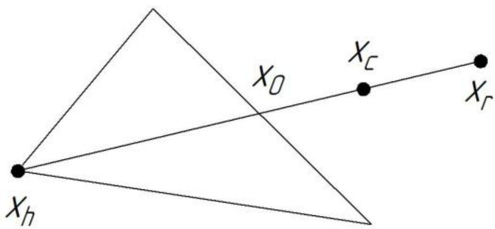

Рисунок 3.8 – Крок стискання для $f_r < f_h$

Ж

Порівнюємо значення функцій f_c і f_h.

1. Якщо $f_c < f_h$, то замінюємо точку x_h на точку 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 x_h, x_l, x_g).

2. Якщо $f_c > f_h$, то, очевидно, що всі наші спроби знайти значення менше ніж 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 x_l – точки, у якій отримано найменше значення функції.

Отже, точка x_i замінюється на точку $x_i + \left(x_l - x_i\right)/2$, тобто заміняємо точку x_i точкою

$$x_i = \left(x_i + x_l\right)/2. \tag{3.8}$$

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

І

Перевірка збіжності ґрунтується на тому, щоб стандартне відхилення $\sigma(n+1)$-го значення функції було менше за деяке задане мале значення *ε*. У цьому разі обчислюють

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

де $\overline{f} = \sum_{i=1}^{n} f_i / (n+1)$.

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

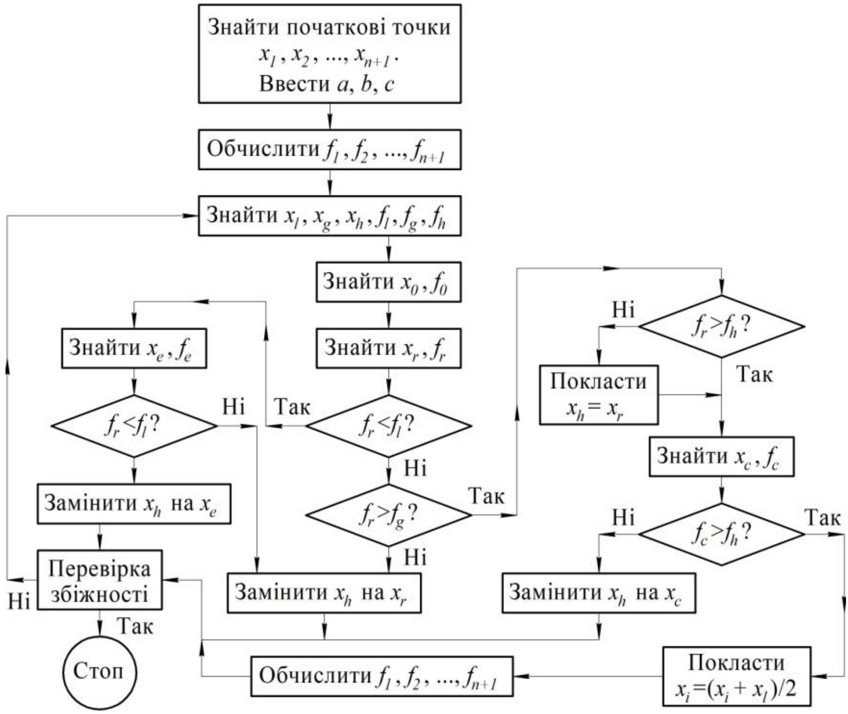

Рисунок 3.9 – Блок-схема методу Нелдера – Міда

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 із різними комбінаціями значень. Ці значення параметрів дозволяють методу бути ефективним.

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.

$$x^{(1)} = (0+0,5176;\ 0+1,9318) = (0,5176; 1,9318);$$

$$x^{(2)} = (0+1,9318;\ 0+0,5176) = (1,9318;\ 0,5176).$$

Ітерація 1

Б

Обчислюємо значення функції в точках початкового симплексу.

$$f\left(x^{(0)}\right) = 5;\quad f\left(x^{(1)}\right) = 0,2374;\quad f\left(x^{(2)}\right) = 3,0658.$$

Отже, $x^{(0)} = x_h$ – найбільше значення функції, $x^{(1)} = x_g$ – наступне за найбільшим значення функції, $x^{(2)} = x_l$ – найменше значення функції.

В

Знайдемо за формулою (3.3) центр тяжіння x_0 усіх точок, за винятком точки x_h

$$x_0 = \frac{1}{n}\sum_{i \neq h} x_i = \frac{1}{2}\left(x_g + x_l\right) = \left(\frac{0,5176+1,9318}{2};\frac{1,9318+0,5176}{2}\right) =$$

$$= (1,2247; 1,2247).$$

Обчислимо

$$f(x_0) = f_0 = (1-1,2247)^2 + (1-1,2247)^2 = 0,6516.$$

Г

Зручніше за все почати переміщення від точки x_h. Відбивши точку x_h щодо точки x_0, одержимо нову точку x_r і знайдемо $f_r = f\left(x_r\right)$ (формула (3.4)).

$$x_r = \left(1+\alpha\right)x_0 - \alpha x_h = \left(1+1\right)x_0 - x_h = \left(2\cdot 1,2247 - 0;\ 2\cdot 1,2247 - 0\right) =$$
$$= \left(2,4494;\ 2,4494\right).$$

Обчислимо

$$f\left(x_r\right) = f_r = \left(1-2,4494\right)^2 + \left(1-2,4494\right)^2 = 2,3027.$$

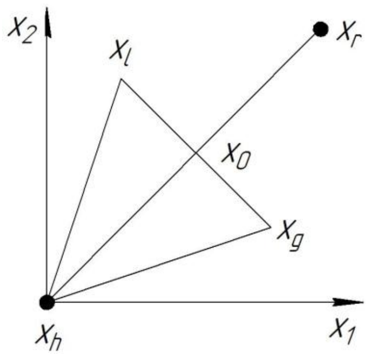

Рисунок 3.10 – Ітерація 1, операція розтягування

Маємо $f_r > f_l$, проте $f_r < f_g$, тобто x_r є кращою точкою порівняно з двома іншими точками, і ми замінюємо точку x_h на точку x_r.

І

Перевіряємо збіжність за формулою (3.9)

$$\overline{f} = \sum_{i=1}^{n} f_i / \left(n+1\right) = \left(0,2374 + 3,0658 + 2,3027\right) / \left(2+1\right) = 1,8686.$$

$$\sigma^2 = \sum_{i=1}^{n}\left(f_i - \overline{f} \right)/\left(n+1\right) =$$

$$= \frac{\left(0,2374-1,8686\right)^2 + \left(3,0658-1,8686\right)^2 + \left(2,3027-1,8686\right)^2}{2+1} =$$

$$= 1,4275;$$

$$\sigma = 1,1948 > \varepsilon.$$

Оскільки збіжності не досягнуто, починаємо другу ітерацію і переходимо на крок ***Б***. Для стислості наведемо на другій і наступних ітераціях лише основні результати.

Ітерація 2

$$x_l\left(0,5176; 1,9318\right); \quad f_l = 0,2374;$$

$$x_g\left(2,4494; 2,4494\right); \quad f_g = 2,3027;$$

$$x_h\left(1,9318; 0,5176\right); \quad f_h = 3,0658.$$

Виконавши операцію відбиття, отримуємо точку $x_r\left(1,0352; 3,8636\right); \quad f_r = 3,4742$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок ***Е***.

Виконавши операцію стискання, отримуємо точку $x_c\left(1,7077; 1,3541\right); \quad f_c = 0,9180$.

Перевірка збіжності $\sigma = 0,859 > \varepsilon$, збіжності не досягнуто.

Ітерація 3

$$x_l\left(0,5176; 1,9318\right); \quad f_l = 0,2374;$$

$$x_g\left(1,7077; 1,3541\right); \quad f_g = 0,9180;$$

$$x_h\left(2,4494; 2,4494\right); \quad f_h = 2,3027.$$

Виконавши операцію відбиття, отримуємо точку $x_r\ \left(-0,224; 0,8366\right); \quad f_r = 2,8517$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок ***Е***.

Виконавши операцію стискання, отримуємо точку $x_c\left(1,7810; 2,0462\right); \quad f_c = 0,6121$.

Перевірка збіжності $\sigma = 0,278 > \varepsilon,$ збіжності не досягнуто.

Ітерація 4

$$x_l\left(0,5176; 1,9318\right); \quad f_l = 0,2374;$$

$$x_g\left(1,7810; 2,0462\right); \quad f_g = 0,6121;$$

$$x_h\left(1,7077; 1,3541\right); \quad f_h = 0,9180.$$

Виконавши операцію відбиття, отримуємо точку $x_r\left(0,5909; 2,6239\right);$ $f_r = 0,5566.$ Маємо $f_r > f_l$ та $f_r < f_g$, тобто точка x_r є кращою порівняно з двома іншими точками симплекса, і ми замінюємо точку x_h на x_r.

Перевірка збіжності $\sigma = 0,278 > \varepsilon,$ збіжності не досягнуто.

Ітерація 5

$$x_l\left(0,5176; 1,9318\right); \quad f_l = 0,2374;$$

$$x_g\left(0,5909; 2,6239\right); \quad f_g = 0,5566;$$

$$x_h\left(1,7810; 2,0462\right); \quad f_h = 0,6121.$$

Виконавши операцію відбиття, отримуємо точку $x_r\ \left(-0,6725; 2,5095\right);$ $f_r = 3,0569.$ Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок ***E***.

Виконавши операцію стискання, отримуємо точку $x_c\left(1,1676; 2,1620\right);$ $f_c = 0,0543.$

Перевірка збіжності $\sigma = 0,2075 > \varepsilon,$ збіжності не досягнуто.

Ітерація 6

$$x_l\left(1,1676; 2,1620\right); \quad f_l = 0,0543;$$

$$x_g\left(0,5176; 1,9318\right); \quad f_g = 0,2374;$$

$$x_h\left(0,5909; 2,6239\right); \quad f_h = 0,5566.$$

Виконавши операцію відбиття, отримуємо точку $x_r\ (1{,}0943; 1{,}4699)$; $f_r = 0{,}29$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок ***E***.

Виконавши операцію стискання, отримуємо точку $x_c\,(0{,}9685; 1{,}7584)$; $f_c = 0{,}0594$.

Перевірка збіжності $\sigma = 0{,}085 > \varepsilon$, збіжності не досягнуто.

***Ітерація** 7*

$$x_l\,(1{,}1676;\ 2{,}1620);\quad f_l = 0{,}0543;$$

$$x_g\,(0{,}9685; 1{,}7584);\quad f_g = 0{,}0594;$$

$$x_h\,(0{,}5176; 1{,}9318);\quad f_h = 0{,}2374.$$

Виконавши операцію відбиття, отримуємо точку $x_r\ (1{,}6185; 1{,}9886)$; $f_r = 0{,}3827$. Оскільки $f_r > f_l$ та $f_r > f_g$, переходимо на крок ***E***.

Виконавши операцію стискання, отримуємо точку $x_c\,(0{,}7928; 1{,}946)$; $f_c = 0{,}0458$.

Перевірка збіжності $\sigma = 0{,}0097 < \varepsilon$, збіжності досягнуто.

Отже, мінімум функції (з точністю до заданого значення ε) є в точці (0,7928; 1,946) та становить 0,0458.

Завдання до підрозділу 3.1.4

Виконати пошук мінімуму або максимуму функції (див. додаток Б) за методом Нелдера – Міда.

3.2 Градієнтні методи. Метод найшвидшого спуску

За допомогою розглянутого в попередньому підрозділі методу покоординатного спуску здійснюють пошук із заданої точки в напрямку, паралельному одній із осей, до точки мінімуму в цьому напрямку. Далі пошук проводять у напрямку, паралельному другій осі тощо. Напрямки, звичайно, фіксовані. Є доцільним модифікувати цей метод так, щоб на кожному етапі пошук точки мінімуму здійснювався вздовж «найкращого» напрямку.

Не зрозуміло, який напрямок є «найкращим», але відомо, що напрямок градієнта $\mathrm{grad}\, f(x)$ є напрямком найшвидшого зростання функції $f(x)$. Отже, протилежний напрямок є напрямком найшвидшого убування функції.

Метод оптимізації, у якому довжину кроку λ вибирають 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 $f(x_1, x_2, ..., x_n)$ починають із заданої точки x_0 із координатами $(x_1, x_2, ..., x_n)_0$. Визначають початкове значення функції $f(x_0)$ (вважаючи $\lambda = 0$).

Б

Визначити часткові похідні функції (проєкції градієнта на координатні напрямки) $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}$.

Взяти за напрямок пошуку напрямок

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

Обчислити d у початковій точці $(x_1, x_2, ..., x_n)_0$, тобто отримати d_0.

В

Знайти значення λ_i, що мінімізує функцію

$$\varphi(\lambda_i) = f(x_i + \lambda_i d_i),$$

де i – номер ітерації.

Для пошуку мінімуму функції $\varphi(\lambda_i)$ у напрямі d_i із точки x_i можна використовувати методи одновимірного пошуку. Як доводить досвід, гарні результати дає застосування методу квадратичної інтерполяції. Довжину кроку λ_i вибирають так, щоб крок «перекрив» мінімум функції $\varphi(\lambda)$. Умова «перекриття» мінімуму виконана, якщо $\varphi(\lambda) \geq \varphi(0)$.

Якщо мінімум не потрапив у відрізок (0, λ), то λ подвоюється, і це повторюється стільки разів, скільки необхідно для виконання умови «перекриття».

Переконавшись, що відрізок (0, λ) 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 із великою точністю, зазвичай, достатньо досягти убування функції $\varphi(\lambda)$. Для цього необхідно виконати 2–3 квадратичних ітерації.

Г

Присвоїти $x_{i+1} = x_i$, і якщо $\left|\text{grad } f\left(x_{i+1}\right)\right|$ є достатньо малим, то процес закінчується. В іншому разі потрібно повернутися на крок ***Б***. Зважаючи, що в процесі пошуку відбувається наближення до екстремуму, для підвищення ефективності процедури доцільно після кожної ітерації зменшувати довжину кроку λ.

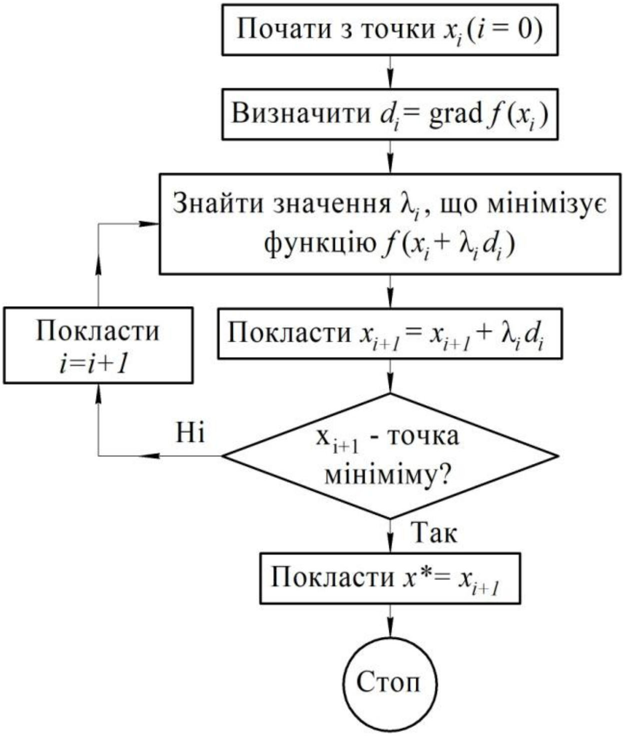

Рисунок 3.11 – Блок-схема методу найшвидшого спуску

Потрібно відзначити, що метод найшвидшого спуску не рекомендований як серйозна оптимізаційна процедура. На перший погляд він є привабливим, проте для практичного застосування «працює» надто повільно. Річ у тім, що властивість найшвидшого спуску є лише локальною властивістю, і тому в деяких випадках доводиться часто змінювати напрям пошуку, що й призводить у підсумку до неефективної обчислювальної процедури. Метод найшвидшого спуску не використовує другі похідні цільової функції. У 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 їхній розгляд не внесено до цього курсу.

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 x_0

$$f(x_0)=(4-1)^2+((-1)-3)^2+4(2+5)^2=221.$$

Б

Похідні функції знайдемо аналітично.

$$\frac{\partial f}{\partial x_1}=2x_1-2;\ \ \frac{\partial f}{\partial x_2}=2x_2-6;\ \ \frac{\partial f}{\partial x_3}=8x_3+40.$$

Отримаємо вираз для градієнта функції.

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

У початковій точці

$$d_0=(-2\cdot 4+2)i+(-2\cdot(-1)+6)j+(-8\cdot 2-40)k=$$

$$=-6i+8j-56k.$$

$$|d_0|=|-\text{grad}\ f(x_0)|=\sqrt{6^2+8^2+56^2}=56{,}89.$$

В

Отримаємо вираз для $\varphi(\lambda)\ f(x+\lambda d)$.

$$\varphi(\lambda)=(x_1+\lambda(-2x_1+2)-1)^2+(x_2+\lambda(-2x_2+6)-3)^2+$$

$$+4(x_3+\lambda(-8x_3-40)+5)^2.$$

У точці x_0 маємо

$$\varphi(\lambda)=(4+\lambda(-2\cdot 4+2)-1)^2+(-1+\lambda((-2)\cdot(-1)+6)-3)^2+$$

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

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

$$\varphi(4) = (3 - 6\cdot 4)^2 + (-4 + 8\cdot 4)^2 + 4(7 - 56\cdot 4)^2 = 48\,314;$$

$$\varphi(2) = (3 - 6\cdot 2)^2 + (-4 + 8\cdot 2)^2 + 4(7 - 56\cdot 2)^2 = 11\,250;$$

$$\varphi(0) = (3 - 6\cdot 0)^2 + (-4 + 8\cdot 0)^2 + 4(7 - 56\cdot 0)^2 = 221.$$

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

Внутрішня ітерація 1

Вважаючи $a = 0,$ $b = 2,$ $c = 4,$ $f_a = 221,$ $f_b = 11\,250,$ $f_c = 48\,314,$ скористаємося формулою (2.13).

$$\delta = \frac{1}{2}\left[\frac{(2^2 - 4^2)\cdot 221 + (4^2 - 0^2)\cdot 11\,250 + (0^2 - 2^2)\cdot 48\,314}{(2-4)\cdot 221 + (4-0)\cdot 11\,250 + (0-2)\cdot 48\,314}\right] = 0{,}1528.$$

Відповідне значення функції $\varphi(0{,}1528) = 21{,}72.$

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

Внутрішня ітерація 2

Вважаючи $a = 0,$ $b = 0{,}1528,$ $c = 2,$ $f_a = 221,$ $f_b = 21{,}72,$ $f_c = 11\,250,$ 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 трьох ітерацій звичайно достатньо. Установлюємо $\lambda_{\min} = 0{,}1280.$

Г

Маємо координати нової точки $x_1\left(x_1, x_2, x_3\right)$:

$$x_1 = x_1 + \lambda_{\min}\left(-2x_1 + 2\right) = 4 + 0{,}128 \cdot \left(-2 \cdot 4 + 2\right) = 3{,}232;$$

$$x_2 = x_2 + \lambda_{\min}\left(-2x_2 + 6\right) = -1 + 0{,}128 \cdot \left(-2 \cdot \left(-1\right) + 6\right) = 0{,}024;$$

$$x_3 = x_3 + \lambda_{\min}\left(-8x_3 - 40\right) = 2 + 0{,}128 \cdot \left(-8 \cdot 2 - 40\right) = -5{,}166.$$

Обчислимо значення функції в новій точці x_1 та значення градієнта в цій точці

$$f\left(x_1\right) = \left(3{,}232 - 1\right)^2 + \left(0{,}024 - 3\right)^2 + 4\left(-5{,}166 + 5\right)^2 = 13{,}96.$$

$$\left|d_1\right| = \left|-\text{grad } f\left(x_1\right)\right| =$$

$$= \sqrt{\left(-2 \cdot 3{,}232 + 2\right)^2 + \left(-2 \cdot 0{,}024 + 6\right)^2 + \left(-8 \cdot \left(-5{,}166\right) - 40\right)^2} = 7{,}56.$$

Відмітимо, що значення і $f\left(x_1\right)$, і $\left|d_1\right|$ стали значно менші, ніж вони були в початковій точці. Усе ж таки вони ще є досить великими, і пошук необхідно продовжувати. Для наступних ітерацій наведемо лише основні проміжні результати.

Ітерація 2

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

Після застосування методу квадратичної інтерполяції (одна внутрішня ітерація) отримуємо значення $\lambda_{\min} = 0{,}4576$ та відповідне $\varphi\left(\lambda_{\min}\right) = 0{,}88.$

Маємо координати нової точки x_2 (1,189; 2,747; –4,558).

Отримаємо значення функції в новій точці x_2 та значення градієнта в цій точці

$$f\left(x_2\right) = \left(1{,}189 - 1\right)^2 + \left(2{,}747 - 3\right)^2 + 4\left(-4{,}558 + 5\right)^2 = 0{,}88.$$

$$\left|d_2\right| = \left|-\text{grad } f\left(x_2\right)\right| =$$

$$= \sqrt{\left(-2 \cdot 1{,}189 + 2\right)^2 + \left(-2 \cdot 2{,}747 + 6\right)^2 + \left(-8 \cdot \left(-4{,}558\right) - 40\right)^2} = 3{,}59.$$

Ітерація 3

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

Після застосування методу квадратичної інтерполяції (одна внутрішня ітерація) отримуємо значення $\lambda_{\min} = 0{,}1280$ та відповідне $\varphi(\lambda_{\min}) = 0{,}056$.

Маємо координати нової точки x_3 (1,141; 2,812; –5,011).

Отримаємо значення функції в новій точці x_3 та значення градієнта в цій точці

$$f(x_3) = (1{,}141 - 1)^2 + (2{,}812 - 3)^2 + 4(-5{,}011 + 5)^2 = 0{,}06.$$

$$|d_3| = |-\mathrm{grad}\ f(x_3)| =$$

$$= \sqrt{(-2 \cdot 1{,}141 + 2)^2 + (-2 \cdot 2{,}812 + 6)^2 + (-8 \cdot (-5{,}011) - 40)^2} = 0{,}48.$$

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

Уже після трьох ітерацій маємо точку (1,141; 2,812; –5,011), значення функції в якій становить 0,06, що майже збігається зі справжнім мінімумом функції. Справжній мінімум цієї функції $f(x)$ є в точці (1; 3; –5), у чому можна переконатися, прирівнявши до нуля перші похідні функції (x). Значення функції (x) у цій точці дорівнює нулю.

Завдання до підрозділу 3.2

Виконати пошук мінімуму або максимуму функції (див. додаток Б) за методом найшвидшого спуску.

РОЗДІЛ 4
ОПТИМІЗАЦІЯ ЗА НАЯВНОСТІ ОБМЕЖЕНЬ

4.1 Загальна теорія

4.1.1 Обмеження у вигляді рівностей

У багатьох задачах на пошук найбільших і 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 називають умовним.

За допомогою методу множників Лагранжа, по суті, установлюються необхідні умови, що дозволяють ідентифікувати точки оптимуму в задачах оптимізації з обмеженнями у вигляді рівностей. Водночас задача з обмеженнями перетворюється в еквівалентну задачу безумовної оптимізації, у 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

$$g(x, y) = 0. \tag{4.1}$$

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

$$\frac{dy}{dx} = \frac{d}{dx}h(x) = -\left.\frac{\partial g}{\partial x}\middle/\frac{\partial g}{\partial y}\right. . \tag{4.2}$$

Тоді функцію

$$z = f(x, h(x)) \tag{4.3}$$

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

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

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

Цей результат може бути поданий у такій формі. Якщо покласти

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

за умови $x = x^*$, $y = y^*$, 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 випливає безпосередньо із співвідношення (4.5).

Отримати ці три необхідні умови можна, використовуючи *функцію Лагранжа*, записану в 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 $f(x, y)$ за наявності обмежень можуть бути записані в такому вигляді:

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

Це система трьох рівнянь, розв'язком якої є значення x^*, y^* та λ^* – у 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

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

де на змінну x накладені обмеження

$$g_1(x) = 0, \; g_2(x) = 0, \; ..., \; g_m(x) = 0. \tag{4.8}$$

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

Тоді з точністю до першого порядку h_j матимемо

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

де $\sum_{j=1}^{n} h_j \frac{\partial g_i}{\partial x_j} = 0$ за умови $i = 1, 2, ..., m.$

Цю умову можна записати інакше

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

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

Звідси випливає

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

Отже, якщо визначити функцію Лагранжа у вигляді

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

то необхідні умови мінімуму функції $f(x)$ за наявності обмежень можна записати так:

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

$$\frac{\partial F}{\partial \lambda_i} = g_i(x) = 0 \text{ за умови } i = 1, 2, ..., m. \tag{4.12}$$

Розв’язання цієї розширеної системи, яка складається з $n+m$ рівнянь, що містять $n+m$ невідомих, визначає стаціонарну точку функції $F(x, \lambda)$. Далі реалізують процедуру перевірки на мінімум або максимум, що проводять на основі обчислення елементів матриці Гессе функції $F(x, \lambda)$.

Приклад

Знайти мінімум функції $f\left(x, y\right)=x^2+y^2$ за умови обмеження $x+y=4$.

Функція Лагранжа набуде вигляду

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

Відповідні умови мінімуму можна записати так:

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

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

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

Розв'язанням цієї системи рівнянь є $x=y=2,\ \ \lambda=4.$ Матриця Гессе функції *F* має $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ 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 $f\left(x\right)$ за наявності *m* обмежень $g_i\left(x\right)\le b_i\ \ \left(i=1, 2, ..., m\right)$.

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

$$g_i\left(x\right)+u_i^2=b_i,$$

або

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

Отже, задача зводиться до мінімізації функції $f\left(x\right)$ за наявності *m* обмежень у вигляді рівності $g_i\left(x\right)+u_i^2-b_i=0.$ Відповідно до викладеного в попередньому підрозділі методу сформуємо функцію Лагранжа

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

У стаціонарній точці мають виконуватися такі умови:

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

$$\frac{\partial F}{\partial \lambda_i} = g(x) + u_i^2 - b_i = 0 \text{ за умови } i = 1, 2, ..., m, \tag{4.16}$$

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

Помноживши останнє рівняння на $\frac{u_i}{2}$, отримуємо $\lambda_i u_i^2 = 0$, тобто

$$\lambda_i \left[b_i - g_i(x) \right] = 0 \text{ за умови } i = 1, 2, ..., m. \tag{4.18}$$

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

Отже, необхідні умови мінімуму функції $f(x)$ за наявності обмежень $g_i(x) \leq b_i$ $(i = 1, 2, ..., m)$ мають вигляд

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

$$g_i(x) \leq b_i \text{ за умови } i = 1, 2, ..., m,$$

$$\lambda_i \left[b_i - g_i(x) \right] = 0 \text{ за умови } i = 1, 2, ..., m,$$

$$\lambda_i \geq 0 \text{ за умови } i = 1, 2, ..., m. \tag{4.19}$$

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 рисунку 4.1. На рисунку зображені лінії постійного рівня функції. Можливо, що наявність обмежень буде призводити до появи локального мінімуму. Це може відбутися навіть у разі, коли функція має лише одну точку мінімуму за відсутності обмежень. Таку ситуацію ілюструє рисунок 4.2.

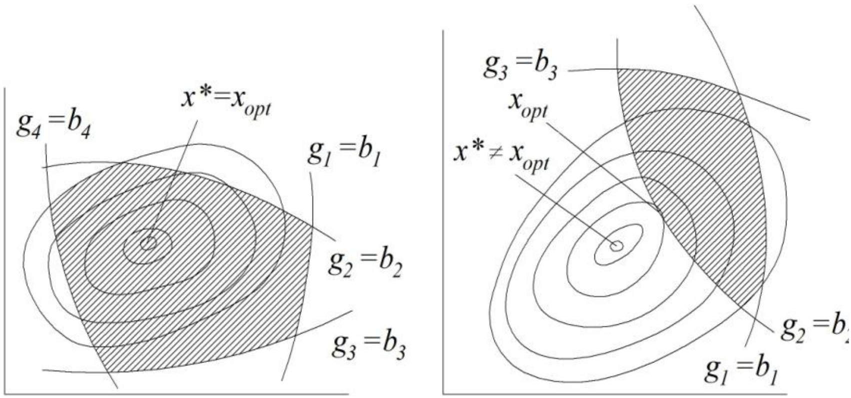

Рисунок 4.1 – Функція та області обмеження

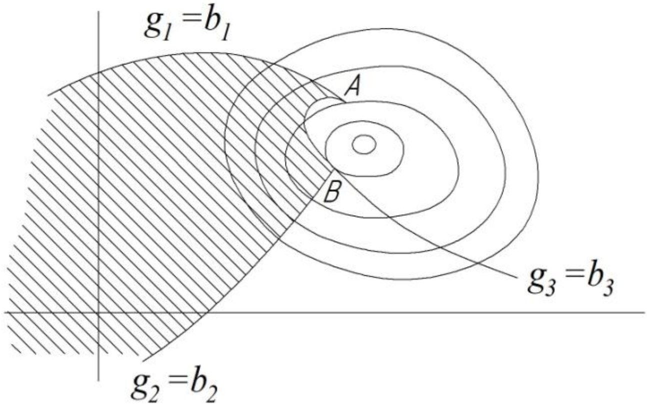

Рисунок 4.2 – Функція та області обмеження.
Локальний мінімум у точці *A*

Приклад

Записати умови Куна – Такера для мінімуму функції $f\left(x\right)=3x_1^2+4x_1x_2+5x_2^2$ за обмежень $x_1\ge 0,\;\; x_2\ge 0$ та $x_1+x_2\ge 4.$

Цю задачу можна вирішити так: мінімізувати функцію $f\left(x\right)=3x_1^2+4x_1x_2+5x_2^2$ за обмежень $-x_1\le 0,\;\; -x_2\le 0$ та $-x_1-x_2\le 4.$ Функція Лагранжа $F\left(x,\lambda,u\right)$ матиме вигляд

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

Необхідні умови мінімуму такі:

$$6x_1+4x_2-\lambda_1-\lambda_3=0,$$
$$4x_1+10x_2-\lambda_2-\lambda_3=0,$$
$$-x_1\le 0,\;\; -x_2\le 0\;\; -x_1-x_2\le 4,$$
$$\lambda_1x_1=0,\;\; \lambda_2x_2=0,\;\; \lambda_3\left(4-x_1-x_2\right)=0,$$
$$\lambda_1,\lambda_2,\lambda_3\ge 0.$$

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

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

Мінімум функції $f\left(x\right)$ за відсутності обмежень дорівнює нулю та є на початку координат. Область обмежень на рисунку 4.3 заштрихована.

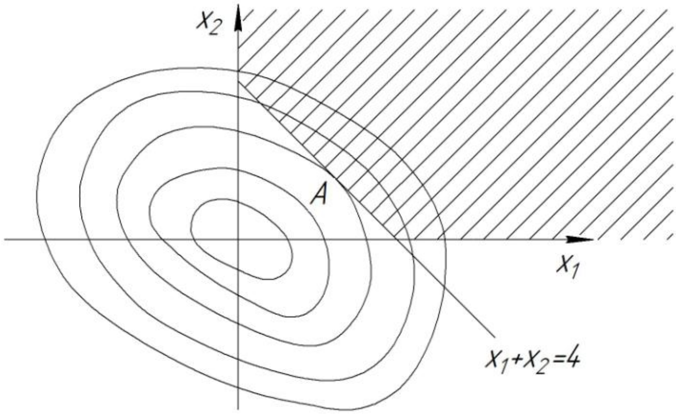

Рисунок 4.3 – Приклад функції та області обмеження

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 під час розв'язання 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 способом. Якщо ні, то цільовій функції привласнюють дуже велике значення. Отже, пошук буде здійснюватися знов у допустимій області в напрямку до мінімальної точки всередині цієї області.

На жаль, загалом такий підхід є недосконалим. За допомогою такого підходу неможливо просуватися вздовж границі області обмежень, оскільки збіжність досягається в першій точці границі. Загальна задача оптимізації за 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 $z = f(x)$ із відповідними обмеженнями, накладеними на *x*, у задачу пошуку мінімуму *без обмежень* функції

$$Z = f(x) + P(x).$$

Функція $P(x)$ є штрафною. Необхідно, щоб за умови порушення обмежень, вона штрафувала функцію *Z*, тобто збільшувала її значення. У цьому разі мінімум *Z* буде всередині області обмежень. Функція $P(x)$, що задовольняє цю умову, може бути не єдиною.

Задачу мінімізації можна сформулювати так: мінімізувати функцію $z = f(x)$ за обмежень $c_j(x) > 0, \quad j = 1, 2, ..., m$ [2].

Зауваження

Обмеження вигляду «менше або дорівнює», $h(x) \le 0$, завжди може бути записане як «більше або дорівнює», – $h(x) \ge 0$, тому в наведеному формулюванні немає втрати загальності.

Функцію $P(x)$ зручно записати так:

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

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

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

Якщо *x* набуває допустимих значень, тобто значень, для яких $c_j(x) \ge 0$, то *Z* набуває значення, більші за відповідні значення $f(x)$ (справжньої цільової функції цієї задачі), і різницю можна зменшити завдяки тому, що r може бути дуже малою величиною. Але якщо x набуває значень, що є хоча і допустимими, проте близькими до границі області обмеження, і щонайменше одна з функцій $c_j(x)$ близька до нуля, то значення функції $P(x)$ і, отже, значення функції Z стають дуже великими. Отже, вплив функції $P(x)$ полягає у створенні «гребню з крутими краями» вздовж кожної границі області обмежень. Так, якщо пошук починається з допустимої точки і здійснюється пошук мінімуму функції $\varphi(x, r)$ без обмежень, то мінімум, звичайно, буде досягатися всередині допустимої області для задачі з обмеженнями. Вважаючи r достатньо малою величиною для того, щоб вплив $P(x)$ був малим у точці мінімуму, ми можемо зробити точку мінімуму функції $\varphi(x, r)$ без обмежень такою, що збігається з точкою мінімуму функції $f(x)$ з обмеженнями.

Приклад

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

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

На рисунку 4.4 зображено графік функції $\varphi(x, r)$ та подано 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 за наявності обмежень. Дійсно, знайдемо першу похідну функції $\varphi(x, r)$

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

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

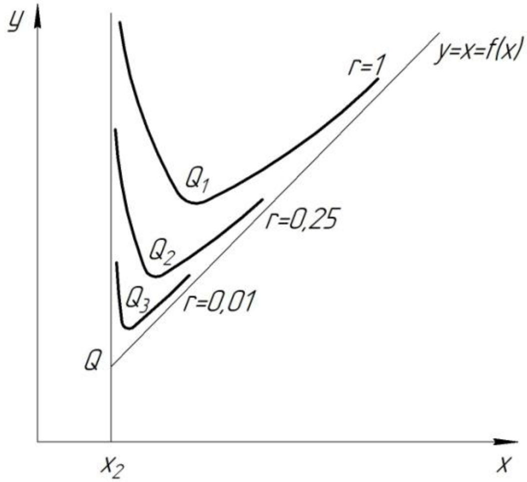

Рисунок 4.4 – Приклад застосування штрафної функції

Тоді

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

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

Отже, функція $\varphi(x, r)$ має мінімум, що дорівнює $2 + 2\sqrt{r}$ за умови $x = 2 + \sqrt{r}$. 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 за умови $r \to 0$ мінімум без обмежень функції $\varphi(x, r)$ наближається до значення 2, і мінімальною точкою є точка $x = 2$.

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

Розглянемо два приклади, взяті з монографії [1]. Оскільки ці приклади мають ілюстративний характер, з метою запобігання втрати наочності в них збережено систему мір, що використав автор.

5.1 Оптимізація системи сонячного опалення для житлового будинку

У зв'язку з усе більшим 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.

Сьогодні існує 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 навколишнього середовища та процесу використання енергії сонячного випромінювання.

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

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

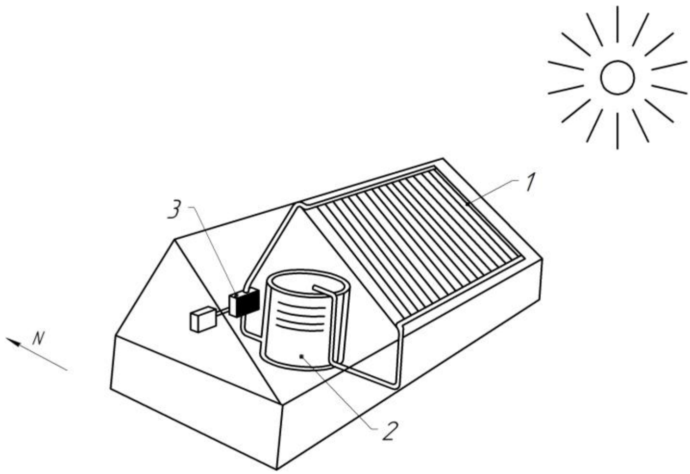

Рисунок 5.1 – Схема будинку із сонячним опаленням:
1 – сонячний колектор; 2 – водяний бак;
3 – регульований радіатор

Характеристики колектора та акумулювальної 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 і дверних прорізах і, звичайно, погодних умов.

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

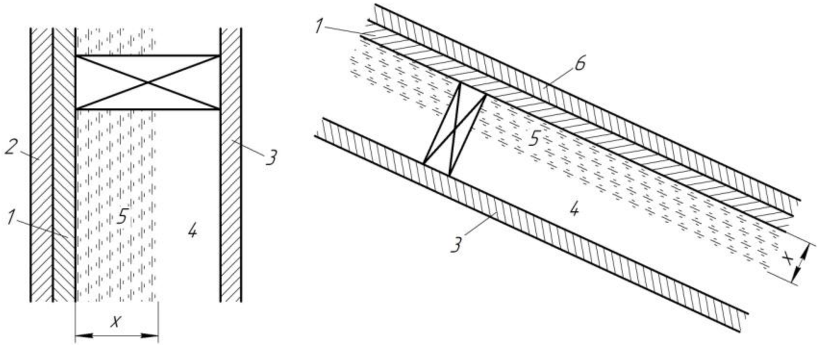

Рисунок 5.2 – Розрізи стіни (ліворуч) та даху (праворуч):
1 – фанера, 1/2 дюйма; 2 – зовнішній щит, 1/2 дюйма;
3 – внутрішній щит, 1/2 дюйма; 4 – повітря; 5 – скловолокно;
6 – покрівельна обшивка

Цільова функція та враховані обмеження наведені нижче.

Розміри будинку (відповідно до ескізу)

Довжина	60 футів
Висота стін	8 футів
Ширина	30 футів
Ширина південного схилу даху	25 футів
Ширина північного схилу даху	15 футів
Троє дверей	7 футів × 3,5 футів
Шість вікон	6 футів × 3 футів

Розрахункові погодні умови

Середня температура приміщення	68 ºF (градієнт у вертикальному напрямку)
Температура зовнішнього повітря	35 ºF
Швидкість вітру	10 миль/год
Коефіцієнт тепловіддачі стін із зовнішнього боку будинку	$h_0 = 6$ БТЕ/год × фут^2 × ºF
Коефіцієнт тепловіддачі стін із внутрішнього боку будинку	$h_i = 1$ БТЕ/год × фут^2 × ºF

Теплоізоляція

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

Ширина простору для скловолокнистої ізоляції:

між зовнішніми та внутрішніми щитами стін між покрівельною обшивкою	3,5 дюйма
та внутрішнім щитом даху (товщина балки)	5,5 дюйма

Втрати тепла внаслідок тепловіддачі від стін

$$\left(T_{вн} - T_{зовн}\right)/RA = \left(33/4{,}5x\right) \text{ (БТЕ/год} \times \text{фут}^2\text{)}.$$

Втрати тепла внаслідок тепловіддачі від даху

$$\left(T_{вн} - T_{зовн}\right)/RA = \left(39/4{,}5x\right) \text{ (БТЕ/год} \times \text{фут}^2\text{)}.$$

Тут x – товщина скловолокнистої ізоляції в дюймах.

Дані для програми оптимізації

Оптимізують вартість системи сонячного опалення будинку з такими параметрами:

Площа стін	2730 футів^2
Площа даху	2400 футів^2 (площа, яку можна використовувати для розміщення колектора, дорівнює 1500 футів^2)
Площа підлоги	1800 футів^2
Площа вікон	108 футів^2
Довжина щілин віконних отворів	108 футів
Довжина щілин дверних отворів	63 фути
Втрати тепла через стіни	$\left(33/4{,}5x\right)$ (БТЕ/год × фут^2)
Втрати тепла через дах	$\left(39/4{,}5x\right)$ (БТЕ/год × фут^2)

Змінні

x_1 – площа поверхні колектора, футів^2;

x_2 – радіус бака, футів;

x_3 – висота бака, футів;

x_4 – товщина ізоляції стін, дюймів;

x_5 – товщина ізоляції даху, дюймів;

x_6 – довжина ущільнювальних прокладок, футів.

Вартість в умовних одиницях

$3,5x_1 + 150$	$+15,7x_2x_3 + 100$
(1) Колектор	(5) Матеріали бічної поверхні бака
$+0,0045x_1^{3/2} + 100$	$+31,4x_2^{1/2}x_3^{1/2} + 100$
(2) Система циркуляції води	(6) Система регулювання бака
$+x_1^{1/2}$	$+137x_4$
(3) Система регулювання колектора	(7) Ізоляція стін
$+6,28x_2^2$	$+72x_5$
(4) Матеріали днищ бака	(8) Ізоляція даху
	$+0,2x_6$
	(9) Ущільнення щілин

Загалом вартість становить (колектор, системи акумулювання, циркуляції, регулювання та ізоляції)

$$P_0 = 450 \text{ умовних одиниць}$$

(незмінні величини із функції вартості P_0 виключаються) =

$$= \underset{(1)}{3,5x_1} + \underset{(2)}{0,0045x_1^{3/2}} + \underset{(3)}{x_1^{1/2}} + \underset{(4)}{6,28x_2^2} + \underset{(5)}{15,7x_2x_3} + \underset{(6)}{31,4x_2^{1/2}x_3^{1/2}} + \underset{(7)}{137x_4} +$$

$$+\underset{(8)}{72x_5} + \underset{(9)}{0,2x_6}$$

Обмеження	Фізичне значення
(1) $x_1 \leq 1\,500$	Розміри колектора обмежені площею південного схилу даху
(2) $x_4 \leq 3{,}5$	Товщина ізоляції стін обмежена шириною простору між зовнішнім і внутрішнім щитами стіни
(3) $x_5 \leq 5{,}5$	Товщина ізоляції даху обмежена товщиною балки
(4) $2x_2x_3 \geq 150$	Площа поверхні бака обмежена умовою природної конвекції
(5) $770x_1 \geq 48 \cdot [2\,730 \cdot \left(33/4{,}5x_4\right) + +2\,400 \cdot \left(39/4{,}5x_5\right) + 1\,800 \cdot \left(1{,}5\right) + +108 \cdot \left(3{,}3/1{,}19\right) - \left(10^2 + 63\right) \times \times \left(0{,}075\right)\left(20\right)\left(0{,}24\right)\left(33\right)x_6]$	Кількість тепла, яке необхідно віддати воді за умови к. к. д. колектора 50 %, для підігріву будинку протягом двох діб
(6) $770x_1 \leq 1{,}2\left(140-68\right)\left(0{,}988\right) \times \times \left(\pi\right)x_2^2x_3$	Підведення тепла за добу = = 1,2 × (теплоємність маси води всередині бака); напрямок знака нерівності визначено вимогою, щоб теплоємність бака була не меншою, ніж добова теплопродуктивність колектора
(7) $x_6 \leq 171$	Довжина ущільнювальних прокладок обмежується периметром віконних і дверних отворів

5.2 Оптимізація лінії електропередачі з проводом кільцевого перерізу

Цей приклад стосується оптимізації високовольтної лінії електропередачі. Зрозуміло, що проєктувальник повинен прагнути до зменшення кількості опор на шляху від електростанції до центру 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, за умови змінного струму найбільша щільність струму припадає на периферійну зону перерізу. Тому центральна зона перерізу провідника може бути видалена без помітного зниження електропровідності. Отже, виявляються два аспекти розв’язання задачі: вибір розмірів поперечного перерізу провідника та визначення оптимальної відстані між опорами.

За заданого значення струму I та за умови лінійної залежності щільності струму від радіуса, $J = kr,$ розміри кільцевого перерізу провідника визначаються із співвідношення $I = 2\pi k\left(r_0^3 - r_i^3\right)/3.$ Беруть, що теоретична функція, що описує криву провисання дротів між опорами, може бути апроксимована параболою $d = wS^2/8t,$ $T = t + wd,$ де w – погонна вага дроту, а інші параметри зазначені на рисунку 5.3.

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

Загальну вартість лінії електропередачі, яка являє собою мінімізовану цільову функцію, записують у вигляді $C_1 mwDl/S + C_2\left(n+1\right).$

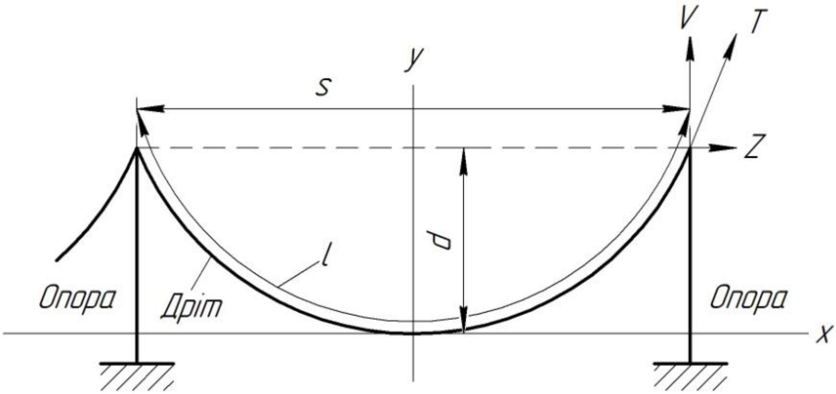

Рисунок 5.3 – Опори та дріт лінії електропередачі

Задані такі сім обмежень:

(1) Величина прогину дроту d не повинна перевищувати 10 футів.

(2) Максимальне напруження розтяжіння в разі використання алюмінієвого сплаву не повинно перевищувати 11 000 футів/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 дюйма.

(6) Максимально припустима товщина стінки кільцевого проводу становить 0,2 дюйма.

(7) Границя міцності матеріалу ізолятора дорівнює 15 000 футів/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

Сьогодні розроблені програмні продукти, що дозволяють розв'язувати оптимізаційні задачі за невеликий відрізок часу і з необов'язковим розумінням методу, який застосовують для розв'язання. Це істотно полегшує їхнє використання для розв'язання прикладних задач. Однак необхідно розібратися в 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 у тому, що Excel дозволяє проводити глибокий аналіз даних і отримувати внаслідок цього 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 розв'язувати 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 із яких може бути автономною програмою.

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 і 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 із розглянутих методів оптимізації в 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 потрібно 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 виконують ітераційно. Після кожної ітерації користувач має можливість побачити проміжні результати пошуку. Після знаходження оптимального рішення Excel може видати звіт за результатами розрахунку.

Завдання до розділу 6

Виконати пошук мінімуму функції (див. додаток В) за допомогою інструменту «Пошук розв’язання» пакета Microsoft Excel.

СПИСОК ЛІТЕРАТУРИ

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

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

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

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

5. Кочевський О. М. Моделі і методи оптимізації : конспект лекцій для студ. спец. 7.000008 «Енергетичний менеджмент» денної та заочної форм навчання / О. М. Кочевський, О. Г. Гусак. – Суми : СумДУ, 2005. – 75 с.

ДОДАТОК А
(обов'язковий)

ВИХІДНІ ДАНІ

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.

Варіант	Функція	Відрізок
1	$f(x)=\ln\left(x^2-2x+2\right)$	$[0;3]$
2	$f(x)=3x/\left(x^{2+1}\right)$	$[0;5]$
3	$f(x)=(2x-1)/(x-1)^2$	$[-0,5;0]$
4	$f(x)=(x+2)e^{1-x}$	$[-2;2]$
5	$f(x)=\ln\left(x^2-2x+4\right)$	$[-1;1,5]$
6	$f(x)=x^3/\left(x^2-x+1\right)$	$[-1;1]$
7	$f(x)=\left((x+1)/x\right)^3$	$[1;2]$
8	$f(x)=\sqrt{x-x^3}$	$[-2;2]$
9	$f(x)=4-e^{-x^2}$	$[0;1]$
10	$f(x)=\left(x^3+4\right)/x^2$	$[1;2]$
11	$f(x)=xe^x$	$[-2;0]$
12	$f(x)=(x-2)e^x$	$[-2;1]$
13	$f(x)=(x-1)e^{-x}$	$[0;3]$
14	$f(x)=x/\left(9-x^2\right)$	$[-2;2]$
15	$f(x)=(1+\ln x)/x$	$[1/e;e]$
16	$f(x)=e^{4x-x^2}$	$[1;3]$

Продовження додатка А

Варіант	Функція	Відрізок
17	$f(x)=\left(x^5-8\right)/x^4$	$[-3;-1]$
18	$f(x)=\dfrac{e^{2x}+1}{e^x}$	$[-1;2]$
19	$f(x)=x\ln x$	$\left[1/e^2;1\right]$
20	$f(x)=x^3e^{x+1}$	$[-4;0]$
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	$f(x)=x^4/4-6x^3+7$	$[16;20]$

ДОДАТОК Б
(обов'язковий)

ВИХІДНІ ДАНІ

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

Варіант	Функція	Точка
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	$f(x, y) = 2xy - 5x^2 - 3y^2 + 2$	$(f_{\max})$

Продовження додатка Б

Варіант	**Функція**	**Точка**
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	$f(x, y) = 2(x + y) - x^2 - y^2$	$(f_{\max})$

ДОДАТОК В
(обов’язковий)

ВИХІДНІ ДАНІ

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

Варіант	**Функція**	***D***
1	$f(x, y) = 3x + y - xy$	$y = x,\ y = 4,\ x = 0$
2	$f(x, y) = xy - x - 2y$	$x = 3,\ y = x,\ y = 0$
3	$f(x, y) = x^2 + 2xy - 4x + 8y$	$x = 0,\ x = 1,$ $y = 0,\ y = 2$
4	$f(x, y) = 5x^2 - 3xy + y^2$	$x = 0,\ x = 1,$ $y = 0,\ y = 1$
5	$f(x, y) = x^2 + 2xy - y^2 - 4x$	$x - y + 1 = 0,$ $x = 3,\ y = 0$
6	$f(x, y) = x^2 + y^2 - 2x - 2y + 8$	$x = 0,\ y = 0,$ $x + y - 1 = 0$
7	$f(x, y) = 2x^3 - xy^2 + y^2$	$x = 0,\ x = 1,$ $y = 0,\ y = 6$
8	$f(x, y) = 3x + 6y - x^2 - xy - y^2$	$x = 0,\ x = 1,$ $y = 0,\ y = 1$
9	$f(x, y) = x^2 - 2y^2 + 4xy - 6x - 1$	$x = 0,\ y = 0,$ $x + y - 3 = 0$
10	$f(x, y) = x^2 + 2xy - 10$	$y = 0,\ y = x^2 - 4$
11	$f(x, y) = xy - 2x - y$	$x = 0,\ x = 3,$ $y = 0,\ y = 4$
12	$f(x, y) = 0{,}5x^2 - xy$	$y = 8,\ y = 2x^2$

Продовження додатка В

Варіант	Функція	*D*
13	$f(x,y)=3x^2+3y^2-2x-2y+2$	$x=0,\ y=0,$ $x+y-1=0$
14	$f(x,y)=2x^2+3y^2+1$	$y=\sqrt{9-\frac{9}{4}x^2},\ y=0$
15	$f(x,y)=x^2-2xy-y^2+4x+1$	$x=-3,\ y=0,$ $x+y+1=0$
16	$f(x,y)=3x^2+3y^2-x-y+1$	$x=5,\ y=0,$ $x-y-1=0$
17	$f(x,y)=2x^2+2xy-0,5y^2-4x$	$y=2x,\ y=2,\ x=0$
18	$f(x,y)=x^2-2xy+2,5y^2-2x$	$x=0,\ x=2,$ $y=0,\ y=2$
19	$f(x,y)=xy-3x-2y$	$x=0,\ x=4,$ $y=0,\ y=4$
20	$f(x,y)=x^2+xy-2$	$y=4x^2-4,\ y=0$
21	$f(x,y)=x^2y(4-x-y)$	$x=0,\ y=0,$ $y=6-x$
22	$f(x,y)=x^3+y^3-3xy$	$x=0,\ x=2,$ $y=-1,\ y=2$
23	$f(x,y)=4(x-y)-x^2-y^2$	$x+2y=4,$ $x-2y=4,\ x=0$
24	$f(x,y)=x^2+2xy-y^2-4x$	$x=3,\ y=0,$ $y=x+1$
25	$f(x,y)=6xy-9x^2-9y^2+$ $+4x+4y$	$x=0,\ x=1,$ $y=0,\ y=2$
26	$f(x,y)=x^2+2xy-y^2-2x+2y$	$y=x+2,\ y=0,$ $x=2$

Продовження додатка В

Варіант	**Функція**	***D***
27	$f\left(x,y\right)=4-2x^2-y^2$	$y=0,\ y=\sqrt{1-x^2}$
28	$f\left(x,y\right)=5x^2-3xy+y^2+4$	$x=-1,\ x=1,$ $y=-1,\ y=1$
29	$f\left(x,y\right)=x^2+2xy+4x-y^2$	$x+y+2=0,$ $x=0,\ y=0$
30	$f\left(x,y\right)=2x^2y-x^3y-x^2y^2$	$x=0,\ y=0,$ $x+y=6$

Електронне навчальне видання

Гусак Олександр Григорович,
Панченко Віталій Олександрович,
Хацко Костянтин Олександрович

МОДЕЛІ І МЕТОДИ ОПТИМІЗАЦІЇ

Конспект лекцій
для студентів спеціальності *144* «*Теплоенергетика*»
(освітня програма **«Енергетичний менеджмент»**)
усіх форм навчання

Відповідальний за випуск М. І. Сотник
Редактор І. О. Кругляк
Комп’ютерне верстання В. О. Панченка

Формат 60×84/16. Ум. друк. арк. 6,28. Обл. вид. арк. 2,23.

Видавець і виготовлювач
Сумський державний університет,
вул. Римського-Корсакова, 2, м. Суми, 40007
Свідоцтво суб’єкта видавничої справи ДК № 3062 від 17.12.2007.