О. М. Кісельова
А.Є. Шевельова

Чисельні методи оптимізації

2008

ЧИСЕЛЬНІ МЕТОДИ ОПТИМІЗАЦІЇ

О. М. Кісельова
А.Є. Шевельова

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДНІПРОПЕТРОВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ОЛЕСЯ ГОНЧАРА

Кафедра обчислювальної математики та математичної кібернетики

О.М. Кісельова, А.Є. Шевельова

Чисельні методи оптимізації

Затверджено Міністерством освіти і науки України як навчальний посібник для студентів вищих навчальних закладів.

Дніпропетровськ
Видавництво Дніпропетровського
національного університету
2008

УДК 519.8

ББК 22.18

К 44

Гриф надано Міністерством освіти і науки України

Лист №1.4/18–Г–2117 від 13.10.08

Рецензенти:

д-р техн. наук, проф. І.В. Бейко

д-р фіз.-мат. наук, проф. В.О. Капустян

д-р техн. наук, проф. О.І. Міхальов

Кісельова О.М.

К 44 Чисельні методи оптимізації: навч. посіб. / О.М. Кісельова, А.Є.Шевельова. – Д.: Вид-во ДНУ, 2008. – 208 с.

ISBN 978-966-551-269-1

Викладено основні теоретичні положення та алгоритми чисельного розв’язання задач неперервної диференційовної та недиференційовної оптимізації.

Наведено основні поняття теорії оптимізації, умови оптимальності, теореми Куна-Такера. Розглянуто методи розв’язання задач одновимірної оптимізації; задач скінченновимірної безумовної та умовної оптимізації; методи недиференційовної безумовної оптимізації. До всіх розділів посібника наведені вправи для самостійної роботи студентів.

Для студентів вищих навчальних закладів галузі «Системні науки і кібернетика», які навчаються за напрямами: «Системний аналіз», «Прикладна математика», а також для науковців, спеціалістів у галузі математики та її застосування.

УДК 519.8

ББК 22.18

ISBN 978-966-551-269-1

Зміст

Передмова

Методи оптимізації широко застосовуються в математиці, економіці, техніці, природознавстві. Тому в наш час математична освіта неможлива без теорії оптимізації.

Теорія оптимізації застосовується в багатьох розділах прикладної математики та системного аналізу, широко використовується при побудові чисельних методів розв’язання екстремальних задач та якісному дослідженні математичних моделей.

Обчислювальні методи розв’язання екстремальних задач останніми роками розвивалися надзвичайно інтенсивно. Розробка нових чисельних методів оптимізації та дослідження, теоретичне обґрунтування існуючих методів відображає ту важливу роль, яку відіграють екстремальні задачі в різноманітних проблемах прикладного характеру.

У першому розділі подані основні відомості з опуклого аналізу, які необхідні при постановці задач, викладенні методів їхнього розв’язання та обґрунтуванні збіжності, поняття теорії оптимізації, умови оптимальності, теорія Куна-Такера. Розділ містить багато прикладів, завдань для самостійного виконання.

Другий – присвячено методам розв’язання задач одновимірної оптимізації, що складають, як відомо, основну складову частину методів нелінійного програмування і частково обумовлює їхню практичну ефективність. Наводяться алгоритми методів, дається геометрична ілюстрація, приклади розв’язання задач.

Третій розділ розглядає задачі багатовимірної безумовної оптимізації та містять основні чисельні методи розв’язання задач оптимізації, які використовують або не використовують похідні цільової функції. Дано обґрунтування збіжності методів, наводяться приклади.

У четвертому розділі викладені методи розв’язання задач умовної оптимізації. А саме: типу проекції, можливих напрямків, лінеаризації, відтинання, штрафних функцій.

Термін «недиференційовна оптимізація» включає широке коло питань, які пов'язані із знаходженням екстремальних значень недиференційовних функцій. Задачі мінімізації або максимізації негладких функцій, природним чином, виникають як при розв'язанні практичних завдань, так і суто математичних проблем.

У п'ятому розділі розглянуті методи недиференційовної оптимізації, які побудовані на інформації про субградієнти та значення функції обчислені за певними точками. Зокрема, алгоритми узагальненого градієнтного спуску, градієнтні методи з розтягом простору в напрямку різниці двох послідовних субградієнтів (R-алгоритм), метод еліпсоїдів та його модифікації.

До всіх розділів посібника наведені задачі та вправи, які дозволяють не тільки закріпити вивчений матеріал, але й містять значну додаткову інформацію.

При написанні навчального посібника була використана наукова навчальна література, список якої наведений в бібліографічних посиланнях. Ця література може бути використана для поглиблення знань як з теоретичних питань, так і з питань практичного застосування методів оптимізації.

Розділ 1. Основні поняття теорії оптимізації, опуклого аналізу, умови оптимальності, поняття двоїстості

§1. Основні поняття теорії оптимізації

Під загальною задачею оптимізації будемо розуміти таку задачу: задані множина X і функція $f(x)$, яка визначена на X. Умовимося записувати задачу на мінімум у вигляді

$$f(x)\to min,\ x\in X. \tag{1.1}$$

При цьому будемо називати $f(x)$ **цільовою функцією**, X – **допустимою множиною**, будь-який елемент $x\in X$ – допустимою точкою задачі (1.1).

У посібнику будуть розглянуті деякі чисельні методи розв'язання скінченновимірних задач оптимізації, тобто задач, допустима множина яких лежить у n-вимірному евклідовому просторі E^n.

Точка x_* називається:

1) точкою **глобального мінімуму** функції $f(x)$ на множині X або глобальним розв'язком задачі (1.1), якщо

$$f(x_*)\le f(x) \text{ для усіх } x\in X. \tag{1.2}$$

2) точкою **локального мінімуму** функції $f(x)$ на множині X або локальним розв'язком задачі (1.1), якщо існує число $\varepsilon>0$ таке, що

$$f(x_*)\le f(x) \text{ при усіх } x\in X\cap U_\varepsilon(x_*), \tag{1.3}$$

де $U_\varepsilon(x_*)=\left\{x\in E^n : \|x-x_*\|\le\varepsilon\right\}$ – куля радіуса $\varepsilon>0$ з центром у точці x_*.

Якщо нерівності (1.2) або (1.3) виконуються як строгі при $x\ne x_*$, то кажуть, що x_* – **точка строгого мінімуму** в глобальному або локальному сенсі.

Ясно, що глобальний розв'язок є і локальним; зворотне твердження є хибним.

Значення функції $f(x_*)$ називається **найменшим (найбільшим) значенням** та позначається як

$$f(x_*)=\min_{x\in X} f(x)\ \left(f(x_*)=\max_{x\in X} f(x)\right).$$

Позначимо через $X_* = \{ x \in X : f(x) = f(x_*) \}$ – **множину всіх точок мінімуму** функції $f(x)$ на X. Залежно від властивостей множини X та функції $f(x)$, множина X_* може складатися з однієї, кількох або безлічі точок, а також можливо, що X_* є порожньою множиною.

Аналогічно з (1.1) будемо записувати задачу максимізації функції $f(x)$ на множині X у вигляді

$$f(x) \to max, \ x \in X. \tag{1.4}$$

Розв'язки задач (1.1), (1.4), точки мінімуму та максимуму функції $f(x)$ на X, називають **точками екстремуму**, а самі задачі (1.1), (1.4) – **екстремальними задачами**.

Ясно, що задача (1.4) еквівалентна задачі

$$-f(x) \to min, \ x \in X$$

в тому сенсі, що відповідні множини глобальних та локальних розв'язків цих задач співпадають. Це дозволяє переносити результати, які отримані для задачі мінімізації до задач максимізації та навпаки. У подальшому, ми будемо розглядати задачу мінімізації.

Приклад. Нехай $f(x) = x^2 + 1$, $x \in [-1,1]$. Такі задачі будуть еквівалентними (рис.1.1):

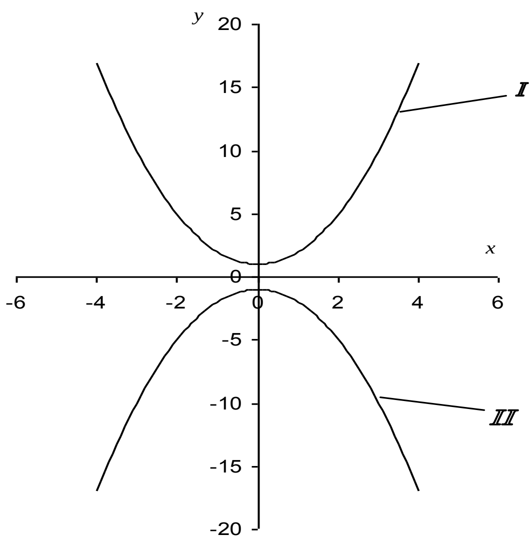


Рис. 1.1

Задача *I*:

$$f(x) \to min, \ x \in [-1,1].$$

Задача *II*:

$$\{-f(x)\} \to max, \ x \in [-1,1].$$

При цьому $\min\limits_{x\in[-1,1]} f(x) = -\max\limits_{x\in[-1,1]} f(x) = 1.$

При дослідженні задач оптимізації, у першу чергу, виникає питання про існування розв'язку задачі. Наведемо, в зв'язку з цим, важливу теорему з математичного аналізу.

Теорема Вейерштраса. Нехай X – замкнена, обмежена множина в E^n, $f(x)$ – неперервна функція на X. Тоді існує точка глобального мінімуму функції $f(x)$ на множині X.

Задача (1.1) є загальною задачею оптимізації. Класифікацію задач оптимізації можна проводити в залежності від властивостей функції $f(x)$ та множини X.

Задача (1.1) називається **задачею безумовної оптимізації**, якщо множина $X = E^n$, тобто, якщо вона має вигляд

$$f(x) \to min,\ x \in E^n. \tag{1.5}$$

Задача (1.1) називається **задачею умовної оптимізації**, якщо множина X є підмножиною простору E^n, тобто $X \subset E^n$.

При розв'язанні задач оптимізації часто застосовують їхню геометричну інтерпретацію, яка основана на понятті **ліній (поверхонь) рівня** функції $f(x)$, тобто множин вигляду

$$\Gamma_\alpha = \left\{ x \in E^n : f(x) = \alpha \right\},\ \alpha \in E^1,$$

де $E^1 = R_1 = \{ x : -\infty < x < +\infty \}$ – числова вісь.

Одним із найважливіших класів задач є **задачі математичного програмування**. Так називають задачу (1.1), коли її допустима множина має вигляд:

$$X = \left\{ x \in X_0 : g_i(x) \le 0, i = \overline{1,m};\ h_j(x) = 0,\ j = \overline{1,l} \right\}.$$

Множина X_0 є множиною простої структури. Наприклад, куля в n-вимірному просторі, координатний паралелепіпед, невід'ємний октант і деякі інші.

Таким чином, задачу математичного програмування можна записати у вигляді

$$f(x) \to min, \tag{1.6}$$

за умовами

$$\begin{array}{l} g_i(x) \le 0, \ \ i = \overline{1, m}; \\ h_j(x) = 0, \ \ j = \overline{1, l}, \quad x \in X_0 \end{array}.$$

При цьому умови типу $g_i(x) \le 0, \quad i = \overline{1,m}$ називаються **обмеженнями–нерівностями**, а типу $h_j(x) = 0, \quad j = \overline{1,l}$ – **обмеженнями–рівностями**. Обидва ці типи умов називаються **функціональними обмеженнями**, а умова $x \in X_0$ носить назву **прямого обмеження**.

Важливими розділами математичного програмування є лінійне програмування та квадратичне програмування.

Задачею **лінійного програмування** (ЛП) називається задача мінімізації (максимізації) лінійної цільової функції при лінійних обмеженнях:

$$\sum_{j=1}^{n} c_j x_j \to min(max), \qquad \sum_{j=1}^{n} a_{ij} x_j \le b_i, \ i = \overline{1,m}, \qquad \sum_{j=1}^{n} a_{ij} x_j = b_i, \ i = \overline{m+1,l},$$

$$x_j \ge 0, \ j = \overline{1,s}, \quad s \le n.$$

Задачею **квадратичного програмування** називається задача мінімізації при лінійних обмеженнях квадратичної цільової функції вигляду:

$$f(x) = \frac{1}{2}(A\,x, x) - (b, x) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{i=1}^{n} b_i x_i,$$

де A – симетрична, невід'ємно визначена матриця розміру $n \times n$, b – вектор з E^n.

У подальшому, говорячи про диференціальні характеристики функції $f(x)$ у точці x, маємо на увазі, що $x \in E^n$, $f(x)$ – числова функція, яка визначена в деякому околі x, $g(x) \in E^n$ – довільний вектор, такий що $\|g\| = 1$.

Функція називається **диференційовною в точці** x, якщо існує вектор $f'(x) \in E^n$ такий, що для будь-якого вектора $g(x) \in E^n, \|g(x)\| = 1$, виконується рівність

$$f(x + \alpha g) = f(x) + \alpha\,(f'(x), g) + o(x, g, \alpha), \tag{1.7}$$

де $o(x, g, \alpha)$ – величина нескінченно мала більш високого порядку, ніж α, тобто $\lim\limits_{\alpha \to 0} \dfrac{o(x, g, \alpha)}{\alpha} = 0$.

Умова (1.7) одночасно визначає **градієнт** $f'(x)$ функції $f(x)$, при цьому $f'(x) = \left(\dfrac{\partial f(x)}{\partial x_1}, \ldots, \dfrac{\partial f(x)}{\partial x_n}\right)$, де $\dfrac{\partial f(x)}{\partial x_i} = \lim\limits_{\alpha \to 0} \dfrac{f(x + \alpha\, e^i) - f(x)}{\alpha}$ – частинна похідна функції $f(x)$ у точці x по аргументу x_i, $e^i = (\,0, \ldots, 0, 1, 0, \ldots, 0\,)$ є i-тим координатним ортом.

Будь-яка функція, що є диференційовною в точці, неперервна в цій точці.

Нехай функція $f(x)$ визначена на певній множині X і в усіх внутрішніх точках має градієнт, який є неперервною вектор-функцією на множині внутрішніх точок. Тоді функція $f(x) \in C^1(X)$, де $C^1(X)$ – клас неперервно-диференційовних (гладких) функцій на X.

Функція називається **диференційовною за напрямком** $g(x) \in E^n, \|g(x)\| = 1$ в точці x, якщо існує скінченна границя

$$f'(x, g) = \lim_{\alpha \to 0} \frac{f(x + \alpha\, g) - f(x)}{\alpha}.$$

Якщо функція є диференційовною в точці, вона диференційовна в цій точці за будь-яким напрямком $g(x)$, при цьому

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

Функція називається **двічі диференційовною в точці** x, якщо разом з градієнтом існує симетрична матриця $f''(x)$ порядку $n \times n$ така що

$f(x + h) = f(x) + (\,f'(x), h) + \dfrac{1}{2}(\,f''(x)h, h) + \beta(h, x)$, де $\dfrac{\beta(h, x)}{\|h\|^2} \to 0$ при $\|h\| \to 0$.

Квадратичну форму $d^2 f(x) = \left(f''(x)h, h\right)$ змінної $h = (h_1, \ldots, h_n) \in E^n$ називають **другим диференціалом** функції $f(x)$ в точці x, а матрицю

$f''(x) = \left(\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j=1,n}$ – **матрицею других похідних (гессіаном, матрицею Гессе)** функції $f(x)$ в точці x.

Якщо всі елементи матриці $f''(x)$ є неперервними функціями змінної x в усіх внутрішніх точках множини X, на якій визначена функція $f(x)$, то функція $f(x)$ належить класу $C^2(X)$ – двічі неперервно-диференційовних (двічі гладких) функцій на X.

§2. Елементи опуклого аналізу

Опуклий аналіз – розділ математики, де вивчаються опуклі множини та опуклі функції. Поняття та твердження опуклого аналізу грають фундаментальну роль у теорії і чисельних методах оптимізації.

Непорожня множина $X \subset E^n$ називається **опуклою**, якщо для будь-яких $y, z \in X$ точки $x_\lambda = \lambda y + (1-\lambda) z \in X$ для всіх λ, $0 \le \lambda \le 1$, тобто X разом з будь-якими своїми точками y, z містить відрізок, який їх з'єднує.

Прикладами опуклих множин є:

1. $X = \left\{ x \in E^n : (c,x) = \alpha \right\}$ – гіперплощина в E^n, α –скаляр, $c \in E^n$ – нормаль до гіперплощини;
2. $X = \left\{ x \in E^n : (c,x) \le \alpha \right\}$ – півпростір в E^n;
3. $X = \left\{ x \in E^n : Ax \le b \right\}$ – многогранна множина, A – матриця розміру $n \times n$, b – n-вимірний вектор;
4. $X = \left\{ x \in E^n : \| x - x_0 \| \le R \right\}$– куля з центром у точці x_0 радіуса R.

Приклад. Доведемо, що куля $X = \left\{ (x_1, x_2) \in E^2 : x_1^2 + x_2^2 \le R^2 \right\}$ є опуклою множиною.

Нехай $(x_1, x_2), (x_1^{'}, x_2^{'}) \in X$. З визначення опуклої множини потрібно довести, що $(\lambda x_1 + (1-\lambda)x_1^{'}, \lambda x_2 + (1-\lambda)x_2^{'}) \in X$, $0 \le \lambda \le 1$.

Виконаємо перетворення:

$$(\lambda x_1 + (1-\lambda)x_1^{'})^2 + (\lambda x_2 + (1-\lambda)x_2^{'})^2 =$$
$$= \lambda^2 (x_1^2 + x_2^2) + 2\lambda(1-\lambda)(x_1 x_1^{'} + x_2 x_2^{'}) + (1-\lambda)^2 \left[(x_1^{'})^2 + (x_2^{'})^2 \right].$$

Використовуючи нерівність Коші-Буняковського

$$x_1 x_1^{'} + x_2 x_2^{'} \le \sqrt{x_1^2 + x_2^2} \sqrt{(x_1^{'})^2 + (x_2^{'})^2}$$

і замінюючи в усіх виразах $x_1^2 + x_2^2$ та $(x_1^{'})^2 + (x_2^{'})^2$ на R^2, отримаємо:

$$\left(\lambda\, x_1 + (1-\lambda)\, x_1^{'}\right)^2 + \left(\lambda\, x_2 + (1-\lambda)\, x_2^{'}\right)^2 \le R^2 \left[\lambda^2 + 2\lambda\,(1-\lambda) + (1-\lambda)^2 \right] = R^2.$$

Функція $f(x)$, яка визначена на опуклій множині $X \subset E^n$ називається **опуклою,** якщо

$$f\left(\lambda\, x_1 + (1-\lambda)\, x_2\right) \le \lambda\, f(x_1) + (1-\lambda) f(x_2) \qquad (1.8)$$

для усіх $x_1, x_2 \in X$ та $\lambda \in [0,1]$.

Функція $f(x)$, яка визначена на опуклій множині $X \subset E^n$ називається **строго опуклою**, якщо нерівність (1.8) виконується як строга для всіх значень $x_1, x_2 \in X$ та усіх $\lambda \in (0,1)$. Функція $f(x)$ **увігнута (строго увігнута)** на $X \subset E^n$, якщо функція $(-f(x))$ є опуклою (строго опуклою) на X.

Зауважимо, що опуклість множини X є істотною в цих означеннях, оскільки для всіх $x_1, x_2 \in X$ та всіх $\lambda \in [0,1]$ точка $\lambda\, x_1 + (1-\lambda)x_2$ повинна належати множині X.

Такі функції є опуклими:

- $f_1(x) = 3x + 4,$
- $f_2(x) = x^2 - 2x,$
- $f_3(x) = |x|,$
- $f_4(x) = -\sqrt{x},$
- $f_5(x) = 2(x_1^2 + x_2^2 - 2x_1 x_2),$
- $f_6(x) = x_1^4 + 2x_2^2 + x_3^2 - 4x_1 - 4x_2 x_3.$

Деякі властивості опуклих функцій:

Нехай X –непорожня опукла множина в E^n.

1. Опукла функція $f(x)$ неперервна в кожній внутрішній точці множини X.
2. Опукла функція $f(x)$ має похідну по будь-якому напрямку в кожній внутрішній точці опуклої множини X.
3. Диференційовна на множині X функція $f(x)$ опукла на цій множині тоді і тільки тоді, коли

$$f(x) - f(\bar{x}) \geq \left(f'(\bar{x}), x - \bar{x}\right) \text{ для всіх } x, \bar{x} \in X. \quad (1.9)$$

Нагадаємо, що графік лінійної функції $l(x) = f(\bar{x}) + \left(f'(\bar{x}),\, x - \bar{x}\right)$ називається **дотичною гіперплощиною** до графіка функції $f(x)$ у точці $(\bar{x}, f(\bar{x}))$. Таким чином, співвідношення (1.9) означає, що графік функції $f(x)$ лежить не нижче за дотичну гіперплощину до нього в точці $(\bar{x}, f(\bar{x}))$ (рис.1.2).

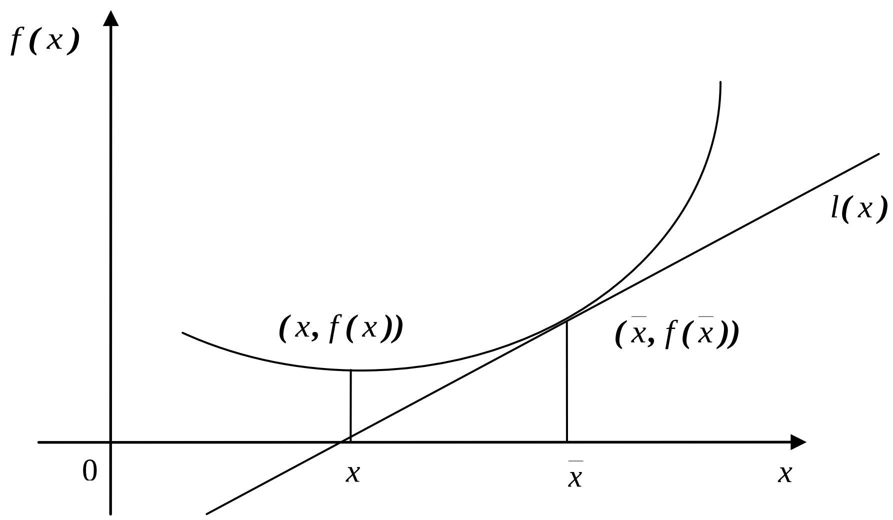


Рис. 1.2

Згідно з властивістю 3, можна отримати інший критерій опуклості в термінах перших похідних.

4. Неперервно-диференційовна множині X функція $f(x)$ опукла тоді і тільки тоді, коли

$$(f'(x) - f'(\bar{x}), x - \bar{x}) \geq 0,\ \forall x, \bar{x} \in X.$$

5. **Критерій опуклості в термінах других похідних**.

Двічі неперервно-диференційовна на множині X функція $f(x)$ опукла тоді і тільки тоді, коли

$$\left(f''(x)h, h\right) \geq 0,\ \forall x \in X,\ \forall h \in E^n. \quad (1.10)$$

Умова (1.10) означає, що матриця Гессе (гессіан) $f''(x)$ невід'ємно визначена для $\forall x \in X$.

Має місце критерій Сильвестра додатної (невід'ємної) визначеності матриці: матриця $f''(x)$ додатно (невід'ємно) визначена в тому і тільки в тому випадку, коли всі її головні (кутові) мінори додатні (невід'ємні).

Нагадаємо, що головними (кутовими) мінорами матриці $A = \{a_{ij}\}$, $i, j = \overline{1, n}$ називаються визначники:

$$\Delta_1 = a_{11},\ \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\ \ldots,\ \Delta_n = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}.$$

Якщо ранг матриці $A = \{a_{ij}\}$, $i, j = \overline{1, n}$ дорівнює r і $r < n$, то вона буде додатно полувизначеною, якщо перші r її головних мінорів додатні, а інші дорівнюють нулю.

Матриця $A = \{a_{ij}\}$, $i, j = \overline{1, n}$ буде від'ємно визначеною в тому і тільки в тому випадку, коли мають місце такі знаки головних (кутових) мінорів:

$$\Delta_1 < 0,\ \Delta_2 > 0,\ \Delta_3 < 0,\ \Delta_4 > 0,\ \ldots$$

Властивість 5 у комбінації з критерієм Сильвестра складають зручний апарат для перевірки опуклості функції невеликого числа змінних, коли обчислення мінорів не викликає труднощів.

Приклад. Розглянемо функцію $f(x) = x_1^2 / x_2$ на множині

$$X = \left\{ x = (x_1, x_2), x \in E^2 : x_2 > 0 \right\}.$$

Маємо таку матрицю других похідних (матрицю Гессе)

$$f''(x) = \begin{bmatrix} \dfrac{2}{x_2} & \dfrac{-2x_1}{x_2^2} \\ \dfrac{-2x_1}{x_2^2} & \dfrac{2x_1^2}{x_2^3} \end{bmatrix}.$$

З критерію Сильвестра витікає, що для $\forall x \in X$ матриця $f''(x)$ невід'ємно визначена. Таким чином, функція $f(x)$ опукла на X .

Доведення цього факту безпосередньо з визначення (1.8) привело би до достатньо громіздких викладок.

6. Кожний локальний мінімум опуклої функції $f(x)$, яка визначена на опуклій множині X, є глобальним.
7. Строго опукла функція $f(x)$ на опуклій множині X досягає мінімуму в єдиній точці.

Субградієнт і субдиференціал опуклої функції

Згідно з властивістю 2 опукла функція має похідні по всім напрямкам у внутрішніх точках області визначення. Але градієнт може не існувати в точках, де функція не є диференційовною. Узагальненням поняття градієнта на випадок негладкої опуклої функції є поняття субградієнта і субдиференціала (множини субградієнтів), які використовуються в теорії негладких опуклих задач оптимізації, а також при розробці чисельних методів їхнього розв'язання.

Нагадаємо, що для опуклої функції $f(x)$, яка є диференційовною в точці $\bar{x}$, виконується нерівність (1.9). Згідно з цією нерівністю, графік функції $f(x)$ лежить не нижче дотичної гіперплощини до нього в точці $(\bar{x}, f(\bar{x}))$ (рис.1.2). Згідно з наступним визначенням під субградієнтом розуміють будь-який вектор, який можна підставити до нерівності (1.9) замість $f'(\bar{x})$.

Нехай функція $f(x)$ визначена на множині $X \subset E^n$. Вектор $g(\bar{x}) \in E^n$ називається **субградієнтом** функції $f(x)$ у точці $\bar{x} \in E^n$, якщо виконується нерівність

$$f(x) - f(\bar{x}) \geq \left(g(\bar{x}),\, x - \bar{x}\right), \text{ для } \forall x \in X. \qquad (1.11)$$

Множина всіх субградієнтів функції $f(x)$ у точці $\bar{x}$ називається **субдиференціалом** $f(x)$ в точці $\bar{x}$ і позначається через $\partial f(\overline{x})$.

Співвідношення (1.11) за аналогією з (1.9) означає, що графік функції $f(x)$ лежить не нижче графіка лінійної функції $l(x) = f(\bar{x}) + (g(\bar{x}), x - \bar{x})$. При цьому графік лінійної функції називається **опорною гіперплощиною** до графіка функції $f(x)$ у точці $(\bar{x}, f(\bar{x}))$.

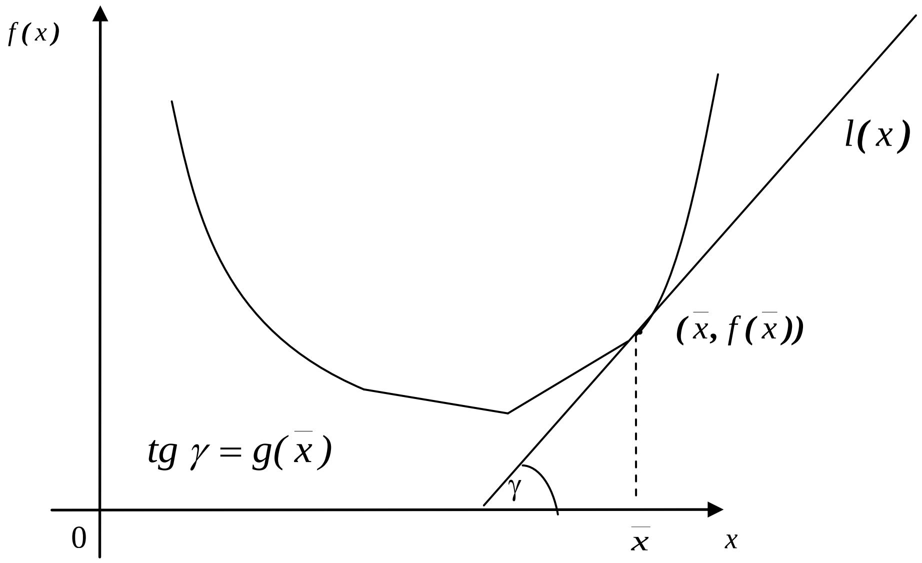


Рис. 1.3

Субградієнт для функції числового аргументу має геометричний сенс тангенса кута нахилу опорної гіперплощини, подібно до того, як похідна, якщо вона існує, є тангенс нахилу дотичної.

З рис. 1.3 неважко зрозуміти, що для функції $f(x)$, яка є опуклою на опуклій числовій множині X, має місце формула

$$\partial f(\bar{x}) = \left[f'_{-}(\bar{x}), f'_{+}(\bar{x}) \right],$$

де $f'_{-}(\bar{x})$, $f'_{+}(\bar{x})$ – ліво- і правосторонні похідні функції $f(x)$ у точці $\bar{x} \in \operatorname{int} X$. Через $int\, X$ позначено сукупність усіх внутрішніх точок множини X.

Деякі властивості субдиференціалів опуклої функції

1. Нехай X – відкрита опукла множина на E^n. Тоді для того, щоб функція $f(x)$, яка визначена на X, мала непорожній диференціал в усіх точках X, необхідно і достатньо, щоб $f(x)$ була опуклою на X.

2. Для опуклої функції $f(x)$, яка визначена на відкритій опуклій множині $X \subset E^n$, субдиференціал $\partial f(x)$ при всіх $x \in X$ є непорожньою, опуклою, замкненою і обмеженою множиною.
3. Для гладкої опуклої функції, яка визначена на опуклій множині $X \subset E^n$, $\partial f(x) = \{ f'(x) \}$ при всіх $x \in int X$, тобто градієнт $f'(x)$ є єдиним субградієнтом функції $f(x)$ у точці x.

Приклад. Знайти субградієнт функції $f(x) = min[f_1(x), f_2(x)]$,

$$\text{де } f_1(x) = 4 - |x|, \; f_2(x) = 4 - (x-2)^2, \; x \in E^1.$$

Так як $f_2(x) \geq f_1(x)$ при $1 \leq x \leq 4$, то $f(x)$ може бути записана у вигляді

$$f(x) = \begin{cases} 4 - x, & 1 \leq x \leq 4 \\ 4 - (x-2)^2, & x < 1, x > 4 \end{cases}.$$

На рис. 1.4 увігнута функція $f(x)$ зображена маркірованою лінією.

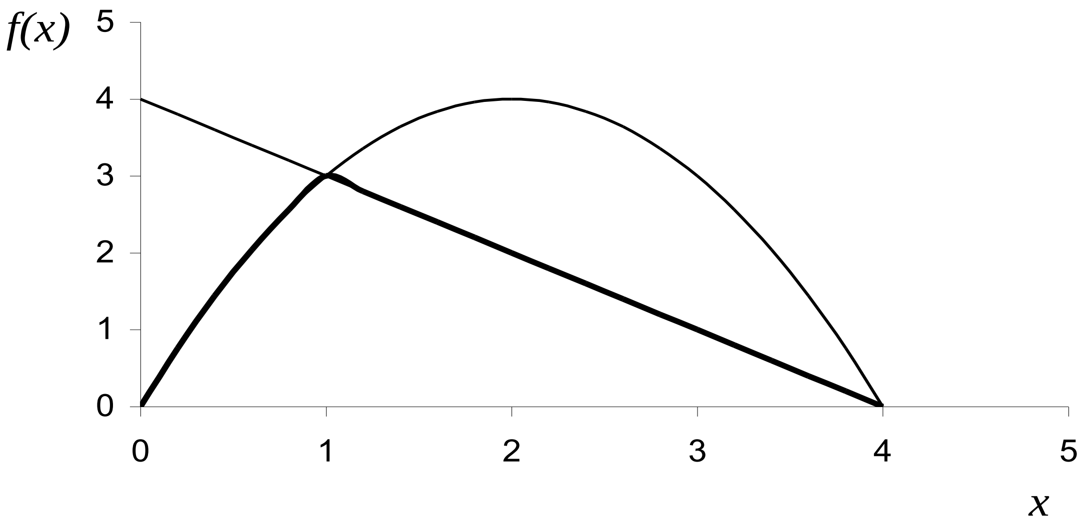


Рис. 1.4

У будь-якій точці x відкритого інтервалу $1 < x < 4$ субградієнтом до функції $f(x)$ є число $g = -1$. За точками $x = 1$ та $x = 4$ субградієнти визначаються не єдином чином, тому що в цих точках існує багато опорних гіперплощин. У точці $x = 1$ сімейство субградієнтів визначається співвідношенням $\lambda f_1'(1) + (1-\lambda) f_2'(1) = -\lambda + 2(1-\lambda) = 2 - 3\lambda$ для $\lambda \in [0,1]$. Іншими словами, будь-яке число g з інтервалу $[-1,2]$ є субградієнтом функції

$f(x)$ у точці $x = 1$. У точці $x = 4$ сімейство субградієнтів визначається співвідношенням $\lambda f_1'(4) + (1-\lambda) f_2'(4) = -\lambda - 4(1-\lambda) = 3\lambda - 4$ для $\lambda \in [0,1]$, тобто, будь-яке число g з інтервалу $[-4,-1]$ є субградієнтом функції $f(x)$ у точці $x = 4$.

Нехай $f_1(x)$, $f_2(x)$ – диференційовні опуклі функції,

$$f(x) = max[\ f_1(x), f_2(x)\ ]$$

і існує така точка $\bar{x}$, в якій $f(\bar{x}) = f_1(\bar{x}) = f_2(\bar{x})$. Вектор ξ є субградієнтом функції $f(x)$ у точці $\bar{x}$ тоді і тільки тоді, коли

$$\xi = \lambda\ f_1'(\bar{x}) + (1-\lambda)\, f_2'(\bar{x}),\ \lambda \in [0,1].$$

Строго квазіопуклі функції

У багатьох практично важливих випадках не потрібно вимагати опуклості досліджуваної функції. Достатньо значно слабших припущень про квазіопуклість або строгу квазіопуклість цільової функції. Строго квазіопуклі функції особливо важливі в математичному програмуванні, тому що для них будь-який локальний мінімум на опуклій множині є глобальним.

Нехай X – непорожня опукла множина в E^n. Функція $f(x)$ називається **квазіопуклою** функцією на множині X, якщо для всіх $y, z \in X$ виконується нерівність

$$f(\lambda\, y + (1-\lambda)\, z) \le max\{\ f_1(y),\ f_2(z)\ \} \text{ при будь–якому } \lambda \in (0,1). \quad (1.12)$$

Функція $f(x)$ називається **строго квазіопуклою** функцією на множині X, якщо нерівність (1.12) виконується як строга при $f(y) \ne f(z)$. На рис. 1.5 наведені на площині приклади графіків квазіопуклих (рис.1.5.*а*) і строго квазіопуклих функцій (рис.1.5.*б*– 1.5.*г*).

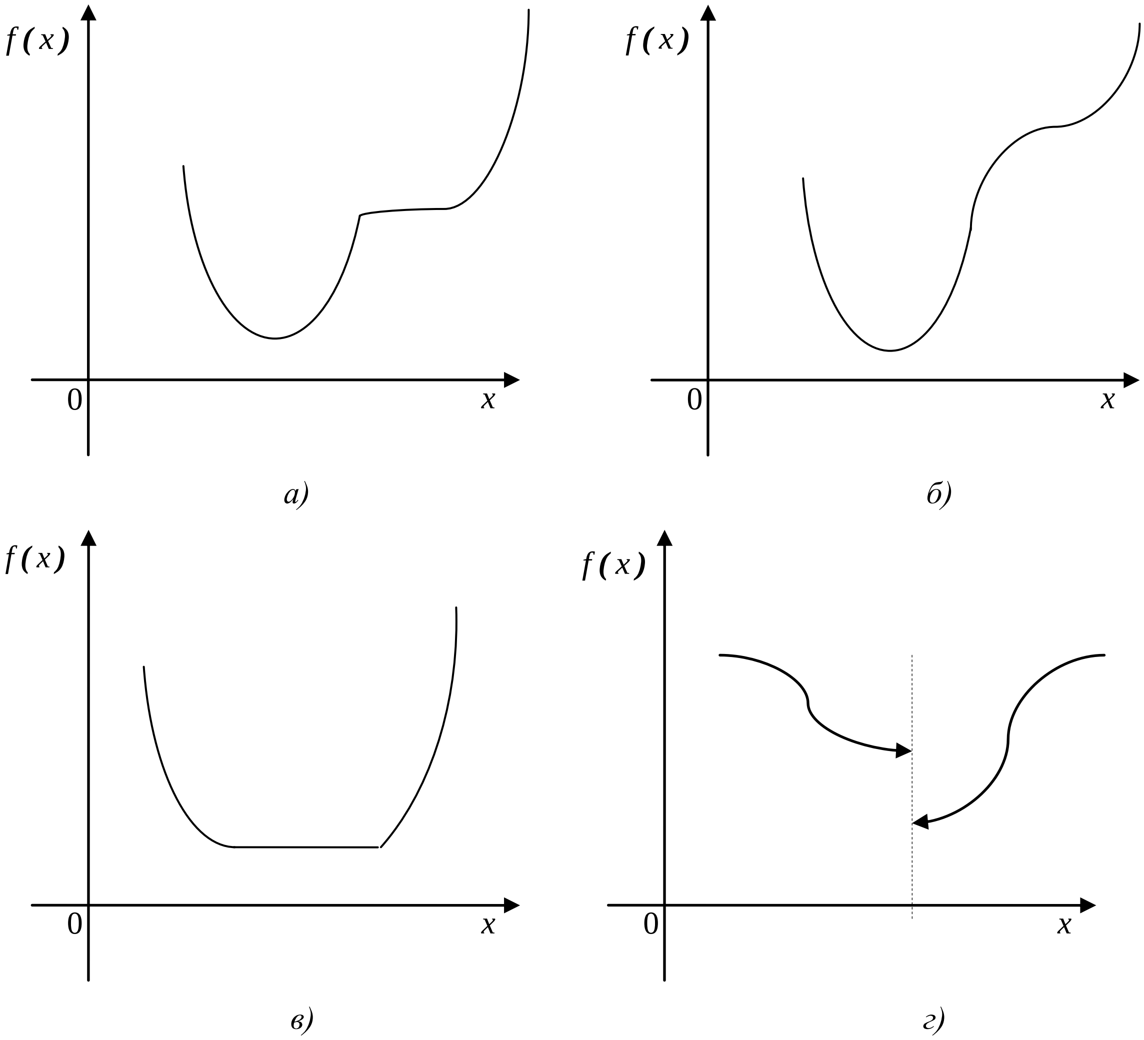


Рис. 1.5

§3. Умови оптимальності

При вивченні будь-якого типу задач оптимізації важливе місце займає питання про умови оптимальності або умови екстремуму. Розрізняють необхідні умови оптимальності, тобто умови, яким повинні задовольняти точки, що є розв'язками задачі, і достатні умови оптимальності, тобто умови, з яких випливає, що дана точка є розв'язком задачі. Інтерес до умов оптимальності пояснюється тим, що вони, по-перше, складають основу якісних методів теорії оптимізації, тобто методів, спрямованих на вивчення властивостей екстремальних задач; по-друге, використовуються при побудові і обґрунтуванні чисельних методів розв'язання цих задач; по-третє, дозволяють у простих випадках, явним чином, розв'язати задачу.

Наведемо умови оптимальності для деяких класів задач оптимізації.

1. Нехай задана загальна задача оптимізації

$$f(x) \to min\,, \ x \in X. \tag{1.13}$$

Необхідна умова оптимальності в термінах напрямків (геометрична умова оптимальності

Вектор $h \in E^n$ задає **напрямок спадання** функції $f(x)$ у точці $x_* \in E^n$, якщо $f(x_* + \alpha h) < f(x_*)$ при всіх достатньо малих $\alpha > 0$. Множина таких $h \in E^n$ позначається як $U(x_*, f)$.

Достатня і необхідна ознака напрямків спадання функції

Лема. Нехай функція $f(x)$ диференційовна в точці $x \in E^n$. Якщо вектор $h \in E^n$ задовольняє умові

$$\left(f^{'}(x), h\right) < 0, \tag{1.14}$$

то $h \in U(x, f)$. Якщо $h \in U(x, f)$, то

$$\left(f^{'}(x), h\right) \le 0. \tag{1.15}$$

Доведення. Нехай виконана умова (1.14). Тоді

$$f(x + \alpha\, h) - f\left(x^*\right) = \left(f^{'}(x), \alpha\, h\right) + o(\alpha) = \alpha\left(\left(f^{'}(x), h\right) + \frac{o(\alpha)}{\alpha}\right) < 0$$

при всіх достатньо малих $\alpha > 0$, тобто $h \in U(x, f)$.

Нехай $h \in U(x, f)$ і $\left(f^{'}(x), h\right) > 0$. Тоді за допомогою попереднього твердження переконуємося, що h – напрямок зростання. Отримане протиріччя показує, що і в розглядуваному випадку справедлива нерівність (1.15).

Лема доведена.

Геометрично умова (1.14) означає, що вектор h складає тупий кут з градієнтом $f^{'}(x)$.

Вектор $h \in E^n$ задає **можливий напрямок** відносно множини X у точці $x_* \in X$, якщо $x_* + \alpha\, h \in X$ при всіх достатньо малих $\alpha > 0$. Множину всіх таких $h \in E^n$ позначимо через $V(x_*, X)$.

Наведемо необхідну умову локальної оптимальності в задачі (1.13) в термінах напрямків, яка не потребує будь-яких припущень відносно X та f.

Теорема 1. Якщо x_* – локальний розв'язок задачі (1.13), то

$$U(x_*, f) \cap V(x_*, X) = \varnothing. \tag{1.16}$$

Теорема 1 означає, що з точки x_*, яка є локальним розв'язком, жодний напрямок спадання (спуску) не є можливим.

Диференціальна умова оптимальності
в задачі мінімізації (1.13) на опуклій множині

Теорема 2. Нехай в задачі (1.13) множина X опукла, функція $f(x)$ диференційовна в точці $x_* \in X$, Тоді:

1) якщо x_* – локальний розв'язок задачі (1.13), то
$$\left(f'(x_*), x - x_*\right) \geq 0,\ x \in X, \tag{1.17}$$
2) якщо функція $f(x)$ опукла на X і виконується (1.17), то x_* – глобальний розв'язок задачі (1.13),
3) якщо $x_* \in int\, X$, то (1.17) еквівалентно умові $f'(x_*) = 0$.

Будь-яка точка $x_* \in X$, яка задовольняє умові (1.17) або (1.18), називається **стаціонарною точкою.**

З теореми 2 витікає відома з класичного аналізу необхідна умова локальної оптимальності для задачі безумовної оптимізації $X \in E^n$: $f'(x_*) = 0$, яка буде і достатньою умовою глобальної оптимальності, якщо функція $f(x)$ опукла на E^n.

Геометрично умова (1.17) означає, що градієнт $f'(x_*)$ (якщо він не дорівнює нулю) складає нетупий кут з вектором, який направлений з x_* у будь-яку точку $x \in X$. Ілюстрацією є рис. 1.6. З нього видно, що точка $\bar{x}$, яка не задовольняє умові (1.17) не може бути розв'язком: з неї існують можливі напрямки, вздовж яких функція $f(x)$ спадає.

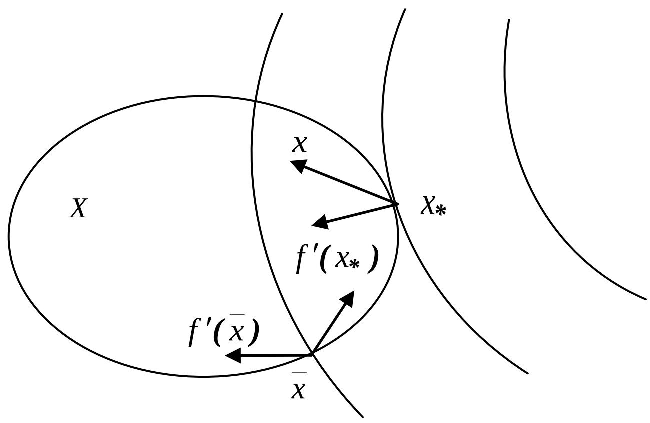


Рис. 1.6

§4. Поняття про чисельні методи оптимізації

Будь-який чисельний метод (алгоритм) розв'язання задачі оптимізації заснований на точному або наближеному обчисленні її характеристик (а саме: значення цільової функції, функцій, які задають допустиму множину, а також їхніх похідних). На основі отриманої інформації будується наближення до розв'язку задачі – шуканої точки мінімуму x_*, або, якщо така точка не є єдиною, – до множини точок мінімуму X_*.

Інколи, якщо тільки це і потрібно, будується наближення до мінімального значення цільової функції $f_* = \min\limits_{x \in X} f(x)$.

Для кожної конкретної задачі питання про те, які характеристики потрібно вибрати для обчислення, вирішується залежно від властивостей цільової функції, обмежень та можливостей зберігання та обробки інформації.

Алгоритми, які використовують лише інформацію про значення цільової функції, називаються **алгоритмами нульового порядку**; алгоритми, які використовують також інформацію про значення перших похідних – **алгоритмами першого порядку**; алгоритми, що використовують, крім того, інформацію про другі похідні, – **алгоритмами другого порядку**.

У подальшому, для запису методу оптимізації будемо використовувати співвідношення вигляду

$$x^{(k+1)} = x^{(k)} + \alpha_k h^{(k)},\ \alpha_k \in E^1,\ k = 0,1,2,\ldots \qquad (1.19)$$

або у покоординатній формі

$$x_1^{(k+1)} = x_1^{(k)} + \alpha_k h_1^{(k)},$$

$$\dots\dots\dots\dots\dots\dots$$

$$x_n^{(k+1)} = x_n^{(k)} + \alpha_k h_n^{(k)},$$

$$\alpha_k \in E^1,\ k = 0,1,2,\dots,\ x = (x_1,\dots,x_n).$$

Конкретний алгоритм визначається: вибором початкової точки $x^{(0)}$, правилами вибору векторів $h^{(k)}$ та чисел α_k на основі отриманої в результаті обчислень інформації, а також умовою зупинки ітераційного процесу.

Вектор $h_k = (h_1^{(k)},\dots,h_n^{(k)})$ визначає напрямок $(k+1)$-го кроку ($(k+1)$-ої ітерації) методу мінімізації, а коефіцієнт α_k – довжину цього кроку. Звичайно, назва методу мінімізації визначається способом вибору $h^{(k)}$, а його різні варіанти зв'язані з різними способами вибору α_k.

Серед методів мінімізації можна умовно виділити скінченнокрокові та нескінченнокрокові методи. Скінченнокроковими або скінченними називають методи, які гарантують знаходження розв'язку задачі за скінченне число кроків. Скінченні методи вдається побудувати лише для деяких спеціальних типів задач оптимізації, наприклад, задач лінійного та квадратичного програмування. Для нескінченнокрокових методів досягнення розв'язку гарантується тільки в граничному значенні.

Збіжність методів оптимізації

Важливою характеристикою нескінченнокрокових методів є збіжність.

Метод (1.19) **збігається,** якщо $x^{(k)} \to x_*$ при $k \to \infty$, де x_* – розв'язок задачі (1.1). У випадку, коли точка мінімуму x_* не єдина, під збіжністю методу розуміється збіжність послідовності $\{x^{(k)}\}_0^\infty$ до множини X_* точок мінімуму функції $f(x)$.

Коли $f\left(x^{(k)}\right) \to f(x_*)$, то кажуть, що метод (1.19) **збігається за функцією**, при цьому послідовність $\{x^{(k)}\}_0^\infty$ називається **мінімізуючою**.

Мінімізуюча послідовність $\{x^{(k)}\}_0^\infty$ може і не збігатися до точки мінімуму функції $f(x)$. Наприклад, для функції $f(x) = x^2 / (1 + x^4)$, $x \ge 0$, графік якої зображено на рис. 1.7, мінімізуюча послідовність $x^{(k)} = k$ не збігається до точки мінімуму $x_* = 0$.

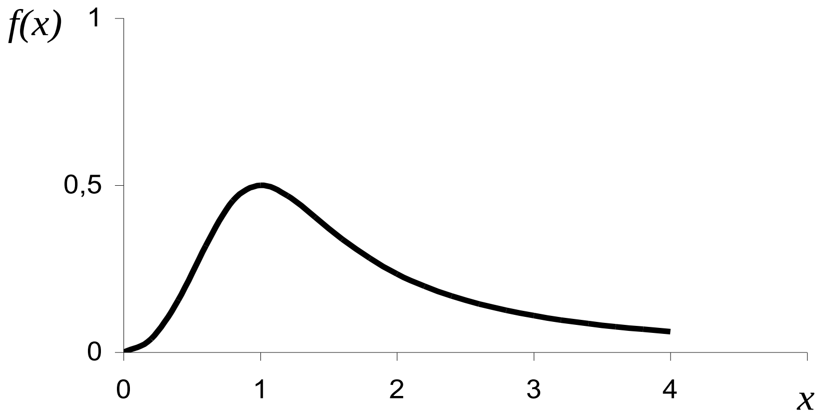


Рис. 1.7

Ефективність збіжного методу можна охарактеризувати за допомогою поняття швидкості збіжності.

Нехай $x^{(k)} \to x_*$ при $k \to \infty$.

Послідовність $x^{(k)}$ збігається до x_* **лінійно (з лінійною швидкістю, зі швидкістю геометричної прогресії)**, якщо існують такі сталі $q \in (0,1)$ та k_0, що

$$\| x^{(k+1)} - x_* \| \le q \| x^{(k)} - x_* \| \text{ при } k \ge k_0.$$

Послідовність $x^{(k)}$ збігається до x_* **надлінійно (з надлінійною швидкістю)**, якщо

$$\| x^{(k+1)} - x_* \| \le q_{k+1} \| x^{(k)} - x_* \|, \ q_k \to +0, \ k \to \infty.$$

Послідовність $x^{(k)}$ збігається до x_* **з квадратичною швидкістю**, якщо існують такі сталі $C \ge 0$ та k_0, що

$$\| x^{(k+1)} - x_* \| \le C \| x^{(k)} - x_* \|^2 \text{ при } k \ge k_0.$$

Для характеристики збіжності послідовності наближених значень $f(x^{(k)})$ до $f(x_*)$ використовуються такі ж терміни: лінійна, надлінійна та квадратична збіжність.

Установлення факту збіжності та оцінка швидкості збіжності дають істотну інформацію про обраний метод мінімізації. Насамперед, вимоги, які доводиться накладати в теоремі про збіжність на цільову функцію, показують область застосування методу. Часто в теоремах про збіжність в явному вигляді формулюються вимоги до початкового наближення $x^{(0)}$. Нарешті, аналіз швидкості збіжності дає корисну кількісну та якісну характеристику методу оптимізації, який вивчається.

У той же час, реальний процес оптимізації не може бути нескінченнокроковим. Звідси випливає, що роль теорем про збіжність є обмеженою. Крім того, в ряді випадків припущення цих теорем важко перевірити. Тому при виборі відповідного методу розв'язання реальних задач доводиться у багатьох випадках керуватися здоровим глуздом, досвідом, інтуїцією, а також результатами чисельних експериментів.

§5. Метод множників Лагранжа

Нехай потрібно розв'язати таку задачу

$$f(x) \to min \tag{1.20}$$

при умовах

$$x \in X = \left\{ x \in E^n : g_i(x) = 0, i = \overline{1, s} \right\}. \tag{1.21}$$

Припустимо, що функції $f(x), g_1(x), \ldots, g_s(x)$ визначені і диференційовні на множині X та $s < n$.

Основна ідея методу множників Лагранжа полягає у переході від задачі на умовний екстремум до задачі знаходження безумовного екстремуму спеціально побудованої функції Лагранжа з наступним застосуванням класичного методу пошуку екстремуму.

Введемо змінні $\lambda_0, \lambda_1, \ldots, \lambda_s$ і побудуємо функцію Лагранжа

$$L(x,\lambda)=\lambda_0\ f(x)+\sum_{i=1}^{s}\lambda_i\ g_i(x). \qquad (1.22)$$

Сформулюємо необхідну умову екстремуму для задачі (1.20),(1.21).

Теорема (Ознака Лагранжа). Для того, щоб вектор $x_* = \left(x_1^*,\ldots,x_n^*\right)$ був розв'язком задачі (1.20), (1.21), необхідно існування вектора $\lambda^* = \left(\lambda_0^*,\lambda_1^*,\ldots,\lambda_s^*\right)$, такого що його компоненти λ_i^* $\left(i=\overline{1,s}\right)$ одночасно не дорівнюють нулеві і такі, що всі частинні похідні першого порядку функції Лагранжа $L(x,\lambda)$ за змінними x_j $\left(j=\overline{1,n}\right)$ в точці x_* дорівнюють нулеві, тобто

$$\frac{\partial L\left(x_*,\lambda^*\right)}{\partial x_k}=\lambda_0^*\ \frac{\partial f\left(x_*\right)}{\partial x_k}+\sum_{i=1}^{s}\lambda_i^*\ \frac{\partial g_i\left(x_*\right)}{\partial x_k}=0,\ k=\overline{1,n},\ \lambda^*\neq 0. \qquad (1.23)$$

Умова (1.23) має ясний геометричний сенс. Вона означає, що градієнти $f'\left(x_*\right), g_1'\left(x_*\right),\ldots, g_s'\left(x_*\right)$ є лінійно залежними. У випадку, коли $s=1$, вектори $f'\left(x_*\right), g_1'\left(x_*\right)$ повинні бути колінеарними.

Якщо $\lambda_0 \neq 0$, то задачу (1.22) називають невиродженою (регулярною), при $\lambda_0 = 0$ – виродженою. У багатьох практичних задачах $\lambda_0 = 1$.

Відзначимо, що вимірність задачі (1.20), (1.21) дорівнює n, а вимірність задачі (1.22) є $n+s+1$.

Алгоритм методу множників Лагранжа

Крок 1. Будуємо функцію Лагранжа

$$L(x,\lambda)=f(x)+\sum_{i=1}^{s}\lambda_i\ g_i(x)$$

та складаємо систему

$$\frac{\partial L(x,\lambda)}{\partial x_k}=0,\ k=\overline{1,n},\ \frac{\partial L(x,\lambda)}{\partial \lambda_i}=g_i(x)=0,\ i=\overline{1,s}.$$

Крок 2. Розв'язавши отриману систему алгебраїчних рівнянь якимось з відомих методів, отримаємо стаціонарні точки x функції Лагранжа.

Крок 3. Залучаючи необхідні умови більш високих порядків або достатні умови, або аналізуючи поведінку цільової функції в околі стаціонарних точок, виберемо серед стаціонарних точок точки мінімуму.

Метод множників Лагранжа для задачі умовної оптимізації з обмеженнями у формі нерівностей

Знайдемо розв'язок задачі

$$f(x) \to min \tag{1.24}$$

при умовах

$$x \in X = \left\{ x \in E^n : g_i(x) \leq 0, i = \overline{1,s} \right\}. \tag{1.25}$$

Припустимо, що функції $f(x), g_1(x), \ldots, g_s(x)$ диференційовні та $s < n$.

У цьому випадку застосовується такий алгоритм:

Алгоритм методу

Крок 1. Розглянемо задачу

$$f(x) \to min, \tag{1.26}$$

$$g_i(x) \leq 0, \quad i = \overline{1,s}. \tag{1.27}$$

Задачу (1.26), (1.27) розв'язуємо класичним методом, тобто знаходимо стаціонарні точки, які є розв'язками системи алгебраїчних рівнянь

$$\frac{\partial f(x)}{\partial x_k} = 0, \; k = \overline{1,n},$$

потім перевіряємо чи задовольняють ці точки умові (1.27). Якщо не задовольняють, виключаємо їх з подальшого розгляду, у протилежному випадку – досліджуємо чи є ці стаціонарні точки точками мінімуму цільової функції.

Крок 2. Розв'язуємо задачу

$$f(x) \to min,$$

$$g_i(x) = 0, \quad i = \overline{1,s}$$

методом множників Лагранжа для обмежень у формі рівностей.

Крок 3. Серед всіх знайдених точок вибираємо точки мінімуму.

Алгоритм описаний.

Приклад 1. Знайти мінімум функції

$$f(x, y) = 4x^2 + y^2, \text{ при умові } x + y = 5.$$

Складаємо функцію Лагранжа

$$L(x, y, \lambda) = 4x^2 + y^2 + \lambda(x + y - 5).$$

Складаємо систему рівнянь

$$\frac{\partial L(x, y, \lambda)}{\partial x} = 8x + \lambda = 0,$$

$$\frac{\partial L(x, y, \lambda)}{\partial y} = 2y + \lambda = 0,$$

$$\frac{\partial L(x, \lambda)}{\partial \lambda} = x + y - 5 = 0.$$

Розв’язуємо отриману систему

$$\begin{cases} 8x + \lambda = 0, \\ 2y + \lambda = 0, \\ x + y - 5 = 0, \end{cases} \Rightarrow \begin{cases} x = 1, \\ y = 4, \\ \lambda = -8. \end{cases}$$

Для дослідження отриманої стаціонарної точки застосуємо критерій Сильвестра. Згідно з цим критерієм, маємо, що цільова функція є опуклою і в точці $(1,4)$ має єдиний мінімум.

Приклад 2. Капітан Грант , який стоїть на береговій скалі висотою h метрів, кидає у море з початковою швидкістю v м/с пляшку з листом про допомогу. Під яким кутом до площини земної поверхні повинен бути здійснений кидок, щоб пляшка улетіла в море як можна далі? Опором повітря та ростом капітана можна знехтувати.

Фіксуємо систему координат (рис 1.8). Нехай $\left(x(t), y(t)\right)$ – траєкторія польоту пляшки, кинутої під кутом α до горизонту. Для довільної точки $M(x, y)$ маємо

$$x = vt\cos\alpha, \; y = vt\sin\alpha - \frac{gt^2}{2} + h.$$

Тоді

$$x(0) = 0, \; y(0) = h, \; x'(0) = v\cos\alpha, \; y'(0) = v\sin\alpha, \; x''(0) = 0, \; y''(0) = -g.$$

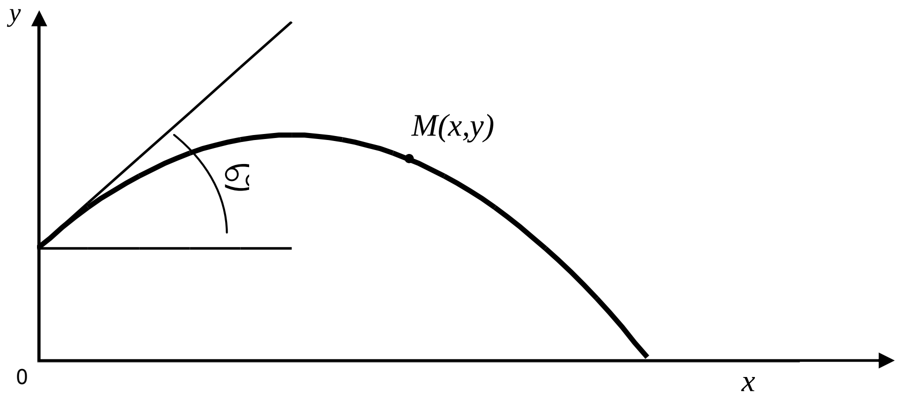


Рис. 1.8

Дальність польоту визначається моментом часу t_*, для якого $y(t_*)=0$. Отже, шуканий кут α^* і відповідний момент часу t_* утворюють розв'язок задачі:

$$vt\cos\alpha \to max,$$

$$vt\sin\alpha - \frac{gt^2}{2} + h = 0,$$

$$t \geq 0, \frac{-\pi}{2} \leq \alpha \leq \frac{\pi}{2}.$$

Знайдемо розв'язок задачі методом множників Лагранжа. Складемо функцію Лагранжа

$$L(x,y) = vt\cos\alpha + \lambda\left(vt\sin\alpha - \frac{gt^2}{2} + h\right).$$

Обчислимо частинні похідні функції Лагранжа за змінними t, α, λ. Зайдемо стаціонарні точки функції Лагранжа, розв'язавши систему трансцендентних рівнянь:

$$\frac{\partial L}{\partial t} = v\cos\alpha + \lambda\,(v\sin\alpha - gt) = 0,$$

$$\frac{\partial L}{\partial \alpha} = -vt\sin\alpha + \lambda vt\cos\alpha = 0,$$

$$\frac{\partial L}{\partial \lambda} = vt\sin\alpha - \frac{gt^2}{2} + h = 0.$$

Розділимо друге рівняння на vt та виразимо з нього λ. Отримаємо $\lambda = tg\alpha$,

$$\begin{cases} v\cos\alpha + tg\alpha(v\sin\alpha - gt) = 0 \\ vt\sin\alpha - \dfrac{gt^2}{2} + h = 0 \end{cases},$$

$$\begin{cases} v\cos^2\alpha + v\sin^2\alpha - gt\sin\alpha = 0 \\ vt\sin\alpha - \dfrac{gt^2}{2} + h = 0 \end{cases},$$

$$\begin{cases} t = \dfrac{v}{g\sin\alpha} \\ \sin\alpha = \dfrac{v}{\sqrt{2(v^2 + hg)}} \end{cases},$$

$$\begin{cases} \alpha = \arcsin\dfrac{v}{\sqrt{2(v^2 + hg)}} \\ t = \dfrac{v}{g\sin\alpha}, \lambda = tg\alpha \end{cases}.$$

§6. Основна задача опуклого програмування

Під задачею **опуклого програмування** розуміють задачу мінімізації опуклої цільової функції $f(x)$ на опуклій множині X:

$$f(x) \to min,\ x \in X, \tag{1.28}$$

де

$$X = \left\{x \in E^n : x \in X_0, g_i(x) \le 0, i = \overline{1,m},\quad g_i(x) = 0, i = \overline{m+1,s}\right\}. \tag{1.29}$$

Тут X_0 – задана опукла множина з E^n; функції $f(x), g_1(x), \ldots, g_m(x)$ визначені та опуклі на X_0; функції $g_i(x) = (a_i, x) - b_i$ – лінійні функції; b_i – задані числа; a_i – задані вектори з E^n при $i = \overline{m+1, s}$. Множина X_0 називається множиною простої структури. Це така множина, для якої можна легко (без трудомісткої обчислювальної роботи) перевірити включення $x \in X_0$, легко знайти проекцію довільної точки простору E^n на множину X_0. Прикладами множин простої структури є:

1) невід’ємний октант: $X_0 = \{x \in E^n : x_j \ge 0, j = \overline{1,n}\}$;

2) n-вимірний паралелепіпед: $X_0 = \{x \in E^n : a_j \le x_j \le b_j, j = \overline{1,n}\}$;

3) куля: $X_0 = \{x \in E^n : \|x - c\| \le R\}$;

4) гіперплощина: $X_0 = \{x \in E^n : (c, x) = d\}$.

Важливе місце в теорії опуклого програмування займає теорема Куна-Такера. Ця теорема дає необхідну та достатню умови оптимальності в задачі (1.28), (1.29) та є узагальненням методу множників Лагранжа. Для формулювання теореми Куна-Такера побудуємо функцію Лагранжа.

Введемо змінні $\lambda_1, \ldots, \lambda_s$ і побудуємо функцію Лагранжа

$$L(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x), \ x \in X_0. \tag{1.30}$$

Змінні $\lambda = (\lambda_1, \ldots, \lambda_s)$, які мають назву множників Лагранжа, належать множині $\Lambda_0 = \{\lambda \in E^s : \lambda_i \ge 0, i = \overline{1,s}\}$.

Точку $(x^*, \lambda^*) \in X_0 \times \Lambda_0$ називають **сідловою точкою** функції Лагранжа (1.30), якщо

$$L(x_*, \lambda) \le L(x_*, \lambda^*) \le L(x, \lambda^*) \text{ при будь-яких } x \in X_0 \text{ та } \lambda \in \Lambda_0.$$

Відзначимо, що

$$L(x_*, \lambda^*) = \min_{x \in X_0} \max_{\lambda \in \Lambda_0} L(x, \lambda) = \max_{\lambda \in \Lambda_0} \min_{x \in X_0} L(x, \lambda).$$

Ця форма запису відповідає теоремі про мінімакс. Її можна використовувати для вивчення двоїстості в нелінійному програмуванні.

Обмеження $g_i(x) \le 0$ називається **регулярним** на множині X_0, якщо існує точка $z \in X_0$, для якої

$$g_i(z) < 0.$$

Якщо X_0 – опукла множина, функції $g_i(x)$, $i = \overline{1,m}$ опуклі на X_0, то існує точка $\bar{x} \in X_0$, для якої

$$g_i(\bar{x}) < 0 \text{ для будь-яких } i = \overline{1,m}. \tag{1.31}$$

Умову (1.31) називають **умовою регулярності Слейтера**.

Теорема Куна-Такера. Нехай для задачі (1.28), (1.29) виконується умова регулярності Слейтера (1.31). Для того, щоб точка $x_* \in X$ була точкою мінімуму функції $f(x)$ на множині X, необхідно і достатньо, щоб існував такий вектор $\lambda^* \in \Lambda_0$ такий, що точка (x_*, λ^*) була б сідловою точкою функції Лагранжа $L(x,\lambda)$ на множині $X_0 \times \Lambda_0$.

Зауваження 1. Якщо всі функції $g_i(x)$, які задають обмеження (1.29), лінійні, тоді теорема про існування сідлової точки справедлива без додаткових умов типу умов регулярності.

Зауваження 2. Теорема Куна-Такера дає можливість перейти від вихідної задачі опуклого програмування з складними обмеженнями до задачі умовної оптимізації з простими обмеженнями.

Обмеження $g_i(x) \le 0$ називається **активним** у точці x, коли $g_i(x) = 0$.

Теорема Куна-Такера в диференціальній формі.

Нехай $X_0 = \left\{x \in E^n : x_j \ge 0, j = \overline{1,n}\right\}$ і функції $f(x)$, $g_i(x)$, $i = \overline{1,m}$ опуклі та неперервно диференційовні на опуклій множині X_0. Для того, щоб точка (x_*, λ^*) була сідловою точкою функції Лагранжа $L(x,\lambda)$ на множині $X_0 \times \Lambda_0$, необхідно і достатньо, щоб

$$\frac{\partial L(x_*,\lambda^*)}{\partial x} \ge 0, \qquad \frac{\partial L(x,\lambda)}{\partial \lambda} \le 0, \tag{1.32}$$

$$\left(x_*, \frac{\partial L(x_*,\lambda^*)}{\partial x}\right) = 0, \qquad \left(\lambda^*, \frac{\partial L(x_*,\lambda^*)}{\partial \lambda}\right) = 0, \tag{1.33}$$

$$x_* \ge 0,\ \lambda^* \ge 0. \tag{1.34}$$

Умови $\left(\lambda^*, \dfrac{\partial L(x_*,\lambda^*)}{\partial \lambda}\right) = \left(\lambda^*, g(x_*)\right) = 0$ називаються **умовами доповнюючої нежорсткості**. Вони вимагають, щоб $\lambda_i^* = 0$, коли відповідне обмеження не активне, і $\lambda_i^* > 0$ для активних обмежень.

Запишемо умови Куна-Такера для задачі квадратичного програмування:

$$f(x)=\frac{1}{2}(Ax,x)-(b,x)\rightarrow min, \tag{1.35}$$

$$Cx\le d\ ,\ x\ge 0, \tag{1.36}$$

де A – симетрична, додатно визначена матриця, $b\in E^n$, C – матриця $m\times n$, $d\in E^m$.

Теорема. Точка x_* є оптимальним розв’язком задачі квадратичного програмування тоді і тільки тоді, коли існують такі m-вимірні вектори $\lambda,w\ge 0$ та n-вимірний вектор $v\ge 0$, що виконують такі умови:

$$b+Ax_*-C^T\lambda+v=0, \tag{1.37}$$

$$d-Cx_*-w=0, \tag{1.38}$$

$$v^T x_*=0,\ w^T\lambda=0. \tag{1.39}$$

Вправи до розділу 1

1. Використовуючи визначення опуклої множини, довести, що прямокутник $X=\{x\in E^2 : 0\le x_1\le a,\quad 0\le x_2\le b\}$ є опуклою множиною.
2. Чи є множина X опуклою?
   - $X=\{(x_1,x_2)\in E^2 : x_1-x_2^2\le 0,\quad x_2-x_1^2\le 0\}$,
   - $X=\{(x_1,x_2)\in E^2 : x_1^2+x_2^2\le 1,\quad x_1-x_2^2\ge 0\}$.
3. Знайти мінімальне значення k при якому множина X є опуклою: $X=\{(x_1,x_2)\in E^2 : (x_1^2+1)x_2\le 5,\quad x_2\ge k\}$.
4. Довести, що множина $X=\{(x_1,x_2)\in E^2 : x_1x_2\ge 1,\quad x_2>0\}$ опукла.
5. Довести, що множина $X=\{(x_1,x_2)\in E^2 : x_1^2\le x_2\}$ опукла.
6. Показати, що такі функції є опуклими:
   - $f(x)=2x_1^2+x_2^2-x_1x_2+5x_1-6x_2+8$,
   - $f(x)=\sqrt{x_1x_2}$.
7. Дослідити опуклість функцій:

- $f(x)=\dfrac{2}{x}, x>0$,
- $f(x)=exp(2x_1-x_2)$,
- $f(x)=5-x_1^2-x_2^2$,
- $f(x)=2x_1-x_1x_2$,
- $f(x)=-\dfrac{1}{x_1}-\dfrac{1}{x_2}, x_1>0, x_2>0$.

8. Дослідити властивості опуклості функції $f(x)=ax^2$.
9. При яких значеннях a, b, c функція $f(x)$ буде опуклою у просторі E^2:

   $f(x)=ax_1^2+2bx_1x_2+cx_2^2$.
10. Чи є функція $f(x)$ строго квазіопуклою? Використати геометричні побудування:
    - $f(x)=exp(-x)+x^2$,
    - $f(x)=3x-2x^2+x^3+2x^4$.
11. Знайти області опуклості та увігнутості функцій:
    - $f(x)=sin(x_1+x_2+x_3)$,
    - $f(x)=sin(x_1^2+x_2^2+x_3^2)$.
12. Якщо функція $f(x)$ є опуклою, чи буде функція $|f(x)|$ опуклою?
13. Якщо $f_1(x)$, $f_2(x)$ є опуклими, то чи буде опуклим їхній добуток? Розглянути на прикладі функцій $f_1(x)=x$, $f_2(x)=x^2$. Що зміниться, коли від $f_1(x)$, $f_2(x)$ вимагати невід'ємності? Або монотонності?
14. Довести, що $f(x)$ квазіопукла на X тоді і тільки тоді, коли множина $D(y)=\{x\in X : f(x)\le f(y)\}$ опукла для будь-яких $x\in X$.
15. Розв'язати задачу Герона. На площині задані точки *A* та *B* по один бік від прямої *l*. Нехай *C* – така точка на прямій *l*, що сума відстаней від *A* до *C* та від *C* до *B* є найменшою. Довести, що кути між прямою *l* та відрізками *AC* та *BC* є однаковими.
16. Дослідити поведінку функції $f(x,y)=x^2-y^4$ біля початку координат.
17. Розв'язати задачу Пеано. Задана функція

$$f(x,y)=abx^4-(a+b)x^2y+y^2,\ a>0,\ b>0.$$

Знайти вектор-градієнт та матрицю Гессе для початкової точки. Чи є початок координат екстремумом функції? Яке значення має похідна за напрямком поблизу початку координат? (*Вказівка*. Рекомендується розкласти функції на множники).

18. Знайти найкоротшу відстань від еліпсу $2x^2+3y^2=12$ до прямої $x+y=6$.

19. Довести такі правила субдиференцювання функцій:

- коли $g(x)=f(x+x_0)$, то $\partial g(x)=\partial f(x+x_0)$;
- коли $g(x)=\lambda f(x)$, $\lambda>0$, то $\partial g(x)=\lambda\partial f(x)$;
- коли $g(x)=f(\lambda x)$, $\lambda>0$, то $\partial g(x)=\lambda\partial f(\lambda x)$;
- коли $f_i(x)$, $i=\overline{1,m}$ – опуклі функції на E^n, то

$$\partial\left(\sum_{i=1}^{m} f_i(x)\right)=\sum_{i=1}^{m}\partial f_i(x),\ x\in E^n;$$

- коли $f(x)$ – опукла функції на E^n, а A – матриця порядку $m\times n$, то функція $g(x)=f(Ax)$, $x\in E^n$, опукла на E^n і $\partial g(x)=A^T\partial f(y)\big|_{y=Ax}$.

20. Знайти субдиференціали функцій:

- $f(x)=|x-1|$, $x\in E^1$;
- $f(x)=|x-1|+|x+1|$, $x\in E^1$;
- $f(x)=|x_1+x_2|+|x_1-x_2|$, $x\in E^2$;
- $f(x)=max\{x^2,(x+2)\}$, $x\in E^1$;
- $f(x)=max\{|x|,|x-1|\}$, $x\in E^1$;
- $f(x)=|(a,x)-b|$, $x\in E^n$;
- $f(x)=\max\limits_{|t|\le 1}\{|t^2+x_1t+x_2|\}$, $x\in E^2$;
- $f(x)=\max\limits_{|t|\le 1}\{|t^2x_1+tx_2|\}$, $x\in E^2$;
- $f(x)=\max\limits_{0\le|t|\le 1}\{|x_1+t\,x_2|\}$, $x\in E^2$;

- $f(x) = \max\limits_{0 \le |t| \le 1} \{ |x_1 + (1-t)x_2| \}, \ x \in E^2$;

21. Довести, що функція $f(x) = \max\limits_{1` \le i \le m} \{ f_i(x) \}, \ x \in E^n$ опукла на E^n, коли функції $f_1(x), \dots, f_m(x)$ опуклі на E^n. Знайти субдиференціал функції $f(x)$.
22. Визначити умовні екстремуми функцій методом множників Лагранжа. Дати геометричну ілюстрацію:
   - $z = x^2 + y^2$ при $x + y = 1$;
   - $z = 3x^2 + 2y^2 - 3x + 1$ при $x^2 + y^2 = 4$;
   - $z = x^2 - y^2$ при $x - y = 4$;
   - $z = (x-3)^2 + (y-5)^2$ при $y = 2x - 5$;
   - $z = (x-2)^2 + (y-3)^2$ при $x^2 + y^2 \le 52$.
23. Знайти за допомогою методу множників Лагранжа стаціонарні точки при дослідженні умовного екстремуму функцій:
   - $f = x + y + z$ при $\frac{1}{x} + \frac{1}{y} + \frac{1}{z} = 1$;
   - $f = \frac{1}{x} + \frac{1}{y}$ при $\frac{1}{x^2} + \frac{1}{y^2} = 1$;
   - $f = xyzt$ при $x + y + zt = 4$;
   - $f = 2x + 3y^2 + z^2$ при $x + y + z = 8$;
   - $f = xyz$ при $x + y + z = 6$, $xy + yz + xz = 12$;
   - $f = x^2 + y^2 + z^2$ при $x + y + z \le 12$, $x \ge 0$, $y \ge 0$, $z \ge 0$;
   - $f = xyz$ при $x + y + z \le 6$, $xy + yz + xz \le 8$;
   - $f = x - 2y + 2z$ при $x^2 + y^2 + z^2 \le 9$;
   - $f = x^2 + 4y^2 - 4xy - 2yz - 2xz$ при $3x^2 + 6y^2 + 2z^2 \le 1$;
   - $f = xy$ при $x^2 + y^2 = 2$;

- $f = x + y$ при $\frac{1}{x} + \frac{1}{y} = 1$;
- $f = x^3 + y^3$ при $x + y = 12$, $x \ge 0$, $y \ge 0$.

24. Знайти точки локальних і глобальних екстремумів таких функцій на всій області їхнього визначення:
    - $f = x^2 + 2y^2 + 5z^2 - 2xy - 4yz - 2z$;
    - $f = ae^{-x} + be^{-x} + ln(e^x + e^x)$, $a > 0$, $b > 0$;
    - $f = x^4 + y^4 - 4xy$;
    - $f = (x + y)(x - a)(x - b)$;
    - $f = x^2 + y^4 - y^5 - 2xy^2$;
    - $f = x + y + 4 sin\, x\, sin\, y$;
    - $f = xe^x - (1 + e^x) cos\, y$;
    - $f = x^3 + y^2 + z^2 + 12xy + 2z$;
    - $f = xy^2z^3(1 - x - 2y - 3z)$.
25. На площині задані точки *A*, *B*, *C*. Знайти точку, сума відстаней від якої до трьох заданих точок мінімальна.
26. Задача Евкліда. У даний трикутник *ABC* вписати паралелограм *ADEF* (причому $DE \| AC$, $FE \| AB$) найбільшої площини.
27. Розділити задане додатне число на дві частини так, щоб добуток добутку цих частин на їхню різницю був максимальним.

Розділ 2. Чисельні методи одновимірної оптимізації

Необхідність окремого розгляду чисельних методів пошуку екстремуму функцій однієї змінної визначається такими обставинами. По–перше, ці методи використовуються в багатьох алгоритмах пошуку екстремуму функцій, які залежать від кількох змінних. По-друге, класи функцій однієї змінної служать зручною моделлю для теоретичного дослідження ефективності методів оптимізації. По-третє, іноді вдається, використовуючи ті чи інші прийоми, безпосередньо за допомогою алгоритмів одновимірної оптимізації отримати розв'язок багатовимірних задач.

Зауважимо, що універсальних методів, які були б придатні для мінімізації довільних функцій однієї змінної, не існує. Тому доводиться будувати алгоритми, які орієнтовані на різні класи функцій, що зустрічаються у прикладних задачах.

Розглянемо методи одновимірної мінімізації строго квазіопуклих (унімодальних) функцій, тобто функцій, які мають єдиний мінімум (див. розділ 1, §2); а також метод дотичних, який застосовується для знаходження точки мінімуму опуклої функції.

Постановка задачі

Розглянемо задачу мінімізації строго квазіопуклої функції $f(x)$, яка досягає на множині $X = \{x : a \le x \le b\}$ своєї нижньої границі :

$$f(x) \to min, \ x \in [a,b].$$

Ідея методів ділення відрізку навпіл, золотого перерізу, Фібоначчі

Суть методів полягає в побудові послідовності вкладених відрізків, які містять точку мінімуму x_* :

$$[a,b] \supset \ldots \supset [a_k, b_k] \supset \ldots \supset [a_n, b_n],$$

$x_* \in [a_k, b_k]$ для будь-якого k .

Належність точки x_* кожному з цих відрізків забезпечується тим, що функція $f(x)$ є строго квазіопуклою.

§1. Метод ділення відрізку навпіл (метод дихотомії)

На відрізку $[a,b]=[a_1,b_1]$ вибирають дві точки

$$x_1=\lambda_1=\frac{a+b-\delta}{2},\ x_2=\mu_1=\frac{a+b+\delta}{2}=a+b-\lambda_1,$$

де δ – стала, $0<\delta<b-a$. Величина δ вибирається обчислювачем і може характеризувати похибку вимірювання величини x і обмежена знизу можливостями вимірювального приладу. Точки λ_1 та μ_1 розташовані симетрично на відрізку $[a,b]$ відносно його середини та при малих значеннях δ поділяють його майже навпіл – цим і пояснюється назва методу.

Далі обчислюємо та порівнюємо значення цільової функції $f(\lambda_1)$ та $f(\mu_1)$. Якщо $f(\lambda_1)\le f(\mu_1)$, то покладаємо $a_2=a$, $b_2=\mu_1$. Якщо ж $f(\lambda_1)>f(\mu_1)$, то покладаємо $a_2=\lambda_1$, $b_2=b$. У результаті отримуємо відрізок $[a_2,b_2]$, який містить точку мінімуму функції $f(x)$ на $[a,b]$ і має довжину

$$b_2-a_2=\frac{b-a-\delta}{2^1}+\delta.$$

Нехай відрізок $[a_k,b_k]$, який містить точку мінімуму функції $f(x)$ на $[a,b]$ уже відомий і має довжину

$$b_k-a_k=\frac{b-a-\delta}{2^{k-1}}+\delta>\delta,\ k\ge 2.$$

Тоді виберемо точки

$$\lambda_k=\frac{a_k+b_k-\delta}{2},\ \mu_k=\frac{a_k+b_k+\delta}{2}=a_k+b_k-\lambda_k, \tag{2.1}$$

і обчислимо значення цільової функції $f(\lambda_k)$ та $f(\mu_k)$. Якщо $f(\lambda_k)\le f(\mu_k)$, то покладаємо $a_{k+1}=a_k$, $b_{k+1}=\mu_k$ (рис. 2.1,*а*). Якщо ж $f(\lambda_k)>f(\mu_k)$, то покладаємо $a_{k+1}=\lambda_k$, $b_{k+1}=b_k$ (рис. 2.1,*б*). Довжина отриманого відрізку $[a_{k+1},b_{k+1}]$ дорівнює

$$b_{k+1}-a_{k+1}=\frac{b-a-\delta}{2^k}+\delta>\delta.$$

Відрізок $[a_{k+1},b_{k+1}]$ містить точку мінімуму функції $f(x)$ на $[a,b]$.

Коли кількість обчислень значень цільової функції нічим не обмежена, то описаний процес ділення відрізку навпіл можна продовжувати до тих пір, поки не отримаємо відрізок $[a_{k+1}, b_{k+1}]$, який має довжину $b_{k+1} - a_{k+1} < \varepsilon$, де ε – задана точність, $\varepsilon > \delta$.

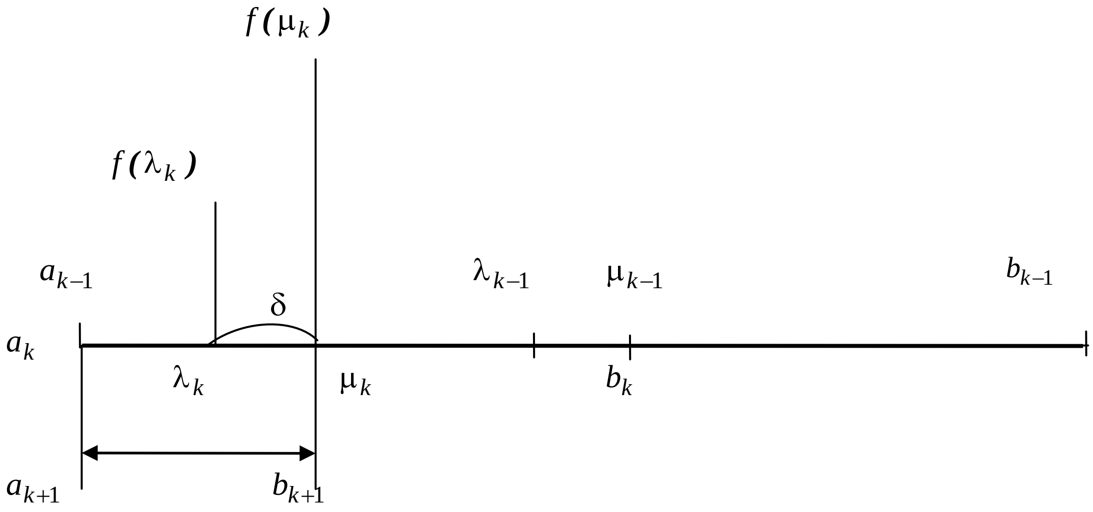


Рис. 2.1, *а*

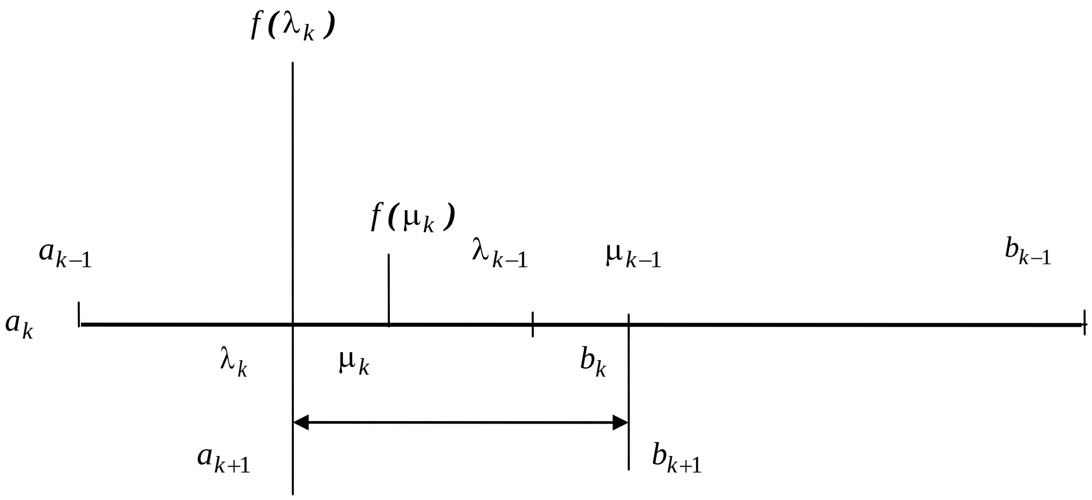


Рис. 2.1, *б*

Кількість ітерацій $k > \log_2 \dfrac{b - a - \delta}{\varepsilon - \delta}$. Так як кожне ділення навпіл потребує двох обчислень значення цільової функції, то для досягнення точності $b_{k+1} - a_{k+1} < \varepsilon$ потрібно всього $n = 2k > 2\log_2 \dfrac{b - a - \delta}{\varepsilon - \delta}$ таких обчислень.

Коли відрізок $[a_{k+1}, b_{k+1}]$ вже визначений, то за точку мінімуму x_* можна взяти точку $\overline{x_n} = \lambda_k$ при $f(\lambda_k) \le f(\mu_k)$ і $\overline{x_n} = \mu_k$ при $f(\lambda_k) > f(\mu_k)$, а значення $f(\overline{x_n})$ може служити наближенням для $f(x_*)$, При такому виборі наближення до точки мінімуму буде допущена похибка

$$\left|x_* - \overline{x_n}\right| \le max\left\{b_{k+1} - \overline{x_n}, \overline{x_n} - a_{k+1}\right\} = \frac{b - a - \delta}{2^k}. \qquad (2.2)$$

Якщо не потрібно обчислювати значення цільової функції в точці $\overline{x_n}$, то замість $\overline{x_n}$ можна взяти точку $y_n = \dfrac{a_{k+1} + b_{k+1}}{2}$ з меншою похибкою

$$\left|x_* - y_n\right| \le \frac{b_{k+1} - a_{k+1}}{2} = \frac{b - a - \delta}{2^{k+1}} + \frac{\delta}{2}. \qquad (2.3)$$

Алгоритм методу ділення відрізка навпіл (методу дихотомії)

Початковий етап. Нехай $[a_1, b_1]$ – відрізок, на якому визначена функція $f(x)$. Вибрати $\delta > 0$ та задати точність $\varepsilon > 0$, причому $\varepsilon > \delta$. Покласти $k = 1$ і перейти до основного етапу.

Основний етап

Крок 1. Якщо $b_k - a_k < \varepsilon$, то зупинитися; точка мінімуму x_* належить інтервалу $[a_k, b_k]$. За x_* прийняти $\dfrac{a_{k+1} + b_{k+1}}{2}$. Інакше, обчислити

$$\lambda_k = \frac{a_k + b_k - \delta}{2}, \quad \mu_k = \frac{a_k + b_k + \delta}{2} = a_k + b_k - \lambda_k$$

і перейти до кроку 2.

Крок 2. Обчислити значення цільової функції $f(\lambda_k)$ та $f(\mu_k)$. Якщо $f(\lambda_k) \le f(\mu_k)$, то покладаємо $a_{k+1} = a_k$, $b_{k+1} = \mu_k$, інакше $a_{k+1} = \lambda_k$, $b_{k+1} = b_k$. Замінити k на $k+1$ і перейти до кроку 1.

Опис алгоритму закінчено.

Приклад. Знайти мінімум функції $f(x)$ методом дихотомії.

$$f(x) = x^2 + 2x \to min, \ x \in [3,5].$$

Нехай $\varepsilon = 0.2$, $\delta = 0.1$. Визначимо перші дві точки:

$$\lambda_1 = \frac{-3+5-0.1}{2} = 0.95, \ \mu_1 = \frac{-3+5+0.1}{2} = 1.05.$$

Так як $f(\lambda_1) \leq f(\mu_1)$, то новий інтервал $[a_2, b_2] = [-3, 1.05]$. Цей процес повторюється згідно сформульованому алгоритму та результати обчислень наведені в таблиці 2.1.

Таблиця 2.1

k	a_k	b_k	λ_k	μ_k	$f(\lambda_k)$	$f(\mu_k)$
1	–3.000	5.000	0.950	1.050	2.803	3.203
2	–3.000	1.050	–1.025	–0.925	–0.999	–0.994
3	–3.000	–0.925	–2.013	–1.913	0.025	–0.167
4	–1.913	–0.925	–1.469	–1.369	–0.779	–0.863
5	–1.369	–0.925	–1.197	–1.197	–0.060	–0.991
6	–1.197	–0.925	–1.111	–1.011	–0.988	–1.000
7	–1.011	–0.925				

Для знаходження точки мінімуму з точністю ε потрібно 12 разів обчислювати значення цільової функції. За наближення до точки мінімуму можна узяти точку $\overline{x_n} = -1.011$, $f(\overline{x_n}) = -1.000$ або точку $y_n = \frac{a_7 + b_7}{2} = -0.968$.

§2. Метод золотого перерізу

Золотим перерізом відрізка називають його ділення на дві частини так, щоб відношення довжини всього відрізка до довжини більшої частини дорівнювало відношенню довжини більшої частини до довжини меншої. Безпосередньо можна перевірити, що золотий переріз відрізка $[a, b]$ роблять дві симетрично розташовані відносно його середини $[a, b]$ точки:

$$x_1 = \lambda_1 = a + \frac{3-\sqrt{5}}{2}(b-a), \ x_2 = \mu_1 = a + \frac{\sqrt{5}-1}{2}(b-a). \tag{2.4}$$

Виявляється, що точка λ_1, у свою чергу, робить золотий переріз відрізка $[a, \mu_1]$, а точка μ_1 – золотий переріз відрізка $[\lambda_1, b]$.

Покладемо $a_1 = a$, $b_1 = b$. На відрізку $[a_1, b_1]$ візьмемо точки λ_1 та μ_1 за формулами (2.4) та обчислимо значення цільової функції $f(\lambda_1)$ та $f(\mu_1)$. Якщо $f(\lambda_1) \le f(\mu_1)$, то покладаємо $a_2 = a_1$, $b_2 = \mu_1$, $\overline{x_2} = \lambda_1$. Якщо ж $f(\lambda_1) > f(\mu_1)$, то покладаємо $a_2 = \lambda_1$, $b_2 = b_1$, $\overline{x_2} = \mu_1$. У результаті отримаємо відрізок $[a_2, b_2]$, який містить точку мінімуму функції $f(x)$ на $[a,b]$ і має довжину

$$b_2 - a_2 = \frac{\sqrt{5}-1}{2}(b-a).$$

Точка $\overline{x_2}$ є наближеним значенням точки мінімуму x_* і робить золотий переріз відрізка $[a_2, b_2]$, наближене оптимальне значення цільової функції $f(\overline{x_2}) = min\{\, f(\lambda_1),\ f(\mu_1)\}$.

Припустимо, що точки $x_1, \ldots, x_n$ уже визначені. Нехай відрізок $[a_n, b_n]$ має довжину

$$b_n - a_n = \left(\frac{\sqrt{5}-1}{2}\right)^{n-1}(b-a),$$

і містить точку мінімуму x_* функції $f(x)$ на $[a,b]$, також визначена точка $\overline{x_n}$, яка робить золотий переріз відрізка $[a_n, b_n]$, така, що $f(\overline{x_n}) = \min\limits_{1 \le i \le n}\{\, f(x_i)\}$. Тоді за наступну точку візьмемо точку $x_{n+1} = a_n + b_n - \overline{x_n}$ (рис. 2.2), яка також робить золотий переріз відрізка $[a_n, b_n]$. Обчислимо значення цільової функції $f(x_{n+1})$. Нехай для визначеності $a_n < x_{n+1} < \overline{x_n} < b_n$.

Якщо:

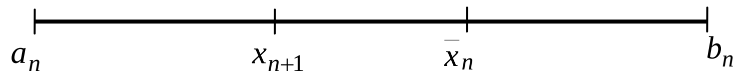
1) $f(x_{n+1}) \le f(\overline{x_n})$, то покладаємо $a_{n+1} = a_n$, $b_{n+1} = \overline{x_n}$, $\overline{x_{n+1}} = x_{n+1}$;
2) $f(x_{n+1}) > f(\overline{x_n})$, то покладаємо $a_{n+1} = x_{n+1}$, $b_{n+1} = b_n$, $\overline{x_{n+1}} = \overline{x_n}$.

Довжина отриманого відрізка $[a_{n+1}, b_{n+1}]$ дорівнює

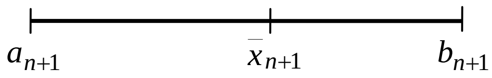
$$b_{n+1} - a_{n+1} = \left(\frac{\sqrt{5}-1}{2}\right)^{n}(b-a),$$

точка $\overline{x_{n+1}}$ робить золотий переріз відрізка $[a_{n+1}, b_{n+1}]$ та

$$f\left(\overline{x_{n+1}}\right)=min\left\{f\left(x_{n+1}\right),\ f\left(\overline{x_n}\right)\right\}=\min_{1\le i\le n+1} f\left(x_i\right).$$



Випадок 1:



Випадок 2:

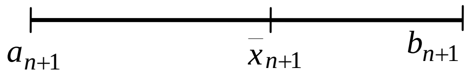


Рис. 2.2

Коли число обчислень значень цільової функції $f\left(x\right)$ заздалегідь не обмежене, то описаний процес можна продовжувати, наприклад, до тих пір, поки не буде виконана нерівність $b_n - a_n < \varepsilon$, де ε – задана точність. Коли число обчислень значень цільової функції $f\left(x\right)$ задане і дорівнює n, то процес закінчується і за розв’язок приймається або точка $\overline{x_n}$ з похибкою

$$\left|x_* - \overline{x_n}\right| \le max\left\{b_n - \overline{x_n},\ \overline{x_n} - a_n\right\}=\left(\frac{\sqrt{5}-1}{2}\right)^n (b-a), \tag{2.5}$$

або (якщо не потрібно обчислювати значення цільової функції в точці $\overline{x_n}$) точка

$$v_n = \frac{a_n + b_n}{2}$$

з похибкою

$$\left|x_* - v_n\right| \le \left(\frac{\sqrt{5}-1}{2}\right)^{n-1} \frac{(b-a)}{2}. \tag{2.6}$$

Рекомендації до чисельної реалізації методу золотого перерізу на ЕОМ

На кожному відрізку $\left[a_n, b_n\right]$, який містить точку $\overline{x_n}$ з попереднього кроку і яку знайдено з похибкою через наближене обчислення числа $\sqrt{5}$, при виборі наступної точки x_{n+1} треба остерігатися користування формулою

$x_{n+1} = a_n + b_n - \overline{x_n}$, тому що тоді похибка з попереднього кроку переноситься на даний крок. Замість цього краще безпосередньо провести золотий переріз відрізка $[a_n, b_n]$ і за точку x_{n+1} узяти ту з точок $a_n + \frac{3-\sqrt{5}}{2}(b_n - a_n)$, $a_n + \frac{\sqrt{5}-1}{2}(b_n - a_n)$, яка найбільш віддалена від $\overline{x_n}$.

Алгоритм методу золотого перерізу

Початковий етап. Нехай $[a_1, b_1]$ – відрізок, на якому визначена функція $f(x)$. Задати точність $\varepsilon > 0$. Обчислити

$$\lambda_1 = a_1 + \frac{3-\sqrt{5}}{2}(b_1 - a_1),\ \mu_1 = a_1 + \frac{\sqrt{5}-1}{2}(b_1 - a_1),$$

значення цільової функції $f(\lambda_1)$ та $f(\mu_1)$, покласти $n=1$ і перейти до основного етапу.

Основний етап

Крок 1. Якщо $b_n - a_n < \varepsilon$, то зупинитися; точка мінімуму x_* належить інтервалу $[a_n, b_n]$. За x_* прийняти $v_n = \frac{a_n + b_n}{2}$. Інакше, якщо $f(\lambda_k) > f(\mu_k)$, то перейти до кроку 2, а якщо $f(\lambda_k) \le f(\mu_k)$, то перейти до кроку 3.

Крок 2. Покласти

$$a_{n+1} = \lambda_n,\ b_{n+1} = b_n,\ \lambda_{n+1} = \mu_n,\ \mu_{n+1} = a_{n+1} + \frac{\sqrt{5}-1}{2}(b_{n+1} - a_{n+1}).$$

Обчислити значення цільової функції $f(\mu_{n+1})$ і перейти до кроку 4.

Крок 3. Покласти $a_{n+1} = a_n,\ b_{n+1} = \mu_n,\ \mu_{n+1} = \lambda_n$,

$$\lambda_{n+1} = a_{n+1} + \frac{3-\sqrt{5}}{2}(b_{n+1} - a_{n+1}).$$

Обчислити значення цільової функції $f(\lambda_{n+1})$ і перейти до кроку 4.

Крок 4. Замінити n на $n+1$ і перейти до кроку 1.

Опис алгоритму закінчений.

Приклад. Знайти мінімум функції $f(x)$ методом золотого перерізу.

$$f(x) = x^2 + 2x \to min,\ x \in [3,5].$$

Очевидно, що цільова функція $f(x)$ є строго квазіопуклою. Знайдемо точку мінімуму функції $f(x)$ з точністю $\varepsilon = 0.2$. Результати обчислень наведені в таблиці 2.2. За вісім ітерацій, які містять дев'ять обчислень значень цільової функції, отримано відрізок $[a_n, b_n] = [-1.112, -0.936]$ і за точку мінімуму може бути узята точка $v_n = \frac{a_9 + b_9}{2} = -1.024$.

Таблиця 2.2

k	a_k	b_k	λ_k	μ_k	$f(\lambda_k)$	$f(\mu_k)$
1	–3.000	5.000	0.056	1.944	<u>0.115</u>	<u>7.667</u>
2	–3.000	1.944	–1.112	0.056	<u>–0.987</u>	0.115
3	–3.000	0.056	–1.832	–1.112	<u>–0.308</u>	–0.987
4	–1.832	0.056	–1.112	–0.664	–0.987	<u>–0.887</u>
5	–1.832	–0.664	–1.384	–1.112	<u>–0.853</u>	–0.987
6	–1.384	–0.664	–1.112	–0.936	–0.987	<u>–0.996</u>
7	–1.384	–0.936	–1.208	–1.112	<u>–0.957</u>	–0.987
8	–1.208	–0.936	–1.112	–1.032	–0.987	–0.999
9	–1.112	–0.936				

§3. Метод Фібоначчі

Подібно до методу золотого перерізу метод Фібоначчі потребує двох обчислень цільової функції на першій ітерації, а на кожній наступній ітерації тільки по одному обчисленню значення цільової функції. Нагадаємо, що в методі ділення відрізка навпіл на кожній ітерації проводиться по два обчислення значень цільової функції. Але, на відміну від методів ділення відрізка навпіл та методу золотого перерізу, метод Фібоначчі потребує попереднього завдання числа n обчислень значень цільової функції.

Метод Фібоначчі пов'язаний з числами Фібоначчі. Будемо позначати метод Фібоначчі через Φ_n. Відомо, що числа Фібоначчі визначаються співвідношеннями

$$F_{n+2} = F_{n+1} + F_n, \ n = 1,2,3,\ldots; \ F_1 = F_2 = 1. \tag{2.7}$$

За допомогою індукції легко показати, що n-е число Фібоначчі представимо у вигляді:

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right). \tag{2.8}$$

З формули (2.8) витікає, що для великих n

$$F_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

При $n = 1$ отримаємо метод Φ_1:

$x_1 = \dfrac{a+b}{2}$, за точку мінімуму x_* виберемо точку $\overline{x_1}$, допускаючи при цьому похибку $\left|x_* - \overline{x_1}\right| \le \dfrac{b-a}{2} = \dfrac{b-a}{F_3}$.

Нехай $n = 2$. Тоді метод Φ_2 починається з вибору двох точок

$$x_1 = a + \frac{F_n}{F_{n+2}}(b-a), \ x_2 = a + \frac{F_{n+1}}{F_{n+2}}(b-a) = a + b - x_1,$$

симетричних одна одній відносно середини відрізка $[a,b]$.

Якщо $f(x_1) \le f(x_2)$, то покладаємо $a_2 = a$, $b_2 = x_2$, $\overline{x_2} = x_1$. Якщо ж $f(x_1) > f(x_2)$, то покладаємо $a_2 = x_1$, $b_2 = b$, $\overline{x_2} = x_2$. У результаті отримаємо відрізок $[a_2, b_2]$, який містить точку мінімуму функції $f(x)$ на $[a,b]$ і має довжину

$$b_2 - a_2 = b - x_1 = x_2 - a = \frac{F_{n+1}}{F_{n+2}}(b-a).$$

Точка $\overline{x_2}$ є наближеним значенням точки мінімуму x_* і має такі властивості:

1) $a_2 < \overline{x_2} < b_2$ за способом вибору точки $\overline{x_2}$;

2) $f\left(\overline{x_2}\right) = \min\limits_{i=1,2} f\left(x_i\right);$

3) $\overline{x_2}$ співпадає з однією з точок x_1 або x_2 на відрізку $[a,b]$;

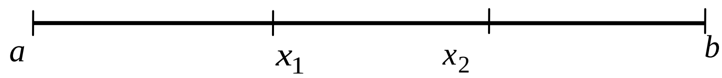
4) $\overline{x_2}$ співпадає з однією з точок $x_2^{'}$ або $x_2^{''}$:

$$x_2^{'} = a_2 + \frac{F_n}{F_{n+1}}(b_2 - a_2) = a_2 + \frac{F_n}{F_{n+2}}(b-a), \text{ коли } f\left(x_1\right) \le f\left(x_2\right),$$

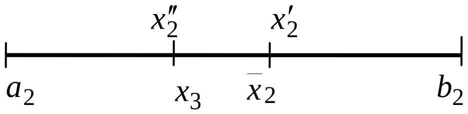
$$x_2^{''} = a_2 + \frac{F_{n-1}}{F_{n+1}}(b_2 - a_2) = a_2 + \frac{F_{n-1}}{F_{n+2}}(b-a), \text{ коли } f\left(x_1\right) > f\left(x_2\right),$$

симетричних одна одній на відрізку $[a_2, b_2]$ відносно його середини (рис. 2.3).

$$f\left(\overline{x_{n+1}}\right) = \min\left\{ f\left(x_{n+1}\right),\ f\left(\overline{x_n}\right)\right\} = \min_{1 \le i \le n+1} f\left(x_i\right).$$



Випадок 1: $f\left(x_1\right) \le f\left(x_2\right)$



Випадок 2: $f\left(x_1\right) > f\left(x_2\right)$

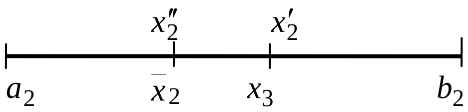


Рис. 2.3

Далі на відрізку $[a_2, b_2]$ вибираємо наступну точку x_3: $x_3 = a_2 + b_2 - \overline{x_2}$, яка співпадає з однією з точок $x_2^{'}$ або $x_2^{''}$, відмінної від $\overline{x_2}$. Порівнюючи $f\left(x_3\right)$ з $f\left(\overline{x_2}\right)$, знаходимо новий відрізок $[a_3, b_3]$ і так далі.

У загальному випадку нехай точки $x_1, \ldots, x_k$, $2 \le k \le n$ вже вибрані, нехай знайдено відрізок $[a_k, b_k]$ довжини

$$b_k - a_k = \frac{F_{n-k+3}}{F_{n+2}}(b-a),$$

який містить точку мінімуму функції $f(x)$ на $[a,b]$ і точку $\overline{x_k}$, яка має такі властивості:

1) $a_k < \overline{x_k} < b_k$;

2) $f\left(\overline{x_k}\right) = \min\limits_{1 \le i \le k} f(x_i)$;

3) $\overline{x_k}$ співпадає з однією з точок $x_1, \ldots, x_k$;

4) $\overline{x_k}$ співпадає з однією з точок:

$$x_k^{'} = a_k + \frac{F_{n-k+2}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+2}}{F_{n+2}}(b - a), \qquad (2.9)$$

коли $a_k = a_{k-1}$(тобто відрізок $[a_k, b_k]$ знаходиться зліва на відрізку $[a_{k-1}, b_{k-1}]$, рис. 2.3, випадок 1),

$$x_k^{''} = a_k + \frac{F_{n-k+1}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+1}}{F_{n+2}}(b - a), \qquad (2.10)$$

коли $b_k = b_{k-1}$(тобто відрізок $[a_k, b_k]$ знаходиться справа на відрізку $[a_{k-1}, b_{k-1}]$, рис. 2.3, випадок 2). Точки $x_k^{'}$ або $x_k^{''}$ симетричні одна одній на відрізку $[a_k, b_k]$ відносно його середини, тобто $x_k^{''} = a_k + b_k - x_k^{'}$.

Коли $k < n$, то на відрізку $[a_k, b_k]$ вибираємо точку $x_{k+1} = a_k + b_k - \overline{x_k}$, яка симетрична $\overline{x_k}$ і співпадає з однією з точок $x_k^{'}$ або $x_k^{''}$, відмінної від $\overline{x_k}$. Процес продовжується доки не буде вичерпано кількість ітерацій.

Нехай $k = n$. Тоді процес закінчуємо і отримуємо

$$x_n^{'} = a_n + \frac{b - a}{F_{n+2}}, \; x_n^{''} = a_n + \frac{b - a}{F_{n+2}},$$

на відрізку $[a_n, b_n]$ точка $\overline{x_n}$ співпадає з $x_n^{'} = x_n^{''}$ і

$$\overline{x_n} = a_n + \frac{b - a}{F_{n+2}}, \; b_n - a_n = \frac{2(b - a)}{F_{n+2}}. \qquad (2.11)$$

За розв’язок приймається точка $\overline{x_n}$ з похибкою

$$\left| x_* - \overline{x_n} \right| \le \frac{b - a}{F_{n+2}}, \qquad (2.12)$$

Рекомендації до чисельної реалізації методу Фібоначчі на ЕОМ

Помітимо, що число $\frac{F_n}{F_{n+2}}$ взагалі є нескінченим періодичним десятковим дробом. Тому перша точка x_1 буде задаватися наближено. Похибка у визначенні першої точки методу Φ_n призводить до швидкого зростання похибки на наступних кроках і при вже не дуже великих n результати будуть відрізнятися від правильних. Тому, на практиці, на кожному відрізку $[a_k, b_k]$, який містить точку $\overline{x_k}$ з попереднього кроку, при виборі наступної точки x_{k+1} потрібно остерігатися користуватися формулою $x_{k+1} = a_k + b_k - \overline{x_k}$, тому що тоді похибка з попереднього кроку переноситься на даний крок. Замість цього краще безпосередньо обчислити точки $x_k^{'}$ та $x_k^{''}$ і за точку x_{k+1} взяти ту з точок, яка найбільш віддалена від $\overline{x_k}$.

Найчастіше відомі довжина відрізка, на якому розшукується точка мінімуму і точність обчислення $\varepsilon > 0$. Кількість ітерацій, необхідних для досягнення заданої точності, дорівнює найменшому з чисел n, яке задовольняє нерівності

$$\frac{b-a}{F_{n+2}} \le \varepsilon \le \frac{b-a}{F_{n+1}},$$

або
$$F_{n+1} \le \frac{b-a}{\varepsilon} \le F_{n+2}.$$

Відзначимо, що довжина відрізка $[a, b]$, на якому можна знайти точку x_* з заданою точністю $\varepsilon > 0$, виконавши при цьому n обчислень значень цільової функції $f(x)$, не перевищує $\varepsilon \cdot F_{n+2}$.

Алгоритм методу Фібоначчі

Початковий етап.

Крок 1. Задати точність $\varepsilon > 0$ обчислення точки мінімуму цільової функції $f(x)$ на відрізку $[a, b]$, покласти $F_1 = F_2 = 1$.

Крок 2. Покласти $j = 1$.

Крок 3. Обчислити $F_{j+2}=F_{j+1}+F_j$.

Крок 4. Якщо $F_{j+1}\le\frac{b-a}{\varepsilon}\le F_{j+2}$, то покласти $n=j$, перейти до кроку 5, інакше покласти $j=j+1$ і перейти до кроку 3.

Крок 5. Обчислити точки

$$\lambda_1=a+\frac{F_n}{F_{n+2}}(b-a),\ \mu_1=a+\frac{F_{n+1}}{F_{n+2}}(b-a).$$

Крок 6. Обчислити значення цільової функції $f(\lambda_1)$ та $f(\mu_1)$. Якщо $f(\lambda_1)\le f(\mu_1)$, то покласти $a_2=a$, $b_2=\mu_1$ і перейти до кроку 7, інакше $a_2=\lambda_1$, $b_2=b$ і перейти до кроку 7.

Крок 7. Покласти $k=2$.

Основний етап.

Крок 8. Якщо $f(\lambda_{k-1})\le f(\mu_{k-1})$, то обчислити точку

$$\lambda_k=a_k+\frac{F_{n-k+1}}{F_{n+2}}(b-a),$$

обчислити значення цільової функції $f(\lambda_k)$ і перейти до кроку 9, інакше покласти $\lambda_k=\mu_{k-1}$, $f(\lambda_k)=f(\mu_{k-1})$, $a_k=\lambda_{k-1}$ і перейти до кроку 10.

Крок 9. Покласти $\mu_k=\lambda_{k-1}$, $f(\mu_k)=f((\lambda_{k-1})$ і перейти до кроку 11.

Крок 10. Обчислити точку

$$\mu_k=a_k+\frac{F_{n-k+2}}{F_{n+2}}(b-a),$$

обчислити значення цільової функції $f(\mu_k)$ і перейти до кроку 11.

Крок 11. Якщо $f(\lambda_k)\le f(\mu_k)$, то покласти $a_{k+1}=a_k$, $b_{k+1}=\mu_k$, і перейти до кроку 12, інакше покласти $a_{k+1}=\lambda_k$, $b_{k+1}=b_k$, і перейти до кроку 12.

Крок 12. Якщо $k<n$, то покласти $k=k+1$ і перейти до кроку 8, інакше покласти $t_n=\frac{a_n+b_n}{2}$ і закінчити обчислення.

Алгоритм описаний.

Приклад. Знайти мінімум функції $f(x)$ методом Фібоначчі.

$$f(x) = x^2 + 2x \to min,\ x \in [3{,}5].$$

Відзначимо, що цільова функція $f(x)$ є строго квазіопуклою на заданому проміжку, а точка мінімуму дорівнює $x_* = -1$. Знайдемо наближене значення точки мінімуму функції $f(x)$ з точністю $\varepsilon = 0.2$. Визначимо число n обчислень значень цільової функції

$$F_{n+1} \le \frac{8}{0.2} \le F_{n+2}.$$

Так як $F_9 = 34$, $F_{10} = 55$, маємо, що $n = 8$. Результати обчислень наведені в таблиці 2.3. За точку мінімуму може бути взята середина останнього відрізка $\overline{x_n} = \dfrac{a_8 + b_8}{2} = -0.964$.

Таблиця 2.3

k	a_k	b_k	λ_k	μ_k	$f(\lambda_k)$	$f(\mu_k)$
1	–3.0000	5.0000	0.0545	1.9455	$\underline{0.112}$	$\underline{7.676}$
2	–3.0000	1.9455	–1.1091	0.0545	$\underline{–0.988}$	0.112
3	–3.0000	0.0545	–1.8364	–1.1091	$\underline{–0.300}$	–0.988
4	–1.8364	0.0545	–1.1091	–0.6727	–0.988	$\underline{–0.893}$
5	–1.8364	–0.6727	–1.3999	–1.1091	$\underline{–0.840}$	–0.988
6	–1.3999	–0.6727	–1.1091	–0.9366	–0.988	$\underline{–0.999}$
7	–1.1091	–0.6727	–0.9636	–0.8182	$\underline{–0.999}$	–0.967
8	–1.1091	–0.8182	–0.9636	–0.9636	–0.999	–0.999

§4. Порівняння методів лінійного пошуку

Нехай довжина вихідного відрізка дорівнює $(b_1 - a_1)$. При заданій величині останнього відрізка $(b_n - a_n)$, яка задовольняє потрібній точності $\varepsilon > 0$, необхідне число обчислень цільової функції n може бути визначено як найменше додатне ціле, яке задовольняє таким співвідношенням:

1. метод ділення відрізку навпіл $\left(\dfrac{1}{2}\right)^{\frac{n}{2}} \le \dfrac{b_n - a_n}{b_1 - a_1}$,

2. метод золотого перерізу $\left(\dfrac{\sqrt{5}-1}{2}\right)^{n-1} \le \dfrac{b_n - a_n}{b_1 - a_1}$,

3. метод Фібоначчі $F_n \ge \dfrac{b_1 - a_1}{b_n - a_n}$.

Для фіксованого значення відношення $\dfrac{b_1 - a_1}{b_n - a_n}$ найменше число обчислень цільової функції відповідає найбільш ефективному алгоритму. Тому, з цієї точки зору, найбільш ефективним алгоритмом є метод Фібоначчі, далі – метод золотого перерізу, потім – метод ділення відрізка навпіл. Зауважимо, що для достатньо великих n значення $\dfrac{1}{F_n} \to \left(\dfrac{\sqrt{5}-1}{2}\right)^{n-1}$, так що методи Фібоначчі та золотого перерізу є майже рівноцінними.

§5. Одновимірний «економічний» пошук

Одновимірний пошук є основою багатьох алгоритмів для розв'язання задач нелінійного програмування. Як правило, алгоритми нелінійного програмування являють собою таку процедуру. Задається точка $x^{(k)}$, визначається вектор напрямку $p^{(k)}$ і підходяща довжина кроку α_k, після цього обчислюється нова точка $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$. Потім цей процес повторюється. Визначення кроку α_k досягається розв'язанням задачі мінімізації функції $f\left(x^{(k)} + \alpha\, p^{(k)}\right)$, яка залежить від змінної α. Це задача одновимірного пошуку. Більшість методів одновимірного пошуку переслідує мету – отримати більш або менш точне наближення до точки глобального мінімуму (або до точки локального мінімуму) функції $g(\alpha) = f\left(x^{(k)} + \alpha\, p^{(k)}\right)$ для $\alpha \ge 0$. Відзначимо, що для задачі одновимірного пошуку $x^{(k)} = \left(x_1^{(k)},\ldots,x_n^{(k)}\right)$, $p^{(k)} = \left(p_1^{(k)},\ldots,p_n^{(k)}\right)$ – відомі вектори простору E^n.

У той же час, збіжність деяких алгоритмів нелінійного програмування може бути досягнута без того, щоб вимагати знаходити близьку точку до точки глобального (або локального) мінімуму для функції $f(x)$ у напрямку $p^{(k)}$.

Пропонуються два методи, які засновані на менш обмежувальних умовах. Ці методи дозволяють отримати точку з потрібними характеристиками за невелике число оцінок цільової функції. Такими методами є:

1) правило Голдстейна, яке застосовується в тих випадках, коли не можна оцінити градієнт цільової функції $g(\alpha)$ або коли отримання оцінки є дуже складним.

2) правило Вольфа-Пауелла, яке потребує оцінки градієнта кожний раз, коли функція обчислена.

Загальний принцип, який лежить в основі цих методів, полягає в наступному:

1) α не повинно вибиратися занадто великим (у протилежному випадку алгоритм може почати осцилювати);

2) α не повинно вибиратися занадто малим (у протилежному випадку алгоритм може збігатися дуже повільно).

У правилі Голдстейна (G) перша умова виконується у силу співвідношення

$$g(\alpha) \le g(0) + m_1 \alpha\, g'(0),\ \text{де}\ m_1 \in (0,1), \tag{2.13}$$

а виконання другої умови забезпечується співвідношенням

$$g(\alpha) \ge g(0) + m_2 \alpha\, g'(0),\ \text{де}\ m_2 \in (m_1, 1). \tag{2.14}$$

Рис. 2.4 дає приклад множини точок, які задовольняють співвідношенням (2.13), (2.14).

Зауважимо, що умова (2.13) забезпечує той факт, що нова отримана точка $\overline{x^{(k)}} = x^{(k)} + \alpha\, p^{(k)}$ задовольняє нерівності $f\left(\overline{x^{(k)}}\right) < f\left(x^{(k)}\right)$ (умова монотонності).

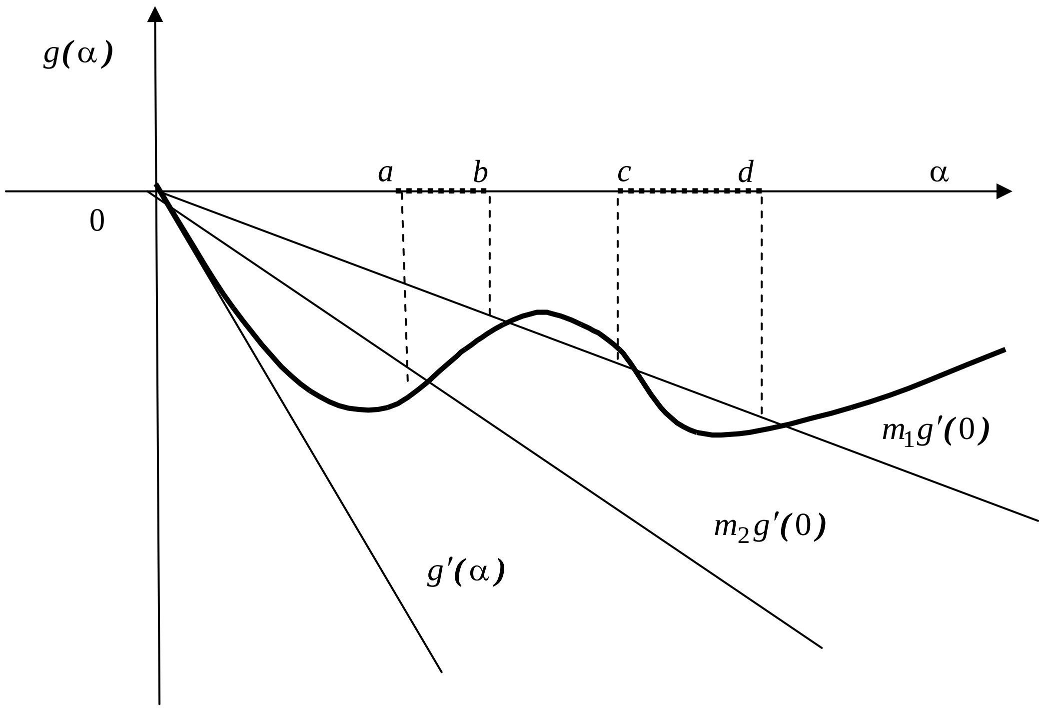


Рис. 2.4

Існує варіант правила Голдстейна, який запропоновано Армійо. У ньому α вибирається таким, що:

1) α задовольняє співвідношенню (2.13);

2) $M\alpha$ не задовольняє співвідношенню (2.13).

При цьому $M > 1$ (звичайно вибирається в границях від 5 до 10).

У правилі Вольфа-Пауелла (WP) знову саме співвідношення (2.13) приводить до того, що α вибирається не занадто великим. Але оскільки передбачається, що обчислення градієнта функції $f(x)$ потребує не набагато більших обчислень, ніж оцінка самої функції, то співвідношення (2.14) може бути замінено умовою

$$g'(\alpha) \geq m_3 g'(0), \tag{2.15}$$

де m_3 – постійний коефіцієнт, який вибирається з інтервалу $(m_1, 1)$. Це має місце, тому що $g'(\alpha)$ в кожній точці α може обчислюватися за формулою

$$g'(\alpha) = p^{(k)} f'\left(x^{(k)} + \alpha\, p^{(k)}\right).$$

На рис. 2.5 показана множина точок, яка задовольняє співвідношенням (2.13), (2.15) правила Вольфа-Пауелла (відрізки $[a, b]$ і $[c, d]$).

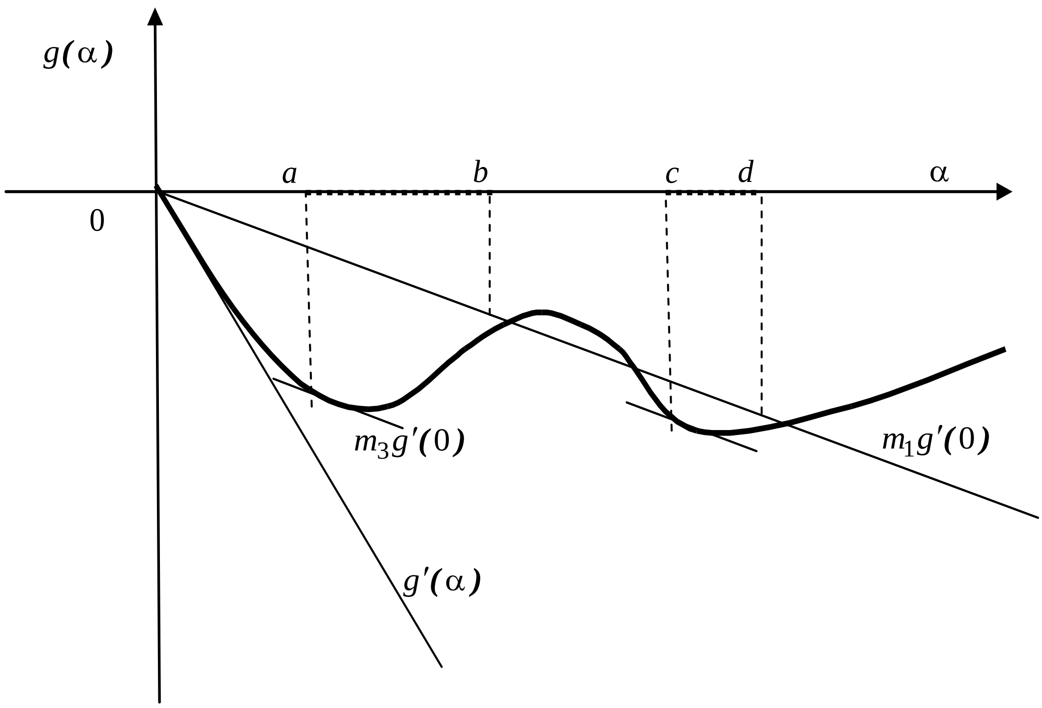


Рис. 2.5

За коефіцієнти m_1 та m_3 звичайно вибираються $m_1 = 0.1$, $m_3 = 0.7$.

Алгоритм методів

Крок 1. Визначити $\alpha_{\min} = 0$, $\alpha_{\max} = +\infty$. Знайти $g'(0) = p^{(k)} f'\left(x^{(k)}\right)$ та присвоїти α початкове значення.

Крок 2. Обчислити $g(\alpha) = f\left(x^{(k)} + \alpha\, p^{(k)}\right)$. Якщо $g(\alpha) \le g(0) + m_1\,\alpha\, g'(0)$, то перейти до кроку 3, інакше покласти $\alpha_{\max} = \alpha$ і перейти до кроку 5.

Крок 3. Обчислити $g'(\alpha) = p^{(k)} f'\left(x^{(k)} + \alpha\, p^{(k)}\right)$.

Якщо $g'(\alpha) \ge m_3\, g'(0)$, то кінець, інакше перейти до кроку 4 (правило WP).

Якщо $g(\alpha) \ge g(0) + m_2\,\alpha\, g'(0)$, то кінець, інакше перейти до кроку 4 (правило G).

Крок 4. Покласти $\alpha_{\min} = \alpha$.

Крок 5. Знайти нове значення α з інтервалу $(\alpha_{min}, \alpha_{max})$ та перейти до кроку 2.

Алгоритм описаний.

Зауважимо, що для вибору початкового значення α необхідно кожний раз, коли це можливо, використовувати інформацію, яку дає алгоритм методу нелінійного програмування.

Проілюструємо це зауваження на прикладі квазіньютонівських методів. Якщо в квазіньютонівському методі перед використанням правила Гольдстейна або правила Вольфа-Пауелла провести оцінку кроку α, а потім використати це значення як початкове, то для отримання методом G або WP задовільної точки $x^{(k)} + \alpha\, p^{(k)}$ може бути достатньо однієї оцінки функції. Ось тому ці методи виправдовують свою назву «економічних» методів.

Приклад. Нехай $f(x) = x_1 - 2x_2 + x_2^2$. Задано $x^{(k)} = (0, 2)$, $p^{(k)} = (2, -2)$. Знайти таке значення α, при якому функція $g(\alpha) = f\left(x^{(k)} + \alpha\, p^{(k)}\right)$ приймає мінімальне значення.

З урахуванням того, що

$$x^{(k)} + \alpha\, p^{(k)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ -2 \end{pmatrix} = \begin{pmatrix} 2\alpha \\ 2(1-\alpha) \end{pmatrix},$$

функцію $g(\alpha)$ можна записати у вигляді $g(\alpha) = -6\alpha + 4\alpha^2$. Далі

1) знайдемо значення α за правилом Гольдстейна. Виберемо $m_1 = 0.1$, $m_3 = 0.6$. Нехай $\alpha_{min} = 0$, $\alpha_{max} = +\infty$. Знайдемо значення $g(0) = 0$, $g'(\alpha) = -6 + 8\alpha$, $g'(0) = -6$. Присвоїмо α початкове значення. Наприклад, $\alpha = 1$, $g(1) = -2$. Перевіримо виконання умов (2.13), (2.14). Умови виконуються. Згідно з алгоритмом $\alpha = 1$ – наближене значення точки мінімуму $\alpha^* = 0.75$. На рис. 2.6 показано, що для даної функції $g(\alpha)$ відрізок $[0.6, 1.35]$ буде задовольняти правилу Гольдстейна.

2) знайдемо значення α за правилом Вольфа-Пауелла. Виберемо $m_1 = 0.1$, $m_3 = 0.7$. Нехай $\alpha_{min} = 0$, $\alpha_{max} = +\infty$, $g(0) = 0$, $g'(\alpha) = -6 + 8\alpha$, $g'(0) = -6$. Присвоїмо початкове значення $\alpha = 3$, $g(3) = 18$. Перевіримо виконання умов (2.13), (2.14). Умова (2.13) не виконується, тому $\alpha_{max} = 3$ і необхідно задати нове значення для α вже в новому інтервалі $(\alpha_{min}, \alpha_{max})$. Нехай $\alpha = 2$, $g(2) = 4$. Умова (2.13) не виконується і $\alpha_{max} = 2$. Нехай $\alpha = 1$, $g(1) = -2$. Умова (2.13) виконується, $g'(1) = 2$. Умова (2.15) також виконується. Тому $\alpha = 1$ –

наближене значення точки мінімуму $\alpha^* = 0.75$. На рис. 2.7 показано відрізок $[0.45, 1.35]$, який задовольняє правилу Вольфа-Пауелла.

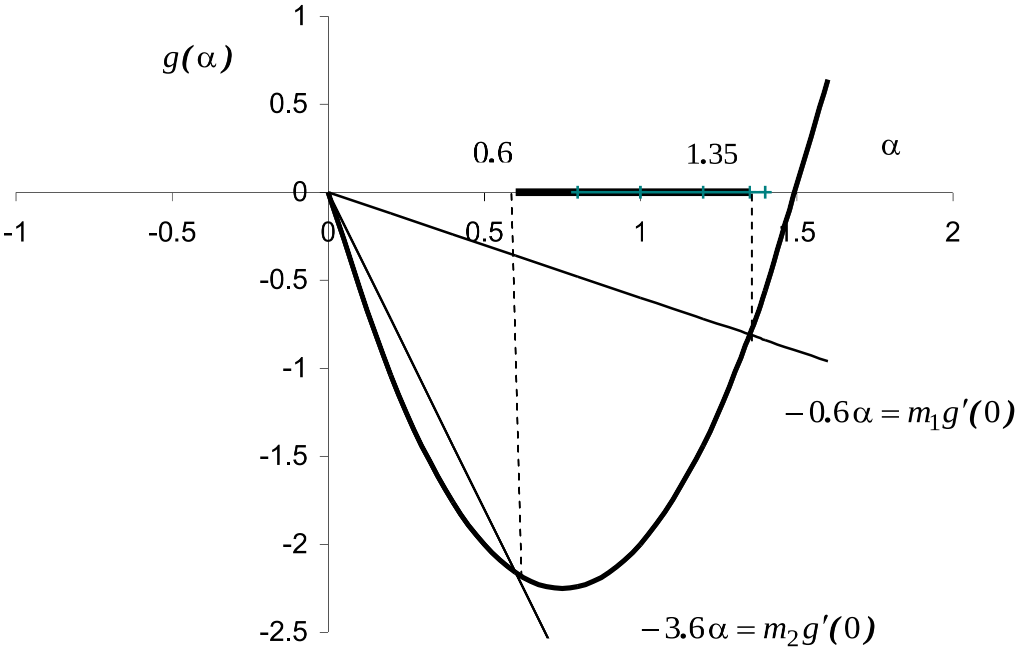


Рис. 2.6

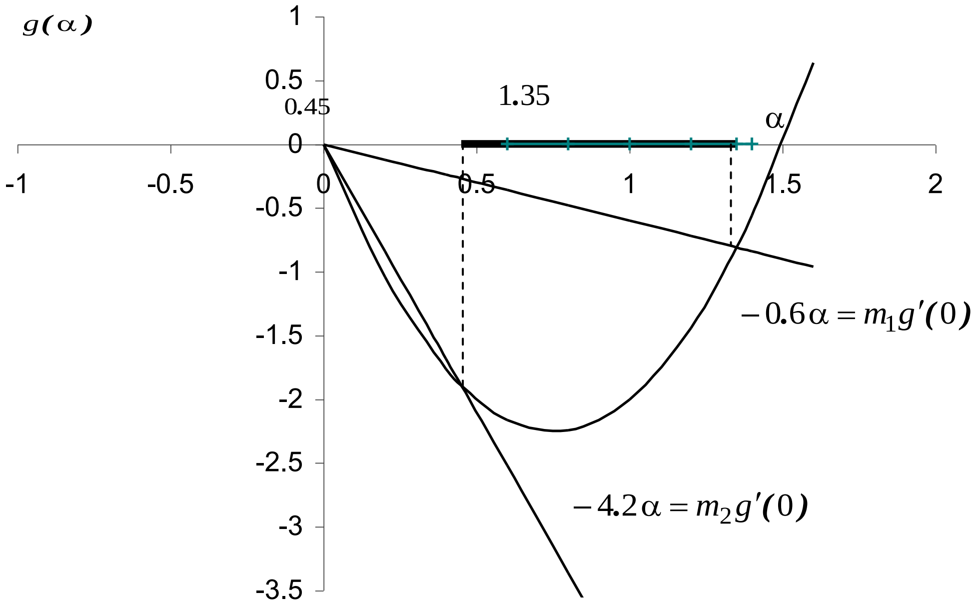


Рис. 2.7

§6. Метод дотичних

Постановка задачі. Мінімізувати функцію $f(x)$ на відрізку $[a,b]$. Функція $f(x)$ є опуклою і диференційовною на $[a,b]$.

Зафіксуємо довільну точку $y \in [a,b]$ і визначимо функцію

$$g(x,y) = f(y) + f'(y)(x-y),\ a \le x \le b,$$

де $g(x,y)$ – лінійна за x функція, яка задає дотичну до графіку функції $f(x)$ у точці y. Відзначимо, що $g(x,y) \le f(x)$, для будь–якого $x \in [a,b]$.

За початкове наближення візьмемо точку $x^{(0)} \in [a,b]$. Наприклад, $x^{(0)} = a$. Складемо функцію $p_0(x) = g\left(x, x^{(0)}\right)$ і визначимо точку $x^{(1)} \in [a,b]$ з умови $p_0\left(x^{(1`)}\right) = \min\limits_{x\in[a,b]} p_0(x)$. Зрозуміло, що при $f'\left(x^{(0)}\right) \ne 0$ буде $x^{(1)} = a$ або $x^{(1)} = b$.

Далі, виберемо нову функцію $p_1(x) = max\left\{p_0(x),\ g\left(x, x^{(1)}\right)\right\}$ і наступну точку $x^{(2)} \in [a,b]$ знайдемо з умови $p_1\left(x^{(2`)}\right) = \min\limits_{x\in[a,b]} p_1(x)$. І так далі...

Якщо точки $x^{(0`)}, x^{(1`)}, \ldots, x^{(n`)},\ n \ge 1$ вже відомі, то складаємо функцію $p_n(x) = max\left\{p_{n-1}(x), g\left(x, x^{(n)}\right)\right\} = \max\limits_{0\le i\le n} g\left(x, x^{(i)}\right)$ і наступну точку $x^{(n+1)} \in [a,b]$ визначимо з умови $p_n\left(x^{(n+1`)}\right) = \min\limits_{x\in[a,b]} p_n(x)$.

Якщо при деякому $n \ge 0$ виявиться, що $f'\left(x^{(n)} + 0\right) \ge 0$, $f'\left(x^{(n)} - 0\right) \le 0$ (при $a < x < b$ ця умова рівносильна умові $f'\left(x^{(n)}\right) = 0$), то точка $x^{(n)}$ – точка мінімуму функції $f(x)$ на $[a,b]$ і ітерації на цьому закінчуються.

З рис. 2.8 видно, що $p_n(x)$ – неперервна, Кусково-лінійна функція і її графік є ламаною, яка складається з відрізків дотичних до графіку функції $f(x)$ в точках $x^{(0)},\ x^{(1)}, \ldots,\ x^{(n)}$. Тому сформульований метод і називається методом дотичних.

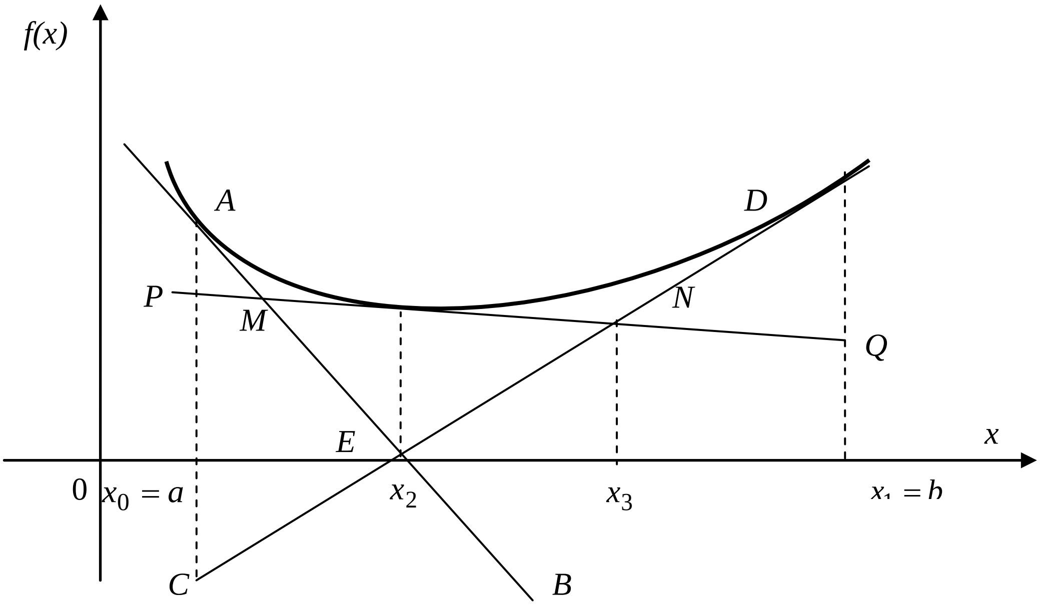


Рис. 2.8

На рис. 2.8 використані такі позначення:

AB – графік функції $g\left(x,x^{(0)}\right)$;

CD – графік функції $g\left(x,x^{(1)}\right)$;

AED– графік функції $p_1(x)$;

PQ– графік функції $g\left(x,x^{(2)}\right)$;

AMND– графік функції $p_2(x)$;

FL– графік функції $g\left(x,x^{(3)}\right)$;

AMRSD– графік функції $p_3(x)$.

Обґрунтування збіжності методу будемо проводити за припущенням, що $f'\left(x^{(k)}\right)\neq 0$, $k=0,1,\ldots$, оскільки у протилежному випадку точка $x^{(k)}$ є точкою мінімуму і процес на цьому кроці закінчується.

Теорема. Для опуклої та неперервно-диференційовної на проміжку $[a,b]$ функції $f(x)$ метод дотичних збігається, тобто

$$\lim_{k\to\infty} f\left(x^{(k)}\right)=\lim_{k\to\infty} p_k\left(x^{(k+1)}\right)=f\left(x^*\right)=\min_{x\in[a,b]} f(x),$$

і справедлива оцінка:

$$0\le f\left(x^{(k+1)}\right)-f\left(x^*\right)\le f\left(x^{(k+1)}\right)-p_k\left(x^{(k+1)}\right),\ k=1,2,\ldots$$

Якщо точка мінімуму x_* є єдиною, то $\lim\limits_{k\to\infty} x^{(k)}=x_*$.

Доведення. З визначення функції $p_k(x)$ та її властивостей маємо

$$p_{k-1}\left(x^{(k`)}\right)=\min_{x\in[a,b]} p_{k-1}(x)\le p_{k-1}\left(x^{(k+1)}\right)\le p_k\left(x^{(k+1)}\right)=$$

$$=\min_{x\in[a,b]} p_k(x)\le p_{k-1}(x_*)\le f(x_*).$$

Таким чином, послідовність $\left\{p_k\left(x^{(k+1)}\right)\right\}$ – монотонно зростаюча та обмежена зверху, внаслідок чого існує $p_*=\lim_{k\to\infty} p_k\left(x^{(k+1)}\right)$.

З нерівності $p_{k-1}(x)\le f(x)$ витікає, що

$$p_*\le f(x_*). \qquad (2.16)$$

Доведемо, що $p_*=f(x_*)$. Нехай $L=\max_{x\in[a,b]}\left|f'(x)\right|$. Зазначимо, що так як функція $f(x)$ опукла, то $L=\max_{x\in[a,b]}\left\{\left|f'(a)\right|,\left|f'(b)\right|\right\}$ і ламана $p_k(x)$ буде задовольняти умові Ліпшиця на відрізку $[a,b]$ з константою Ліпшиця L. Тому що $x^{(k)}\in[a,b]$, $k=0,1,\ldots$, то існує підпослідовність $\left\{x^{(k_i)}\right\}$, яка є збіжною: $\bar{x}=\lim_{i\to\infty} x^{(k_i)}$. З того, що $p_{k_i-1}\left(x^{(k_i)}\right)=\min_{x\in[a,b]} p_{k_i-1}(x)\le p_{k_i-1}\left(x^{(k_i-1)}\right)$ через умову Ліпшиця для функції $p_{k_i-1}(x)$, маємо

$$0\le p_{k_i-1}\left(x^{(k_i-1)}\right)-p_{k_i-1}\left(x^{(k_i)}\right)\le L\left\|x^{(k_i-1)}-x^{(k_i)}\right\|.$$

Можна вважати, що індекси $k_1<k_2<\ldots<k_i$, тоді $k_{i-1}\le k_i-1$,

$$p_{k_i-1}\left(x^{(k_i-1)}\right)=f\left(x^{(k_i-1)}\right),\ p_{k-1}\left(x^{(i)}\right)=f\left(x^{(i)}\right),\ \forall i=\overline{0,\,k-1}.$$

Тому

$$0\le f\left(x^{(k_i-1)}\right)-p_{k_i-1}\left(x^{(k_i`)}\right)\le L\left\|x^{(k_i-1)}-x^{(k_i)}\right\|.$$

Переходячи до границі при $i\to\infty$, отримаємо нерівність $0\le f(\bar{x})-p_*\le 0$, звідки витікає, що $f(\bar{x})=p_*$. Але $f(x_*)\le f(\bar{x})$, тому $f(x_*)=p_*$. Отже, через нерівність (2.16), буде $f_*=p_*$.

Друге твердження теореми є прямим наслідком того, що будь-яка гранична точка $\bar{x}$ послідовності $\{x^{(k)}\}$ є такою, що $f(\bar{x})=f(x_*)=\min_{x\in[a,b]} f(x)$.

Зауваження. Метод дотичних застосовується і для одновимірної мінімізації опуклих функцій, від яких не вимагається неперервна диференційованість. Оскільки опукла функція $f(x)$ має в кожній внутрішній точці y відрізку $[a,b]$ лівосторонню $f'(y-0)$ та правосторонню $f'(y+0)$ похідні, то метод дотичних може бути таким чином пристосований до цього випадку.

За лінійну функцію будемо обирати таку функцію:

$$g(x,x_i)= f(x_i)+\gamma_i(x-x_i),$$

де γ_i – будь-яке число з відрізку $\left[f'(x_i-0), f'(x_i+0)\right]$. Наприклад, γ_i можна вибирати з умови $|\gamma_i| = min\left\{ \left|f'(x_i-0)\right|, \left|f'(x_i+0)\right| \right\}$.

Подальша конструкція методу залишається такою, як описано раніше. Теорема про збіжність залишається справедливою і в цьому випадку.

Алгоритм, який наведено нижче, дозволяє знайти точку мінімуму опуклої диференційованої функції $f(x)$ відрізку $[a,b]$ із заздалегідь заданою точністю $\varepsilon > 0$.

Алгоритм методу

Крок 1. Покладемо $a_1 = a$, $b_1 = b$, $f'(a_1)= f'(a)$, $f'(b_1)= f'(b)$.

k-й крок ($k \ge 1$). Якщо $\frac{1}{2}(b_k - a_k)\le \varepsilon$, то точність досягнута і процес закінчується, точку мінімуму знайдено.

Якщо $\frac{1}{2}(b_k - a_k) > \varepsilon$, то визначаємо абсцису $x^{(k)}$ точки перетину прямих

$$y = f(a_k)+ f'(a_k)(x-a_k),\ y = f(b_k)+ f'(b_k)(x-b_k).$$

Потім обчислюємо $f'\left(x^{(k)}\right)$.

Якщо $f'\left(x^{(k)}\right)=0$, то мінімум знайдемо і процес закінчується.

Якщо $f'\left(x^{(k)}\right)<0$, то $a_{k+1} = x^{(k)}$, $b_{k+1} = b_k$.

Якщо $f'\left(x^{(k)}\right)>0$, то $a_{k+1} = a_k$, $b_{k+1} = x^{(k)}$.

Алгоритм описаний.

Неважко виписати явно вираз для $x^{(k+1)}$, який визначається умовою $g(x,a_k)=g(x,b_k)$ перетину дотичних в точках a_k, b_k при $f'(a_k)<0$, $f'(b_k)>0$:

$$x^{(k+1)}=\frac{f(a_k)-f(b_k)+b_k f'(b_k)-a_k f'(a_k)}{f'(b_k)-f'(a_k)},\ k\ge 1.$$

Теорема. Нехай

1) функція $f(x)$ є двічі неперервно диференційовною функцією на відрізку $[a,b]$,
2) $\inf\limits_{x\in[a,b]} f''(x)>0$,
3) x_* – точка мінімуму функції $f(x)$ на відрізку $[a,b]$,
4) послідовність $\{x^{(k)}\}$ отримана методом дотичних при $x^{(0)}=a$, причому $x^{(k)}\ne x_*$, $k=0,1,\dots$

Тоді для будь-якого $\varepsilon>0$ існує номер $N=N(\varepsilon)$ такий, що

$$\left|x^{(k)}-x_*\right|\le\left(\frac{1+\varepsilon}{2}\right)^{k-N}(b_N-a_N),\ k\ge N.$$

Отримана оцінка показує, що метод дотичних збігається зі швидкістю, яка є не меншою ніж швидкість збіжності геометричної прогресії з знаменником

$$q=\frac{1+\varepsilon}{2}\approx\frac{1}{2}.$$

Переваги методу: за допомогою методу можна отримати розв'язок задач мінімізації для функцій, які задовольняють умові Ліпшиця. Проста та зручна у практичному застосуванні формула, яка дає оцінку невідомої похибки $f\left(x^{(k+1)}\right)-f(x_*)$ через відомі величини, що обчислюються у процесі реалізації методу. Метод збігається при будь-якому виборі початкової точки $x^{(0)}$.

Недоліки методу: він може застосовуватися лише у випадку, коли функція, що мінімізується, є опуклою і значення функції та її похідних обчислюються достатньо просто. Метод дотичних для класу гладких опуклих функцій не кращий за метод ділення відрізка навпіл.

Приклад. Найти мінімум функції $f(x)=x^2-1$ на відрізку $[-1,2]$ з точністю $\varepsilon=0.15$. Результати обчислень наведені в таблиці 2.4, а на рис.2.9 дана геометрична інтерпретація методу.

Таблиця 2.4

k	a_k	b_k	$f(a_k)$	$f(b_k)$	$f'(a_k)$	$f'(b_k)$	$x^{(k)}$	$f'(x^{(k)})$
1	–1.00	2.00	0.00	3.00	–2.00	4.00	0.50	1.00
2	–1.00	0.50	0.00	–0.75	–2.00	1.00	–0.25	–0.50
3	–0.25	0.50	–0.9373	–0.75	–0.50	1.00	0.125	0.25
4	–0.25	0.125	–0.9373	–0.9844	–0.50	0.25	–0.0625	–0.125
5	–0.0625	0.125						

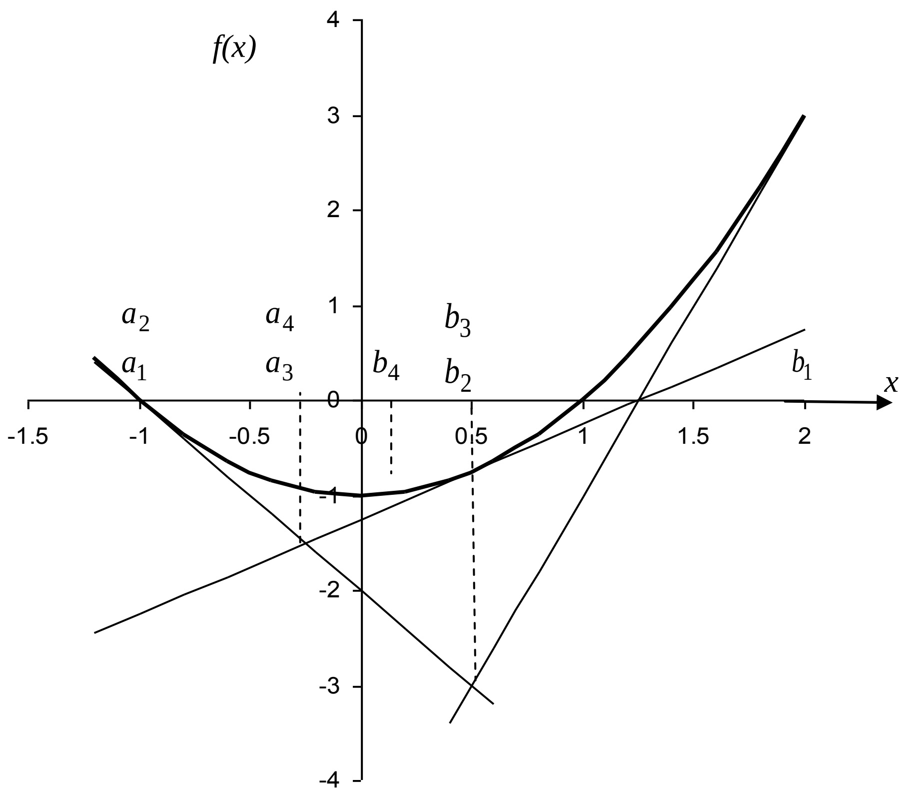


Рис. 2.9

Вправи до розділу 2

1. Для кожного з трьох методів, знаючи два параметра $(b-a)$ – довжину відрізка, ε – задану точність обчислення точки мінімуму x_* функції $f(x)$ відрізка $[a,b]$, N – необхідну кількість обчислень значень функції $f(x)$, знайти третій параметр.

2. Знайти найменше n, починаючи з якого точність методу золотого перерізу більша за точність методу Фібоначчі в 2, 1.5 рази.
3. Знайти найменше N, щоб $\left| F_n - \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n \right| < \varepsilon$ для усіх $n \geq N$.

   Розглянути випадки $\varepsilon = 10^{-3}$, $\varepsilon = 10^{-4}$.
4. Знайти всі точки локального екстремуму функції

$$f(x) = \left| \left| \left| x^2 - 1 \right| - 1 \right| - 1 \right|$$

на відрізку $[a,b]$ при різних a та b. При яких a та b ця функція буде унімодальною на відрізку $[a,b]$?

5. Довести, що лінійна функція $f(x) = cx + d$, де c, d – сталі, $c \neq 0$ досягає свого мінімуму та максимуму на відрізку $[a,b]$ тільки при $x = a$ або $x = b$.
6. Знайти мінімум функції $f(x)$ відрізку $[a,b]$ з точністю $\varepsilon = 10^{-1}$, 10^{-2}, 10^{-4} (табл. 2.5).

Таблиця 2.5

N	$f(x)$	$[a,b]$	N	$f(x)$	$[a,b]$
1	$y = \frac{x}{x^2+1} \to min$	$\left[-\frac{3}{2}, 0\right]$	2	$y = \frac{ln\, x - 1}{x} \to min$	$\left[e, e^3\right]$
3	$y = x + \frac{2}{x} \to min$	$\left[\frac{1}{2}, 3\right]$	4	$y = x + exp(-x) \to min$	$[-2, 2]$
5	$y = x + exp\left(\frac{-x}{2}\right) \to max$	$[1, 3]$	6	$y = x + \frac{4}{x^2} \to min$	$[1, 4]$
7	$y = x^4 - 2x^2 + 3 \to max$	$[-1, 1]$	8	$y = ln\left(x^2+1\right) \to min$	$[-1, 1]$
9	$y = \frac{x}{(x+1)^2} \to min$	$[0, 2]$	10	$y = \frac{x^2+1}{x} \to max$	$\left[-2, -\frac{1}{2}\right]$
11	$y = \frac{x^4}{(x+1)^3} \to min$	$\left[-\frac{1}{2}, 1\right]$	12	$y = \frac{4}{3}x^3 - 4x \to min$	$[0, 2]$
13	$y = \lvert x \rvert + (x-1)^2 \to min$	$[-1, 1]$	14	$y = (x-3)\sqrt{x} \to min$	$[0, 2]$

15	$y=\frac{\ln x}{\sqrt{x}}\to max$	$[e,e^3]$	16	$y=\frac{\exp x}{x+1}\to min$	$\left[-\frac{1}{2},1\right]$
17	$y=\lvert x\rvert+2x^2\to min$	$\left[-\frac{1}{2},2\right]$	18	$y=x^2\exp(-x)\to max$	$[1,3]$
19	$y=x^3-\frac{3}{4}x\to min$	$[0,2]$	20	$y=\exp\left(2x-x^2\right)\to max$	$[0,2]$
21	$y=\sqrt{x}+\sqrt{4-x}\to max$	$[0,4]$	22	$y=x+\sqrt{3-x}\to max$	$[0,3]$
N	$f(x)$	$[a,b]$	N	$f(x)$	$[a,b]$
23	$y=x^4-4x^3+6x^2-$ $-4x\to min$	$[-2,2]$	24	$y=\frac{(x-1)^2}{x}\to min$	$\left[\frac{1}{2},2\right]$
25	$y=3x-2x^2\to max$	$[-1,1]$	26	$y=\frac{x^2}{x+1}\to min$	$\left[-\frac{1}{2},1\right]$
27	$y=\frac{x^2-1}{x^2-4}\to max$	$[-1,1]$	28	$y=2x^3-6x^2-$ $-18x+7\to max$	$[-2,0]$
29	$y=\frac{(x-1)^2}{x}\to max$	$\left[-3,-\frac{1}{2}\right]$	30	$y=x\exp(x)\to min$	$[-3,3]$
31	$y=\frac{1}{4}x^4-\frac{5}{2}x^2+$ $+x+1\to max$	$\left[\frac{1}{2},1\right]$	32	$y=(x-1)\exp(x)\to min$	$[-1,1]$
33	$y=x+\frac{1}{x}\to min$	$\left[\frac{1}{2},2\right]$	34	$y=\lvert x\rvert+(x-2)^2\to min$	$[-1,2]$
35	$y=7-x-x^2\to max$	$[-1,3]$	36	$y=(x+2)\exp\left(\frac{1}{x}\right)\to min$	$[1,3]$
37	$y=\frac{x^2-3x-1}{x+2}\to max$	$[-6,-4]$	38	$y=\lvert x^2-3x+2\rvert-$ $-\lvert 2x-3\rvert\to min$	$[0,1]$
39	$y=2x^2+x\to min$	$[-1,1]$	40	$y=2x^3-3x^2\to max$	$\left[-1,\frac{1}{2}\right]$

41	$y=x^2+exp\,(x)\rightarrow min$	$[-1,0]$	42	$y=\dfrac{x^3+4}{x^2}\rightarrow min$	$[1,3]$
43	$y=(x-5)\,exp(x)\rightarrow min$	$[3,5]$	44	$y=x\,exp\,(-x)\rightarrow max$	$[0,2]$
45	$y=2x^2-5x\rightarrow min$	$[-1,2]$	46	$y=\dfrac{2x^3-32}{x^2}\rightarrow min$	$[3,5]$
47	$y=ln\,(x)-2x^2\rightarrow max$	$\left[\frac{1}{4},2\right]$	48	$y=cos\,(2x)+\dfrac{x}{2}\rightarrow max$	$[0,1]$
49	$y=x^3+5x^2+3x\rightarrow min$	$[-1,1]$	50	$y=\dfrac{x+\sqrt{2}}{x^2-1}\rightarrow max$	$\left[-\frac{3}{4},\frac{1}{2}\right]$
N	$f(x)$	$[a,b]$	N	$f(x)$	$[a,b]$
51	$y=\lvert x^2-3x+2\rvert-$ $-\lvert 2x-3\rvert\rightarrow max$	$[1,2]$	52	$y=\dfrac{1}{(x+2)(x-2)}\rightarrow max$	$\left[-\frac{3}{4},\frac{3}{2}\right]$
53	$y=x^2-\dfrac{1}{4}x^4\rightarrow max$	$\left[-\frac{3}{4},\frac{3}{2}\right]$	54	$y=x+\dfrac{2}{x}-3\rightarrow max$	$\left[-5,-\frac{1}{2}\right]$
55	$y=\lvert x^2-3x+2\rvert\rightarrow max$	$[1,2]$	56	$y=cos(\,x\,)+\dfrac{x}{2}\rightarrow max$	$[-1,1]$
57	$y=3x^4-4x^3+$ $+12x^2+1\rightarrow min$	$[-1,1]$	58	$y=x^3-3x+2\rightarrow min$	$[0,2]$
59	$y=x-ln(1+x)\rightarrow min$	$\left[-\frac{1}{2},1\right]$	60	$y=\dfrac{2x+1}{(x+3)(x-1)}\rightarrow max$	$[2,5]$
61	$y=x\,sin\,(x)+cos\,(x)-$ $-\dfrac{1}{2}x^2\rightarrow min$	$\left[-\frac{\pi}{2},\frac{\pi}{4}\right]$	62	$y=x^4-2x^2+5\rightarrow min$	$\left[-2,-\frac{1}{2}\right]$
63	$y=x-2\sqrt{x}\rightarrow min$	$\left[\frac{1}{2},2\right]$	64	$y=sin\,(2x)-x\rightarrow max$	$\left[0,\frac{\pi}{2}\right]$
65	$y=x^3-3x+7\rightarrow max$	$[-3,0]$	66	$y=\dfrac{exp\,(x)}{x}\rightarrow min$	$\left[-\frac{1}{2},\frac{3}{2}\right]$

67	$y = x^3\left(x^2 - 5\right) \to min$	$[0,3]$	68	$y = cos\,(3x) + \frac{3x}{2} \to max$	$\left[-\frac{\pi}{6}, \frac{\pi}{6}\right]$
69	$y = 2\sqrt{x} + \frac{1}{x} \to min$	$\left[\frac{1}{2}, 2\right]$	70	$y = \frac{x^2}{2} + \frac{2}{3x\sqrt{x}} \to min$	$\left[\frac{1}{2}, 2\right]$
71	$y = x^3 + 5x^2 + 3x \to max$	$[-4,0]$	72	$y = x^3\left(x^2 - 5\right) \to max$	$[-3,0]$
73	$y = \frac{x\sqrt{x}}{3} - \frac{x^2}{16} \to max$	$[12,18]$	74	$y = x^2\,exp\,(-x) \to min$	$[-1,1]$
75	$y = \frac{1}{4}x^4 - \frac{1}{3}x^3 -$ $-6x^2 - 2 \to \max$	$[-1,1]$			

Розділ 3. Чисельні методи безумовної оптимізації

Розглядаються чисельні методи для задачі безумовної оптимізації:

$$f(x)\to min,\ x\in E^n.$$

До теперішнього часу розроблена та досліджена велика кількість методів безумовної мінімізації функцій багатьох змінних. На дану тему існує чимала бібліографія. Ця область екстремальних задач продовжує бурхливо розвиватися, кількість робіт, які присвячені задачам мінімізації функцій скінченого числа змінних швидко збільшується.

У цьому розділі передбачається вивчення та отримання навичок практичної реалізації лише деяких методів безумовної мінімізації, які стали вже класичними та є важливими в методичному плані.

Серед **методів нульового порядку** будуть розглянуті: метод покоординатного спуску, пошуку по деформованому багатограннику, конфігурацій, метод Розенброка.

Серед **методів першого порядку** будуть розглянуті: градієнтні, спряжених напрямків, змінних напрямків, змінної метрики.

Серед **методів другого порядку** буде розглянутий: метод Ньютона.

Знання основ цих методів полегшить майбутньому фахівцю вивчення літератури з безумовної мінімізації, дозволить без особливого напруження зрозуміти суть того чи іншого методу та вибрати відповідний варіант для розв'язання практичної задачі або розробити самому більш зручні модифікації, які будуть краще пристосовані для розв'язання цієї задачі.

При описанні багатьох методів спочатку розглядається випадок квадратичної цільової функції. Може виникнути питання: навіщо витрачати зусилля на мінімізацію квадратичної функції $f(x)=\frac{1}{2}(Ax,x)-(b,x)$, коли ця функція без зусиль мінімізується, даючи в результаті $Ax-b=0$ або $x=A^{-1}b$. Інтерес представляє мінімізація функцій більш складних ніж квадратичні. Квадратичні функції розглядаються з двох причин. По-перше: якщо метод не є придатним для квадратичної функції, то дуже мало шансів, що він придатний для функцій з більш складною структурою. Тому випробування цього методу на квадратичній функції є цілком розумним. По-друге: неквадратичну функцію можна апроксимувати квадратичною функцією, якщо обмежитися в розкладенні у ряд Тейлора членами не вище другого порядку. При цьому мінімум функції може бути отримано за допомогою мінімізації послідовності квадратичних функцій. Такі функції апроксимують вихідну функцію відносно точок, які послідовно наближуються до точки точного мінімуму.

Загальна схема ітераційних методів для розв’язання задачі безумовної мінімізації має вигляд (див. розділ 1, § 3,4):

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)},\ k = 0,1,2,\ldots, \tag{3.1}$$

де $p^{(k)}$ – напрямок спадання функції $f(x)$ (напрямок спуску) в точці $x^{(k)}$, $p^{(k)} \in U\left(x^{(k)}, f\right)$ – множині напрямків зменшення функції $f(x)$ у точці $x^{(k)}$, α_k – параметр, який регулює довжину кроку вздовж $p^{(k)}$.

Методи монотонного спуску

$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right) \tag{3.2}$$

називаються **релаксаційними** методами. Відповідна послідовність $\left\{x^{(k)}\right\}_0^\infty$ називається також релаксаційною. Якщо функція $f(x)$ диференційована в точці $x^{(k)} \in E^n$, то релаксаційність методу (3.1) забезпечується тоді, коли напрямок $p^{(k)}$ складає нетупий кут з напрямком градієнта $f'\left(x^{(k)}\right)$ (див. розділ 1, § 3), тобто $\left(f'(x^{(k)}), p^{(k)}\right) \geq 0$.

Розглянемо основну властивість градієнту функції $f(x)$ в точці x.

Теорема. Нехай $f(x)$ диференційовна в точці $x^{(k)}$ і $f'\left(x^{(k)}\right) \neq 0$. Напрямок найшвидшого зростання функції $f(x)$ в точці $x^{(k)}$ співпадає з напрямком градієнта $f'\left(x^{(k)}\right)$ у цій точці, а напрямок найшвидшого спадання – з напрямком антиградієнта $\left(-f'\left(x^{(k)}\right)\right)$.

Доведення. З диференційовності функції $f(x)$ у точці $x^{(k)}$ витікає, що

$$f\left(x^{(k)}+p^{(k)}\right)-f\left(x^{(k)}\right)=\left(f'\left(x^{(k)}\right),p^{(k)}\right)+o\left(x^{(k)},p^{(k)}\right). \qquad (3.3)$$

Якщо $f'\left(x^{(k)}\right)\neq 0$, то при достатньо малих $\left\|p^{(k)}\right\|$ головна лінійна частина приросту (3.3) буде визначатися диференціалом функції $\left(f'\left(x^{(k)}\right),p^{(k)}\right)$. Справедлива нерівність Коші-Буняковського

$$-\left\|f'\left(x^{(k)}\right)\right\|\cdot\left\|p^{(k)}\right\|\leq\left(f'\left(x^{(k)}\right),p^{(k)}\right)\leq\left\|f'\left(x^{(k)}\right)\right\|\cdot\left\|p^{(k)}\right\|,$$

при цьому, якщо $f'\left(x^{(k)}\right)\neq 0$, права частина нерівності перетворюється на рівність тільки при $p^{(k)}=\beta\, f'\left(x^{(k)}\right)$, де $\beta=const\geq 0$.

Звідси зрозуміло, що при $f'\left(x^{(k)}\right)\neq 0$ напрямок найшвидшого зростання функції $f(x)$ у точці $x^{(k)}$ співпадає з напрямком градієнта $f'\left(x^{(k)}\right)$ у точці $x^{(k)}$, а напрямок найшвидшого спадання – з напрямком антиградієнта $\left(-f'\left(x^{(k)}\right)\right)$.

Умови закінчення ітераційного процесу (3.1)

(Критерії закінчення розрахунків)

На практиці часто використовуються такі умови закінчення розрахунків:

$$\left\|x^{(k+1)}-x^{(k)}\right\|\leq\varepsilon_1, \qquad (3.4)$$

$$\left|f\left(x^{(k+1)}\right)-f\left(x^{(k)}\right)\right|\leq\varepsilon_2, \qquad (3.5)$$

$$\left\|f'\left(x^{(k+1)}\right)\right\|\leq\varepsilon_3. \qquad (3.6)$$

До початку обчислень вибирається одна з умов (3.4)–(3.6) і відповідне їй мале довільне число $\varepsilon_i>0$, $i=\overline{1,3}$. Обчислення закінчуються після $(k+1)$-го кроку, якщо вперше виявляється виконаною обрана умова зупинки. На практиці також використовуються критерії, які складаються в одночасному виконанні двох з умов (3.4)–(3.6) або всіх трьох.

Зрозуміло, що критерій (3.6) відноситься лише до задач безумовної оптимізації. Його виконання означає, що в точці $x^{(k+1)}$ з точністю ε_3 виконано умову стаціонарності (див. розділ 1, § 3).

Замість критеріїв (3.4)–(3.6), які базуються на понятті абсолютної похибки, можна використовувати аналогічні критерії, які базуються на понятті відносної похибки:

$$\left\| x^{(k+1)} - x^{(k)} \right\| \le \delta_1 \left(1 + \left\| x^{(k+1)} \right\| \right),$$

$$\left| f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \right| \le \delta_2 \left(1 + \left| f\left(x^{(k+1)}\right) \right| \right),$$

$$\left\| f'\left(x^{(k+1)}\right) \right\| \le \delta_3 \left(1 + \left| f\left(x^{(k+1)}\right) \right| \right).$$

§1. Метод покоординатного спуску

У практичних задачах оптимізації нерідко зустрічаються випадки, коли функція, що мінімізується, або не має потрібної гладкості, або є гладкою, але обчислення її похідних з потрібною точністю потребує занадто великого об'єму робіт, багато машинного часу. У таких випадках можуть бути корисні методи нульового порядку, тобто методи, які не потребують обчислення похідних. Одним з них є метод покоординатного спуску.

Постановка задачі

$$f(x) \to min,\ x \in E^n .$$

У методі покоординатного спуску за напрямок спуску $p^{(k)}$ вибирається один з координатних векторів $e_1, \dots, e_n$, де $e_i = (0, \dots, 0, \overset{i}{1}, 0, \dots, 0)$. Унаслідок цього в точки $x^{(k)}$ на кожній ітерації змінюється лише одна з координат $x_i^{(k)}$.

Нехай $x^{(0)}$– початкове наближення, а α_0– деяке дійсне число, тоді за методом покоординатного спуску

$$k = 1: \begin{cases} y_1^{(1)} = x^{(1)} = x^{(0)} + \alpha_0 e_1 \\ y_2^{(1)} = x^{(2)} = x^{(1)} + \alpha_1 e_2 \\ \dots \\ y_n^{(1)} = x^{(n)} = x^{(n-1)} + \alpha_{n-1} e_n \end{cases} \tag{3.7}$$

За формулами (3.7) буде реалізовано спуск за n **внутрішніх ітерацій** (n – розмірність простору E^n) з точки $x^{(0)}$ у точку $x^{(n)}$ за ламаною, що складається з відрізків прямих, які паралельні координатним осям (рис. 3.1).

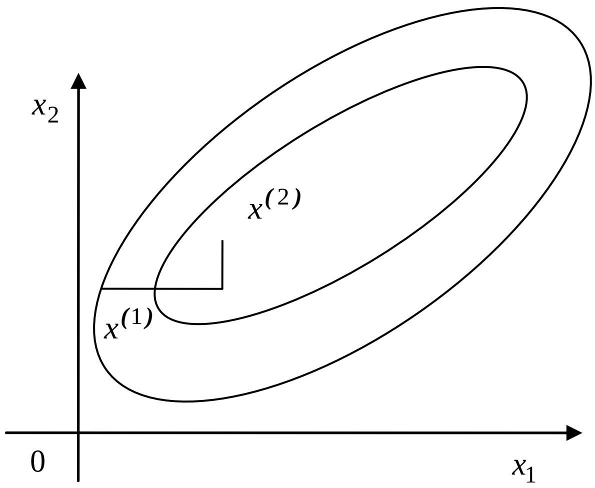


Рис. 3.1

Спуск по усім n координатам за формулами (3.7) складають одну зовнішню ітерацію.

Наступна друга зовнішня ітерація виконується за формулами:

$$k = 2: \begin{cases} y_1^{(2)} = x^{(n+1)} = x^{(n)} + \alpha_n e_1 \\ y_2^{(2)} = x^{(n+2)} = x^{(n+1)} + \alpha_{n+1} e_2 \\ \dots \\ y_n^{(2)} = x^{(2n)} = x^{(2n-1)} + \alpha_{2n-1} e_n \end{cases}$$

і так далі.

Нехай k – номер чергової зовнішньої ітерації ($k = 0,1,\dots$), i – номер тієї координати, по якій проводиться спуск, тобто номер внутрішньої ітерації ($i = 1, 2,\dots, n$). Тоді ітераційна рекурентна формула, яка визначає наступне наближення до точки мінімуму, запишеться у вигляді:

$$y_i^{(2)} = x^{(k-1)n+i} = x^{(k-1)n+i-1} + \alpha_{(k-1)n+i-1} e_i,\ k = 0,1,\dots,\ i = 1, 2,\dots, n \qquad (3.8)$$

Після $i = n$ лічильник числа зовнішніх ітерацій k збільшується на 1, а значення $i = 1$.

До початку обчислень задається мале довільне число $\varepsilon > 0$. Ітераційний процес (3.8) закінчується, коли, наприклад,

$$\left\| y_n^{(k+1)} - y_n^{(k)} \right\| \le \varepsilon .$$

Існують різні способи вибору величини α_k на k-ій ітерації. Основна задача при виборі α_k в релаксаційних процесах мінімізації: забезпечити виконання нерівності $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$.

Розглянемо деякі способи вибору параметра α_k у методі покоординатного спуску.

Спосіб 1. Вибір параметра α_k з умови мінімізації цільової функції вздовж напрямку $p^{(k)}$.

Величини $\alpha_{(k-1)n+i-1}$ в методі (3.8) можна визначити з умови

$$f\left(x^{(k-1)n+i-1} + \alpha_{(k-1)n+i-1} e_i\right) = \min_{\alpha \in E^1} f\left(x^{(k-1)n+i-1} + \alpha\, e_i\right). \qquad (3.9)$$

Спосіб 2. Нехай $\alpha_k = \alpha_{k-1} > 0$. Обчислимо значення функції в точці $x = x^{(k)} + \alpha_k p^{(k)}$ і перевіримо виконання нерівності

$$f\left(x^{(k)} + \alpha_k p^{(k)}\right) < f\left(x^{(k)}\right). \qquad (3.10)$$

Якщо (3.10) виконується, то або приймемо $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, $\alpha_{k+1} = \alpha_k$ і перейдемо до наступної $(k+2)$-ої ітерації, або виберемо $\alpha_k = 2\alpha_{k-1}$. Якщо значення $f(x)$ менше за попереднє, то процес подвоєння можна продовжувати до тих пір, поки буде виконуватися (3.10).

У тому випадку, коли (3.10) не виконується, то обчислимо значення функції $f(x)$ у точці $x = x^{(k)} - \alpha_k p^{(k)}$ і перевіримо виконання нерівності

$$f(x^{(k)} - \alpha_k p^{(k)}) < f(x^{(k)}). \qquad (3.11)$$

У випадку виконання (3.11) або приймаємо $x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)}$, $\alpha_{k+1} = \alpha_k$ і перейдемо до наступної $(k+2)$-ої ітерації, або виберемо, як і в попередньому випадку, $\alpha_k = 2\alpha_{k-1}$. Якщо значення $f(x)$ менше за попереднє, то процес подвоєння можна продовжувати до тих пір, доки буде виконуватися (3.11).

Якщо за одну зовнішню ітерацію, яка складається з n внутрішніх ітерацій при переборі напрямків усіх координатних осей $e_1,\ldots,e_n$ з кроком α_k реалізувалася хоча б одна вдала ітерація, то довжина кроку α_k не дробиться і зберігається на протязі наступного циклу з n ітерацій. Якщо ж серед останніх n ітерацій не було ні однієї вдалої, то крок α_k дробиться: $\alpha_k = \frac{1}{2}\alpha_{k-1}$, і здійснюється перехід до наступної зовнішньої ітерації.

Збіжність методу покоординатного спуску

Відзначимо, що хоч метод (3.8) з вибором α_k способом подвоєння для своєї реалізації не потребує знання градієнта функції, яка мінімізується, але виявляється, що коли функція $f(x)$ не є гладкою, то метод покоординатного спуску може не збігатися до множини точок мінімуму X_* функції $f(x)$ на E^n.

Теорема. Нехай функція $f(x)$ опукла на E^n і належить класу $C^1(E^n)$, а початкове наближення $x^{(0)}$ таке, що множина $M(x^{(0)}) = \{x \in E^n : f(x) \le f(x^{(0)})\}$ обмежена, тоді послідовність $\{x^{(k)}\}$, яка отримана методом (3.8), (3.10), (3.11), мінімізує функцію $f(x)$ на E^n і збігається до множини точок мінімуму X_*.

Алгоритм методу покоординатного спуску

Початковий етап. Вибрати число $\varepsilon > 0$ для зупинки алгоритму, початкову точку $x^{(0)}$, покласти $z_1 = x^{(0)}$, $k = i = 1$ і перейти до основного етапу.

Основний етап.

Крок 1. Вибрати $\alpha_{(k-1)n+i-1}$ таке, щоб виконувалася нерівність

$$f\left(x^{(k-1)n+i-1} + \alpha_{(k-1)n+i-1}e_i\right) < f\left(x^{(k-1)n+i-1}\right).$$

Покласти $x^{(k-1)n+i} = x^{(k-1)n+i-1} + \alpha_{(k-1)n+i-1}e_i$. Якщо $i < n$, то замінити i на $i+1$ та повернутися до кроку 1, якщо $i = n$, то перейти до кроку 2.

Крок 2. Покласти $z_{k+1} = x^{(kn)}$. Якщо $\|z_{k+1} - z_k\| < \varepsilon$, то зупинитися. У протилежному випадку покласти $k = k+1,\ i = 1$та перейти до кроку 1.
Алгоритм описаний.

Приклад. Розглянемо таку задачу: знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x = (x_1, x_2) \in E^2.$$

Зазначимо, що оптимальним розв'язком цієї задачі є точка $x_* = (2.00, 1.00)$, в якій значення цільової функції $f(x_*) = 0$. У табл. 3.1 наведені результати обчислень за методом покоординатного спуску для початкової точки $x^{(0)} = (0.00, 3.00)$. Величина α_k на кожній ітерації отримана внаслідок одновимірної оптимізації функції f уздовж одного з напрямків e_i, $i = \overline{1,n}$, тобто за першим способом. Укажемо також, що помітне зменшення функції отримано за кілька перших ітерацій, тоді як на останніх ітераціях процес явно уповільнюється. Після семи зовнішніх ітерацій отримана точка (2.22, 1.11), значення цільової функції в якій дорівнює 0.0023.

Таблиця 3.1

k	z_k $f(z_k)$	i	e_i	$x^{(k-1)n+i-1}$	$\alpha_{(k-1)n+i-1}$
1	(0.00,3.00) 52	1 2	(1.0,0.0) (0.0,1.0)	(0.00,3.00) (3.13,3.00)	3.13 –1.44
2	(3.13,1.56) 1.63	1 2	(1.0,0.0) (0.0,1.0)	(3.13,1.56) (2.63,1.56)	–0.50 –0.25
3	(2.63,1.31) 0.16	1 2	(1.0,0.0) (0.0,1.0)	(2.63,1.31) (2.44,1.31)	–0.19 –0.09
4	(2.44,1.22) 0.04	1 2	(1.0,0.0) (0.0,1.0)	(2.44,1.22) (2.35,1.22)	–0.09 –0.05
5	(2.35,1.17) 0.015	1 2	(1.0,0.0) (0.0,1.0)	(2.35,1.17) (2.29,1.17)	–0.06 –0.03
6	(2.29,1.14) 0.007	1 2	(1.0,0.0) (0.0,1.0)	(2.29,1.14) (2.25,1.14)	–0.04 –0.02

Закінчення табл. 3.1

k	z_k $f(z_k)$	i	e_i	$x^{(k-1)n+i-1}$	$\alpha_{(k-1)n+i-1}$
7	(2.25,1.12) 0.004	1 2	(1.0,0.0) (0.0,1.0)	(2.25,1.12) (2.22,1.12)	–0.03 –0.01
8	(2.22,1.11) 0.0023				

На рис. 3.2 показано лінії рівня цільової функції, які отримані за методом покоординатного спуску. Уповільнення на останніх ітераціях пояснюється тим, що вздовж дна яру, який показано пунктирною лінією, робляться дуже маленькі кроки за ортогональними напрямками.

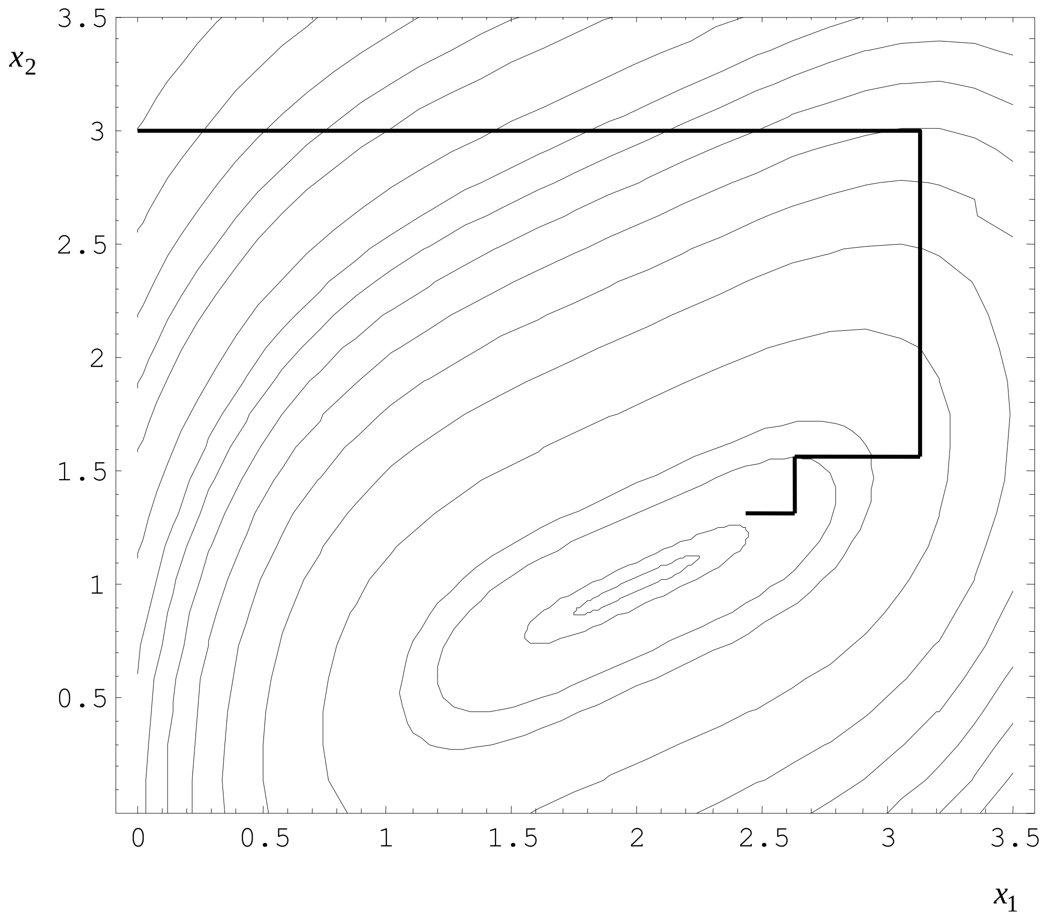


Рис.3.2

§2. Метод конфігурацій, метод Розенброка

Продовжимо вивчення методів нульового порядку для задачі безумовної оптимізації

$$f(x) \to min, \; x \in E^n.$$

1. Метод конфігурацій (метод Хука і Дживса)

Метод Хука і Дживса здійснює два типи пошуку – пошук, що досліджує, та пошук за зразком. У цьому методі спочатку досліджується окіл вибраної точки $x^{(0)}$ (змінюючи по одній або по декілька значень компонент вектора $x^{(0)}$). Після того, як знайдено задовільний напрямок (тобто напрямок зменшення цільової функції), в цьому напрямку проводиться обчислення функції при кроці, який поступово збільшується. Тим самим установлюється **конфігурація пошуку**. Цей процес продовжується доти, доки пошук у цьому напрямку продовжує приводити до точок з меншим значенням цільової функції $f(x)$. Коли в напрямку встановленої конфігурації не вдається знайти точку з меншим значенням цільової функції $f(x)$, то розмір кроку зменшують. Після кількох послідовних зменшень кроку від прийнятої конфігурації відмовляються та починають нове дослідження околу.

Таким чином, згідно цьому методу робляться спроби знайти напрямок яру цільової функції, з тим, щоб рухатись по цьому яру. Перевагою методу є те, що він дозволяє пошуком під час дослідження відновлювати напрямок руху вздовж яру в тих випадках, коли внаслідок скривлення яру встановлена конфігурація втрачається.

Алгоритм методу Хука і Дживса

Нехай $e_1,\ldots,e_n$ – координатні вектори.

Початковий етап. Вибрати число $\varepsilon > 0$ для зупинки алгоритму, початкову точку $x^{(1)}$, покласти $y_1 = x^{(1)}$, $k = j = 1$ і перейти до основного етапу.

Основний етап.

Крок 1. Вибрати α_j – оптимальний розв'язок задачі мінімізації функції $f(y_j + \alpha e_j)$ при умові, що $\alpha \in E^1$ і покласти $y_{j+1} = y_j + \alpha e_j$.

Якщо $j < n$, то замінити j на $j+1$ та повернутися до кроку 1.

Якщо $j = n$, то покласти $x^{(k+1)} = y_{n+1}$.

Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| < \varepsilon$, то зупинитися, у протилежному випадку перейти до кроку 2.

Крок 2. Покласти $p = x^{(k+1)} - x^{(k)}$ і знайти $\hat{\alpha}$ – оптимальний розв'язок задачі мінімізації функції $f\left(x^{(k)} + \alpha p\right)$ при умові $\alpha \in E^1$. Покласти $j = 1$, $y_1 = x^{(k+1)} + \hat{\alpha} p$, замінити k на $k+1$ та перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо таку задачу: знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x = (x_1, x_2) \in E^2.$$

Оптимальним розв'язком цієї задачі є точка $x_* = (2.00, 1.00)$, $f(x_*) = 0$. У табл. 3.2 наведені результати обчислень за методом Хука і Дживса для початкової точки $x^{(0)} = (0.00, 3.00)$. На кожній ітерації пошук, що досліджує, по координатним напрямкам $e_1 = (1.0,\ 0.0)$, $e_2 = (0.0,\ 1.0)$ дає точки y_2 та y_3, а пошук за зразком у напрямку $p = x^{(k+1)} - x^{(k)}$ дає точку y_1. Виняток становить перша ітерація, де $y_1 = x^{(1)}$. Відзначимо, що треба було виконати чотири ітерації для переходу з початкової точки до оптимальної $x^* = (2.00,\ 1.00)$. У цій точці $\left\| x^{(5)} - x^{(4)} \right\| < 2 \cdot 10^{-3}$ і процедура знаходження точки мінімуму цільової функції була зупинена.

На рис. 3.3 показані точки, які отримані за методом Хука і Дживса. Відзначимо, що пошук за зразком покращує збіжність у результаті руху вздовж напрямку, майже паралельному дну яра.

Таблиця 3.2

k	$x^{(k)}$ $f(x^{(k)})$	j	y_j	α_j	y_{j+1}	p	$\hat{\alpha}$
1	(0.00,3.00) 52	1 2	(0.00,3.00) (3.13,3.00)	3.13 –1.44	(3.13,3.00) (3.13,1.56)	– (3.13, –1.44)	– –0.10
k	$x^{(k)}$ $f(x^{(k)})$	j	y_j	α_j	y_{j+1}	p	$\hat{\alpha}$
2	(3.13,1.56) 1.63	1 2	(2.82,1.70) (2.70,1.70)	–0.12 –0.35	(2.70,1.70) (2.70,1.35)	– (–0.43,–0.21)	– –1.50
3	(2.70,1.35) 0.24	1 2	(2.06,1.04) (2.04,1.04)	–0.02 –0.02	(2.04,1.04) (2.04,1.02)	– (–0.66,–0,33)	– 0.06
4	(2.04,1.02) 0.000003	1 2	(2.00,1.00) (2.00,1.00)	0.00 0.00	(2.00,1.00) (2.00,1.00)	–	–
5	(2.00,1.00) 0.0	1 2					

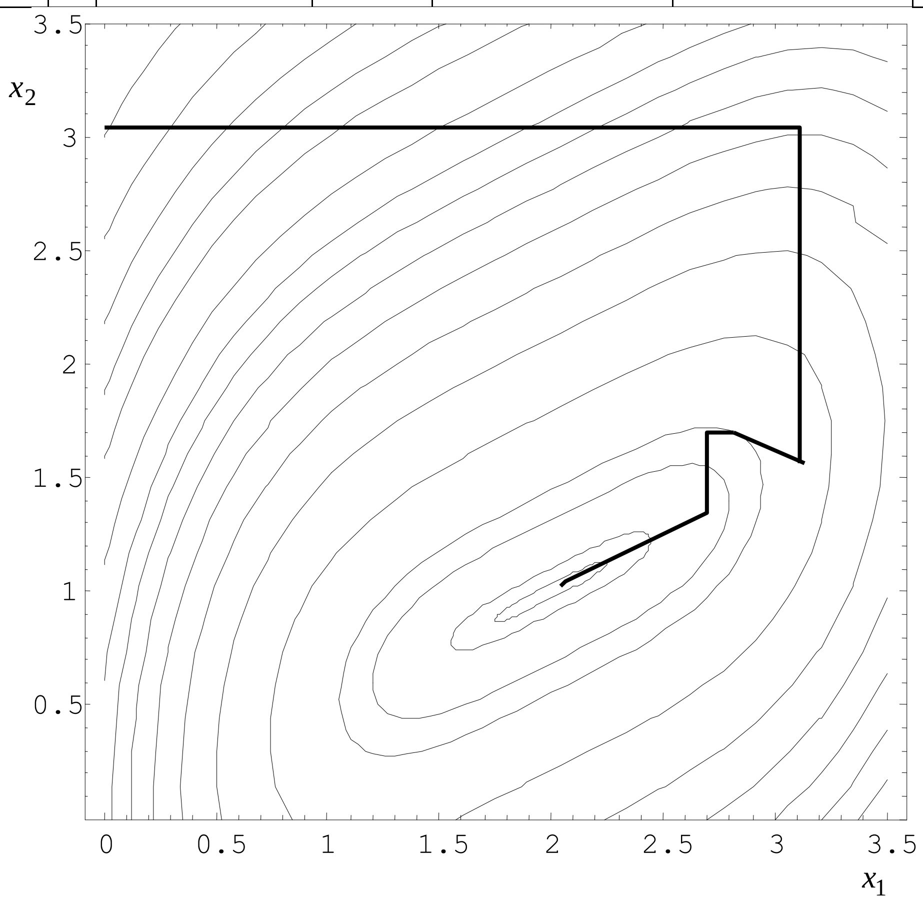


Рис.3.3

2.Метод Розенброка

На кожній ітерації методу здійснюється пошук уздовж взаємно ортогональних напрямків. Коли отримана нова точка в кінці ітерації, будується нова множина ортогональних векторів (рис. 3.4). Якщо, наприклад, цільова функція має вузький скривлений гребінь, то пошук за n взаємно ортогональним напрямкам ефективний тим, що результуючий напрямок прагне розташуватися вздовж осі яру.

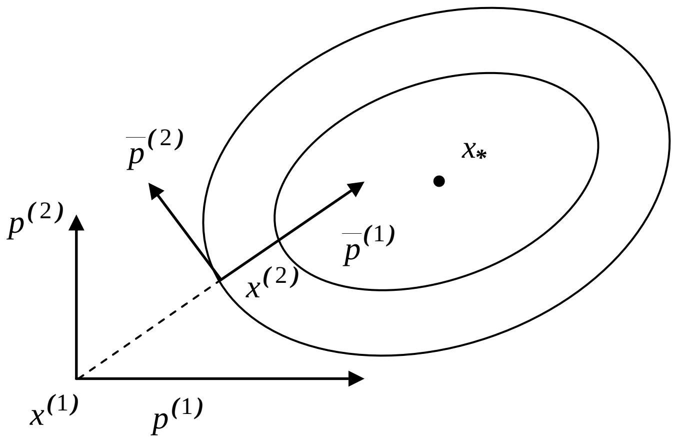


Рис. 3.4

Побудова напрямків пошуку. Нехай $p^{(1)},\ldots,p^{(n)}$ – лінійно незалежні вектори, норма яких дорівнює одиниці. Припустимо, що ці вектори взаємно ортогональні, тобто $\left(p^{(i)},p^{(j)}\right)=0,\ i\neq j$. Починаючи з поточної точки $x^{(k)}$, цільова функція послідовно мінімізується вздовж кожного з напрямків, у результаті чого отримаємо точку $x^{(k+1)}$. Зокрема, $x^{(k+1)}-x^{(k)}=\sum_{j=1}^{n}\alpha_j p^{(j)}$, де α_j – довжина кроку в напрямку $p^{(j)}$. Новий набір напрямків $\bar{p}^{(1)},\ldots,\bar{p}^{(n)}$ будується за допомогою процедури Грама-Шмідта:

$$a_j=\begin{cases} p^{(j)}, \text{якщо} \quad \alpha_j=0, \\ \sum\limits_{j=1}^{n}\alpha_j p^{(j)}, \text{якщо} \quad \alpha_j\neq 0, \end{cases}$$

$$b_j = \begin{cases} a_j, \text{якщо} \quad j = 1, \\ a_j - \sum_{i=1}^{j-1} (a_j, \overline{p}^{(j)}) \overline{p}^{(i)}, \text{якщо} \quad j \ge 2, \end{cases} \tag{3.12}$$

$$\overline{p}^{(j)} = \frac{b_j}{\|b_j\|}.$$

Нові напрямки $\overline{p}^{(j)}$, які отримані за формулами (3.12), також будуть лінійно незалежними та ортогональними.

Алгоритм методу Розенброка

Початковий етап. Вибрати число $\varepsilon > 0$ для зупинки алгоритму, $p^{(1)},\ldots,p^{(n)}$ – координатні напрямки, початкову точку $x^{(1)}$, покласти $y_1 = x^{(1)}$, $k = j = 1$ і перейти до основного етапу.

Основний етап.

Крок 1. Вибрати α_j – оптимальний розв'язок задачі мінімізації функції $f(y_j + \alpha\, p_j)$ при умові, що $\alpha \in E^1$ і покласти $y_{j+1} = y_j + \alpha\, p_j$.

Якщо $j < n$, то замінити j на $j+1$ та повернутися до кроку 1, у протилежному випадку перейти до кроку 2.

Крок 2. Покласти $x^{(k+1)} = y_{n+1}$. Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| < \varepsilon$, то зупинитися, у протилежному випадку покласти $y_1 = x^{(k+1)}$, $j = 1$ замінити k на $k+1$ та перейти до кроку 3.

Крок 3. Побудувати нову множину лінійно незалежних та взаємно ортогональних напрямків відповідно до (3.12). Позначити нові напрямки як $p^{(1)},\ldots,p^{(n)}$ та перейти до кроку 1.

Алгоритм описаний.

Приклад. Знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2, \; x = (x_1, x_2) \in E^2.$$

У табл. 3.3 наведені результати обчислень за методом Розенброка для початкової точки $x^{(0)} = (0.00, 3.00)$. Точка y_2 отримана мінімізацією функції вздовж напрямку $p^{(1)}$ при початковій точці y_1, а y_3 отримана мінімізацією цільової функції вздовж напрямку $p^{(2)}$ при початковій точці y_2. Після першої ітерації маємо $\alpha_1 = 3.13$ та $\alpha_2 = -1.44$. Використовуючи формули (3.12), за нові напрямки пошуку отримаємо (0.91,–0.42) та (–0.42,–0.91). Після чотирьох ітерацій отримана точка (2.21,1.10), в якій значення цільової функції дорівнює 0.002. При цьому $\left\| x^{(4)} - x^{(3)} \right\| < 0.15$ і процедура була зупинена.

На рис. 3.5 показано процес мінімізації. Цікаво порівняти цей рисунок з рисунком 3.3.

Таблиця 3.3

k	$x^{(k)}$ $f(x^{(k)})$	j	y_j $f(y_j)$	α_j	p_j	y_{j+1} $f(y_{j+1})$
1	(0.00,3.00) 52.00	1	(0.00,3.00) 52.00	3.13	(1.0,0.0)	(3.13,3.00) 9.87
		2	(3.13,3.00) 9.87	–1.44	(0.0,1.0)	(3.13,1.56) 1.63
2	(3.13,1.56) 1.63	1	(3.13,1.56) 1.63	–0.34	(0.91,–0.42)	(2.82,1.70) 0.79
		2	(2.82,1.70) 0.79	0.51	(–0.42,–0.91)	(2.61,1.24) 0.16
3	(2.61,1.24) 0.16	1	(2.61,1.24) 0.16	0.38	(–0.85,–0.52)	(2.23,1.04) 0.05
		2	(2.29,1.04) 0.05	–0.10	(0.52,–0.85)	(2.24,1.13) 0.004
4	(2.24,1.13) 0.004	1	(2.24,1.13) 0.004	0.04	(–0.96,–0.28)	(2.20,1.12) 0.003
		2	(2.20,1.12) 0.003	0.02	(0.28,–0.96)	(2.21,1.10) 0.002

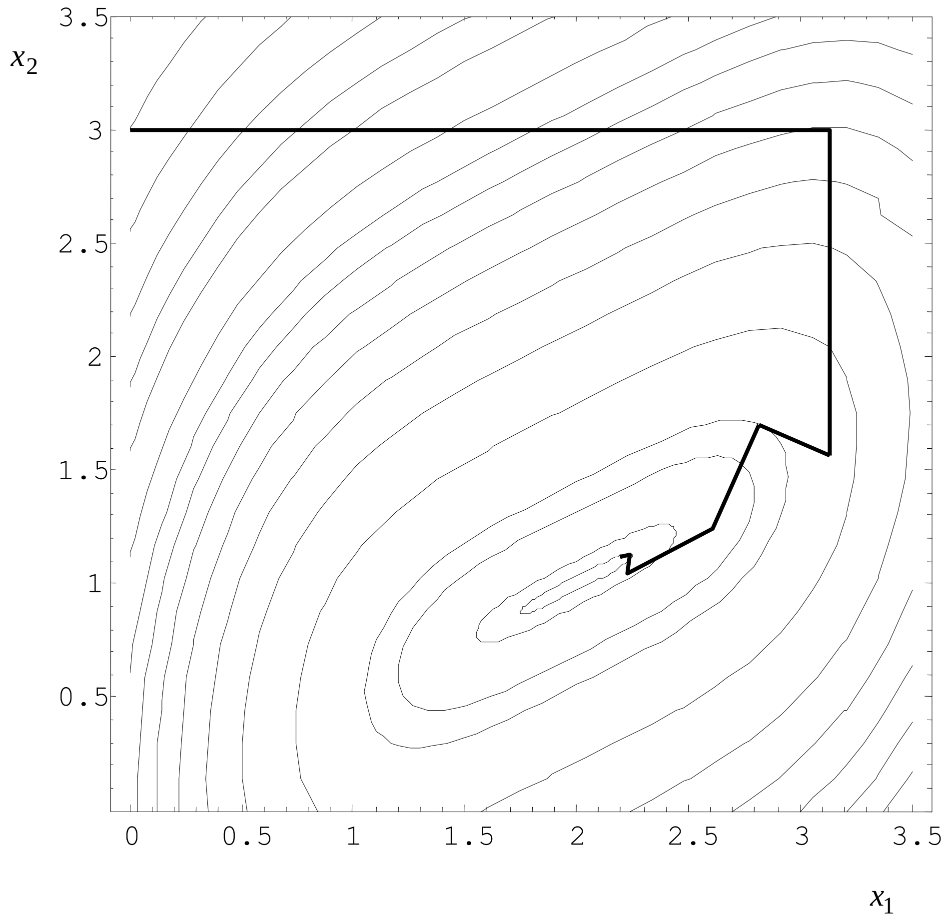


Рис.3.5

§3. Метод пошуку по деформованому багатограннику (Метод Нелдера–Міда)

Правильні (регулярні) багатогранники у просторі E^n є **симплексами**. Для випадку двох змінних регулярний симплекс представляє собою рівносторонній трикутник, у випадку трьох змінних – тетраедр. Кількість вершин симплекса на одиницю більша розмірності простору задачі. Розглянемо основні етапи методу, в якому використовуються симплекси.

1. Вибирається початковий симплекс та обчислюються значення цільової функції в його вершинах.
2. Від вершини симплекса з найбільшим значенням цільової функції здійснюється перехід через центр тяжіння симплекса до нової точки, яка симетрична даній. Знаходиться значення функції в новій вершині.
3. Перевіряється критерій закінчення розрахунків для нового симплекса.

Зауважимо, що сформульований метод може зациклюватися в околі точки мінімуму.

Методом, який усуває вказаний недолік і враховує кривизну ліній рівня цільової функції, є метод пошуку по деформованому багатограннику.

Розглянемо задачу мінімізації функції n незалежних змінних з використанням $n+1$ вершини деформованого багатогранника в E^n.

Нехай точка $x_i^{(k)} = \left(x_{i_1}^{(k)}, \ldots, x_{i_n}^{(k)}\right)$, $i = \overline{1, n+1}$ є i-ю вершиною багатогранника в E^n на k-м етапі пошуку $k = 0, 1, \ldots$ і нехай значення цільової функції в точці $x_i^{(k)}$ дорівнює $f(x_i^{(k)})$. Визначимо

$$f\left(x_h^{(k)}\right) = max\left\{ f\left(x_1^{(k)}\right), \ldots, f\left(x_{n+1}^{(k)}\right) \right\}, \; f\left(x_l^{(k)}\right) = min\left\{ f\left(x_1^{(k)}\right), \ldots, f\left(x_{n+1}^{(k)}\right) \right\},$$

де $f\left(x_h^{(k)}\right)$, $f\left(x_l^{(k)}\right)$ – вершини багатогранника, в яких функція досягає відповідно свого максимального та мінімального значення.

Координати центру тяжіння вершин визначаються за формулою

$$x_{n+2,j}^{(k)} = \frac{1}{n}\left[\left(\sum_{i=1}^{n+1} x_{i,j}^{(k)}\right) - x_{h,j}^{(k)}\right], \; j = \overline{1, n}, \tag{3.13}$$

де індекс j позначає координатний напрямок.

Початковий багатогранник звичайно вибирається у вигляді регулярного симплекса, але це не обов'язково. Процедура відшукування вершини у просторі E^n, в якій функція $f(x)$ має «краще» значення, складається з описаних нижче операцій.

1. **Відображення** – проектування $x_h^{(k)}$ через центр тяжіння у відповідності з співвідношенням

$$x_{n+3}^{(k)} = x_{n+2}^{(k)} + \alpha\left(x_{n+2}^{(k)} - x_h^{(k)}\right), \tag{3.14}$$

де $\alpha > 0$ є коефіцієнтом відображення.

2. **Розтягання.** Якщо $f\left(x_{n+3}^{(k)}\right) \le f\left(x_l^{(k)}\right)$, то вектор $\left(x_{n+3}^{(k)} - x_{n+2}^{(k)}\right)$ розтягається у відповідності з співвідношенням

$$x_{n+4}^{(k)} = x_{n+2}^{(k)} + \gamma\left(x_{n+3}^{(k)} - x_{n+2}^{(k)}\right), \tag{3.15}$$

де $\gamma > 1$ є коефіцієнтом розтягнення.

Якщо $f\left(x_{n+4}^{(k)}\right) < f\left(x_l^{(k)}\right)$, то $x_h^{(k)}$ замінюється на $x_{n+4}^{(k)}$ і процедура продовжується з операції відображення при $k = k+1$. У протилежному випадку $x_h^{(k)} = x_{n+3}^{(k)}$ і також перехід до операції відображення при $k = k+1$.

3. **Стискання**. Якщо $f\left(x_{n+3}^{(k)}\right) > f\left(x_i^{(k)}\right)$ для всіх $i \ne h$, то вектор $\left(x_h^{(k)} - x_{n+2}^{(k)}\right)$ стискається відповідно з співвідношенням

$$x_{n+5}^{(k)} = x_{n+2}^{(k)} + \beta\left(x_h^{(k)} - x_{n+2}^{(k)}\right), \tag{3.16}$$

де $0 < \beta < 1$ є коефіцієнтом стискання. Потім замінюємо $x_h^{(k)}$ на $x_{n+5}^{(k)}$ і повертаємось до операції відображення для продовження пошуку на $(k+1)$–ому кроці.

4. **Редукція**. Якщо $f\left(x_{n+5}^{(k)}\right) > f\left(x_h^{(k)}\right)$, то всі вектори $\left(x_i^{(k)} - x_l^{(k)}\right)$ зменшуються в два рази відповідно з співвідношенням

$$x_i^{(k)} = x_l^{(k)} + \frac{1}{2}\left(x_i^{(k)} - x_l^{(k)}\right),\ i = \overline{1, n+1}. \tag{3.17}$$

Потім повертаємось до операції відображення для продовження пошуку на $(k+1)$-ому кроці.

Критерій закінчення пошуку, який використовується Нелдером та Мідом, складається у перевірці умови

$$\sqrt{\frac{1}{n+1}\sum_{i=1}^{n+1}\left[f\left(x_i^{(k)}\right) - f\left(x_i^{(k)}\right)\right]^2} \le \varepsilon. \tag{3.18}$$

Деформівний багатогранник у протилежність жорсткому симплексу адаптується до топографії цільової функції, витягуючись уздовж похилих поверхонь, змінюючи напрямок у вигнутих впадинах та стискаючись в околі мінімуму.

Коефіцієнти α, β, γ звичайно вибираються такими: $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$.

Алгоритм методу Нелдера-Міда

Початковий етап. Задати $n -$ розмір простору, число $\varepsilon > 0$ для зупинки алгоритму. Вибрати початковий багатогранник. Обчислити значення цільової функції у вершинах початкового багатогранника.

Основний етап

Крок 1. Позначити через $x_h^{(k)}$ – вершину багатогранника, в якій значення функції $f\left(x_h^{(k)}\right) = max\left\{f\left(x_1^{(k)}\right), \ldots, f\left(x_{n+1}^{(k)}\right)\right\}$ і через $x_l^{(k)}$ – вершину багатогранника, в якій значення функції $f\left(x_l^{(k)}\right) = min\left\{f\left(x_1^{(k)}\right), \ldots, f\left(x_{n+1}^{(k)}\right)\right\}$.

Крок 2. Знайти координати центру тяжіння та значення цільової функції в цій точці

$$x_{n+2,j}^{(k)} = \frac{1}{n}\left[\left(\sum_{i=1}^{n+1} x_{i,j}^{(k)}\right) - x_{h,j}^{(k)}\right], \ j = \overline{1,n}.$$

Крок 3. Перевірити виконання критерію закінчення (3.18):

$$\sqrt{\frac{1}{n+1}\sum_{i=1}^{n+1}\left[f\left(x_i^{(k)}\right) - f\left(x_i^{(k)}\right)\right]^2} \le \varepsilon.$$

Якщо критерій виконано, то кінець: $x_* \approx x_l^{(k)}$, $f(x_*) \approx f\left(x_l^{(k)}\right)$, інакше перейти до кроку 4.

Крок 4. Виконати операцію відображення і знайти точку $x_{n+3}^{(k)}$ за формулою $x_{n+3}^{(k)} = x_{n+2}^{(k)} + \alpha\left(x_{n+2}^{(k)} - x_h^{(k)}\right)$. Обчислити значення функції $f\left(x_{n+3}^{(k)}\right)$.

Крок 5. Якщо маємо $f\left(x_{n+3}^{(k)}\right) \le f\left(x_l^{(k)}\right)$, то перейти до кроку 6, інакше перейти до кроку 8.

Крок 6. Виконати операцію розтягнення і знайти точку $x_{n+4}^{(k)}$:

$x_{n+4}^{(k)} = x_{n+2}^{(k)} + \gamma\left(x_{n+3}^{(k)} - x_{n+2}^{(k)}\right)$. Обчислити значення функції $f\left(x_{n+4}^{(k)}\right)$.

Крок 7. Якщо $f\left(x_{n+4}^{(k)}\right) < f\left(x_l^{(k)}\right)$, то присвоїти $x_h^{(k)} = x_{n+4}^{(k)}$ і перейти до кроку 12, інакше $x_h^{(k)}$ замінити на $x_{n+3}^{(k)}$ і перейти до кроку 12.

Крок 8. Якщо $f\left(x_{n+3}^{(k)}\right)> f\left(x_i^{(k)}\right)$ для всіх $i \neq h$, то перейти до кроку 9, інакше $x_h^{(k)}$ замінити на $x_{n+3}^{(k)}$ і перейти до кроку 12.

Крок 9. Якщо $f\left(x_{n+3}^{(k)}\right)< f\left(x_h^{(k)}\right)$, то перейти до кроку 10, інакше $x_h^{(k)}$ замінити на $x_{n+3}^{(k)}$ і перейти до кроку 10.

Крок 10. Виконати операцію стискання і знайти точку $x_{n+5}^{(k)}$ за формулою $x_{n+5}^{(k)} = x_{n+2}^{(k)} + \beta\left(x_h^{(k)} - x_{n+2}^{(k)}\right)$. Обчислити значення функції $f\left(x_{n+5}^{(k)}\right)$.

Крок 11. Якщо $f\left(x_{n+5}^{(k)}\right)> f\left(x_h^{(k)}\right)$, то виконати операцію редукції і обчислити нові значення $x_i^{(k)}$ та $f\left(x_i^{(k)}\right)$:

$$x_i^{(k)} = x_l^{(k)} + \frac{1}{2}\left(x_i^{(k)} - x_l^{(k)}\right),\ i = \overline{1, n+1},$$

інакше $x_h^{(k)}$ замінити на $x_{n+5}^{(k)}$.

Крок 12. Покладемо $x_i^{(k+1)} = x_i^{(k)}$, $f\left(x_i^{(k+1)}\right)= f\left(x_i^{(k)}\right)$. Присвоїти k значення $k+1$ та перейти до кроку 1.

Алгоритм описаний.

Приклад. Знайдемо точку мінімуму функції двох змінних

$$f(x_1, x_2) = 4(x_1 - 5)^2 + (x_2 - 6)^2$$

методом Нелдера-Міда при $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$, $\varepsilon = 0.01$.

Оптимальним розв'язком цієї задачі є точка $x_* = (5.00, 6.00)$, $f(x_*) = 0$.

Нехай $x_1^{(0)} = (8,9)$, $f\left(x_1^{(0)}\right) = 45$, $x_2^{(0)} = (10,11)$, $f\left(x_2^{(0)}\right) = 125$, $x_3^{(0)} = (8,11)$, $f\left(x_3^{(0)}\right) = 65$.

Знайдемо координати центру тяжіння: $x_4^{(0)} = (8,10)$, $f\left(x_4^{(0)}\right) = 52$.

Координати точки $x_5^{(0)} = 2x_4^{(0)} - x_h^{(0)} = (6,9)$, $f\left(x_5^{(0)}\right) = 13$. Так як умова $f\left(x_{n+5}^{(k)}\right) \leq f\left(x_h^{(k)}\right)$ виконана, проводиться операція розтягнення $x_6^{(0)} = (4,8)$, $f\left(x_6^{(0)}\right) = 8$. Згідно умові $f\left(x_6^{(k)}\right) < f\left(x_l^{(k)}\right)$ покладаємо $x_h^{(0)} = x_6^{(0)} = (4,8)$. Перевіряємо критерій закінчення процесу. Нові вершини трикутника мають координати: $x_1^{(1)} = x_1^{(0)} = (8,9)$, $x_2^{(1)} = x_6^{(0)} = (4,8)$, $x_3^{(1)} = x_3^{(0)} = (8,11)$. Подальші обчислення наведені в таблиці 3.4. Точки $x_h^{(k)}$ та $x_l^{(k)}$ підкреслені рискою зверху і знизу відповідно.

Таблиця 3.4

k	$x_1^{(k)}$ $f(x_1^{(k)})$	$x_2^{(k)}$ $f(x_2^{(k)})$	$x_3^{(k)}$ $f(x_3^{(k)})$	$x_4^{(k)}$ $f(x_4^{(k)})$	$x_5^{(k)}$ $f(x_5^{(k)})$	$x_6^{(k)}$ $f(x_6^{(k)})$	$x_7^{(k)}$ $f(x_7^{(k)})$
0	$\underline{(8,9)}$ 45	$\overline{(10,11)}$ 125	(8,11) 65	(8,10) 52	(6,9) 13	(4,8) 8	
1	(8,9) 45	$\underline{(4,8)}$ 8	$\overline{(8,11)}$ 65	(6,8.5) 10.25	(4,6) 4	(2,3.5) 42.25	
2	$\overline{(8,9)}$ 45	(4,8) 8	$\underline{(4,6)}$ 4	(4,7) 5	(0,5) 101		(6,8) 8
3	$\overline{(6,8)}$ 8	(4,8) 8	$\underline{(4,6)}$ 4	(4,7) 5	(2,6) 36		(5,7.5) 2.25
4	$\underline{(5,7.5)}$ 2.25	$\overline{(4,8)}$ 8	(4,6) 4	(4.5,6.75) 1.5625	(5,5.5) 0.25		
5	(5,7.5) 2.25	$\underline{(5,5.5)}$ 0.25	$\overline{(4,6)}$ 4	(5,6.5) 0.0625	(6,7) 5		
6	(5,7.5) 2.25	(5,5.5) 0.25	(4.5,6,25) 1.0625				

§4. Градієнтні методи

Постановка задачі

$$f(x) \to min, \ x \in E^n.$$

На відміну від методів нульового порядку будемо припускати, що $f(x)$ диференційовна на E^n.

У градієнтних методах за напрямок спуску $p^{(k)}$ з точки $x^{(k)}$ (див. формулу (3.1)) вибирається антиградієнт функції $f(x)$ у точці $x^{(k)}$, тобто

$$x^{(k+1)} = x^{(k)} - \alpha_k f'\left(x^{(k)}\right),\ k = 0,1,2,\ldots,\ \alpha_k > 0 \qquad (3.19)$$

або в координатній формі

$$x_i^{(k+1)} = x_i^{(k)} - \alpha_k \frac{\partial f\left(x^{(k)}\right)}{\partial x_i},\ i = 1,2,\ldots,n.$$

Різні способи вибору величини α_k у методі (3.19) визначають різні варіанти градієнтних методів. Укажемо кілька найбільш уживаних на практиці способів вибору крокового множника.

1. Метод найшвидшого спуску

На промені $\left\{x \in E^n : x = x^{(k)} - \alpha f'\left(x^{(k)}\right), \alpha \geq 0\right\}$, який направлений за антиградієнтом з точки $x^{(k)}$, введемо функцію $g_k(\alpha) = f\left(x^{(k)} - \alpha f'\left(x^{(k)}\right)\right)$, $\alpha \geq 0$ і визначимо α_k з умови

$$g_k(\alpha_k) = \min_{\alpha \geq 0} g_k(\alpha),\quad \alpha_k > 0. \qquad (3.20)$$

Метод (3.19), (3.20), в якому кроковий множник α_k вибирається з умови мінімізації функції $f(x)$ вздовж напрямку антиградієнта, носить назву **методу найшвидшого спуску**. У тих випадках, коли точне визначення величини α_k з умови (3.20) є важким або неможливим (наприклад, мінімум в (3.20) при деяких k не досягається), удаються до інших способів вибору α_k.

2. Градієнтний метод з дробленням кроку

Процес (3.19) з дробленням кроку проходить таким чином. Вибираються певні константи $\beta > 0$, $0 < \lambda < 1$ (часто $\lambda = \frac{1}{2}$). Для коефіцієнта $\alpha = \beta$ перевіряється умова монотонності

$$f\left(x^{(k)} - \alpha f'\left(x^{(k)}\right)\right) < f\left(x^{(k)}\right). \qquad (3.21)$$

Якщо воно виконано, то покладають $\alpha_k = \alpha$. Якщо ні, то проводиться дроблення кроку, тобто приймається $\alpha = \lambda\,\beta$, і знову перевіряється виконання умови (3.21). Процес дроблення продовжується доти, поки умова (3.21) не виявиться виконаною. Цей процес не може бути нескінченим, так як $(-f'(x^{(k)}))$ – напрямок спадання. Перше α, для якого умова (3.21) виявиться виконаною, приймається за α_k.

3. **Градієнтний метод з адаптивним вибором кроку** α_k

Виберемо α_k з умови:

$$f\left(x^{(k)} - \alpha\, f'\left(x^{(k)}\right)\right) - f\left(x^{(k)}\right) \le -\varepsilon\, \alpha_k \left\| f'\left(x^{(k)}\right) \right\|^2, \qquad (3.22)$$

де $0 < \varepsilon < 1$ – довільна стала (однакова на всіх ітераціях), яка називається параметром методу. Якщо ε дуже мале, то метод може збігатися повільно, а якщо ε дуже велике, то ускладнюється вибір методу α_k з умови (3.22). Наведемо алгоритм вибору α_k на k-й ітерації:

1) вибираємо деяке довільне α (однакове на всіх ітераціях) і визначаємо точку $x = x^{(k)} - \alpha\, f'\left(x^{(k)}\right)$,
2) обчислимо $f(x) = f\left(x^{(k)} - \alpha\, f'\left(x^{(k)}\right)\right)$,
3) перевіримо нерівність (3.22) при $\alpha_k = \alpha$,
4) якщо нерівність (3.22) виконано, то значення α вибираємо за α_k і переходимо до наступної ітерації. У протилежному випадку проводимо дроблення α (помножуючи його на довільне число $0 < \lambda < 1$) поки нерівність (3.22) не буде виконаною. Тоді переходимо до наступної ітерації.

Відзначимо, що з виконання нерівності (3.22) випливає виконання нерівності (3.21). Вимога (3.22) більш сувора, ніж (3.21), але має той же сенс. Цей градієнтний метод найбільш часто використовується на практиці.

4. **Апріорне завдання величини** α_k

Виберемо α_k з умови:

$$\alpha_k > 0,\ k = 0,1,\ldots,\ \sum_{k=0}^{\infty} \alpha_k = \infty,\ \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \qquad (3.23)$$

Наприклад, за α_k можна узяти $\alpha_k = \frac{1}{k+1}$. Вибір α_k за формулою (3.23) є дуже простим для реалізації, але не гарантує виконання умови монотонності $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$, та збігається повільно.

Спосіб вибору α_k за формулою (3.23) застосовується також у субградієнтному методі (методі узагальненого градієнтного спуску), який є аналогом градієнтного методу для мінімізації негладких функцій. У цьому методі послідовність точок $x^{(0)}, x^{(1)}, \ldots$ визначається за правилом

$$x^{(k+1)} = x^{(k)} - \alpha_k g\left(x^{(k)}\right),\ k = 0,1,2,\ldots,\ g\left(x^{(k)}\right) \in \partial f\left(x^{(k)}\right), \qquad (3.24)$$

де субградієнт $g\left(x^{(k)}\right)$ (див. означення розділ 1, § 2) вибирається з множини $\partial f\left(x^{(k)}\right)$ субградієнтів довільним способом.

Виходячи із теоретичних оцінок та обчислювальної практики, субградієнтний метод (3.23), (3.24) збігається порівняно повільно і не є достатньо ефективним способом мінімізації негладких функцій.

Збіжність градієнтних методів

Розглянуті чотири варіанти градієнтних методів за своїми властивостями близькі один до одного. Вони можуть використовуватися для мінімізації функцій, які належать одному класу функцій, при цьому швидкість збіжності (в тих випадках, коли її вдається оцінити) приблизно однакова.

Тому сформулюємо теореми про збіжність тільки для методу найшвидшого спуску, який визначається формулами (3.19), (3.20).

Теорема 1. Нехай

1) функція $f(x)$ диференційовна на E^n;
2) функція $f(x)$ обмежена знизу на E^n, тобто $inf\ f(x) = f_* > -\infty$;
3) її градієнт задовольняє умові Ліпшиця:

$$\left\| f'(\bar{x}) - f'(\bar{z}) \right\| \le L \left\| \bar{x} - \bar{z} \right\|, \text{ для } \forall\, \bar{x}, \bar{z} \in E^n, \qquad (3.25)$$

де L – константа Ліпшиця.

Тоді для послідовності $\{x^{(k)}\}_0^\infty$, яку отримано за формулами (3.19), (3.20), має місце $\left\| f'\left(x^{(k)}\right)\right\| \to 0$, $k \to \infty$ для будь-якої початкової точки $x^{(0)}$.

4) Якщо при цьому множина $M\left(x^{(0)}\right)=\left\{x \in E^n : f(x) \le f\left(x^{(0)}\right)\right\}$ – обмежена, то

$$\lim_{k\to\infty} \rho\left(x^{(k)}, S_*\right)=0,$$

де $S_* = \left\{x \in M\left(x^{(0)}\right): f'(x)=0\right\}$ – множина стаціонарних точок функції $f(x)$ на $M\left(x^{(0)}\right)$, $\rho\left(x^{(k)}, S_*\right)$ – відстань від точки $x^{(k)}$ до множини S_* стаціонарних точок функції $f(x)$.

У випадку виконання умов теореми 1 метод найшвидшого спуску забезпечує тільки збіжність до множини стаціонарних точок S_*, але при таких умовах неможливо оцінити швидкість збіжності і стверджувати збіжність до множини точок мінімуму X_*.

Теорема 2. Нехай виконані умови теореми 1 і, окрім того, функція $f(x)$ опукла на E^n, тоді для послідовності $\left\{x^{(k)}\right\}_0^\infty$, яку отримано за формулами (3.19), мають місце такі твердження:

1) $\lim\limits_{k\to\infty} f\left(x^{(k)}\right)= f_*$, тобто послідовність $\left\{x^{(k)}\right\}_0^\infty$ є мінімізуючою,

2) $\lim\limits_{k\to\infty} \rho\left(x^{(k)}, X_*\right)=0$, послідовність $\left\{x^{(k)}\right\}_0^\infty$ збігається до множини точок мінімуму X_*,

3) справедлива така оцінка за функцією:

$$0 \le f\left(x^{(k)}\right) - f(x_*) \le 4D^2L\frac{1}{k}, \quad k=1,2,\ldots,$$

де $D = \sup\limits_{x,x' \in M(x_0)} \left\| x - x' \right\|$ – діаметр множини $M\left(x^{(0)}\right)$.

У теоремі 2 оцінюється швидкість збіжності за функцією. Оцінка швидкості збіжності за змінною можлива лише при більш жорстких припущеннях відносно властивостей функції $f(x)$.

Теорема 3. Нехай $f(x)$ – двічі неперервно-диференційовна функція, її матриця других частинних похідних $f''(x)$ задовольняє умові:

$$m\|z\|^2 \le \left(f''(x)\,z, z\right) \le M\,\|z\|^2,\ 0<m\le M,\ \forall x, z \in E^n. \qquad (3.26)$$

Тоді для послідовності $\left\{x^{(k)}\right\}_0^\infty$, яку отримано за формулами (3.19), (3.20), збігається до точки мінімуму зі швидкістю геометричної прогресії із знаменником $q=\dfrac{M-m}{M+m}$, тобто при достатньо великих k виконується нерівність

$$\left\|x^{(k+1)}-x_*\right\| \le \frac{M-m}{M+m}\left\|x^{(k)}-x_*\right\|. \qquad (3.27)$$

Функція $f(x)$, яка задовольняє умові (3.26), є **сильно опуклою**.

Геометричний зміст градієнтних методів

1. Геометричний зміст методу найшвидшого спуску.

Твердження:

1) у точці $x^{(k+1)}$, яка визначається умовами (3.19), (3.20), градієнт $f'\left(x^{(k+1)}\right)$ є ортогональним лінії рівня $\Gamma^{(k+1)}$;
2) напрямки спуску $\left(-f'\left(x^{(k)}\right)\right)$ та $\left(-f'\left(x^{(k+1)}\right)\right)$ на двох послідовних ітераціях взаємно ортогональні.

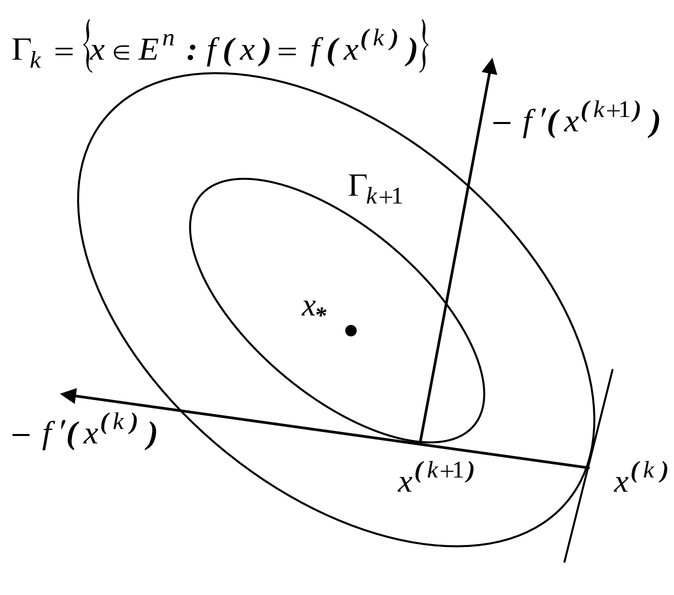


Рис. 3.6

2. Геометричний зміст інших градієнтних методів.

Зауважимо, що перша частина твердження про те, що в точці $x^{(k+1)}$ градієнт $f'\left(x^{(k+1)}\right)$ є ортогональним лінії рівня $\Gamma^{(k+1)}$, має місце для всіх градієнтних методів, а друга – про взаємну ортогональність напрямків спуску – тільки для методу найшвидшого спуску.

Градієнтний метод (відмінний від методу найшвидшого спуску) генерує послідовність точок $\left\{x^{(k)}\right\}_0^\infty$, які утворюють зигзагоподібну траєкторію $x^{(0)}, x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ з ортогональною ланкою в кожній точці $x^{(0)}, x^{(1)}, x^{(2)}, \ldots, x^{(k)}$, відповідній лінії рівня.

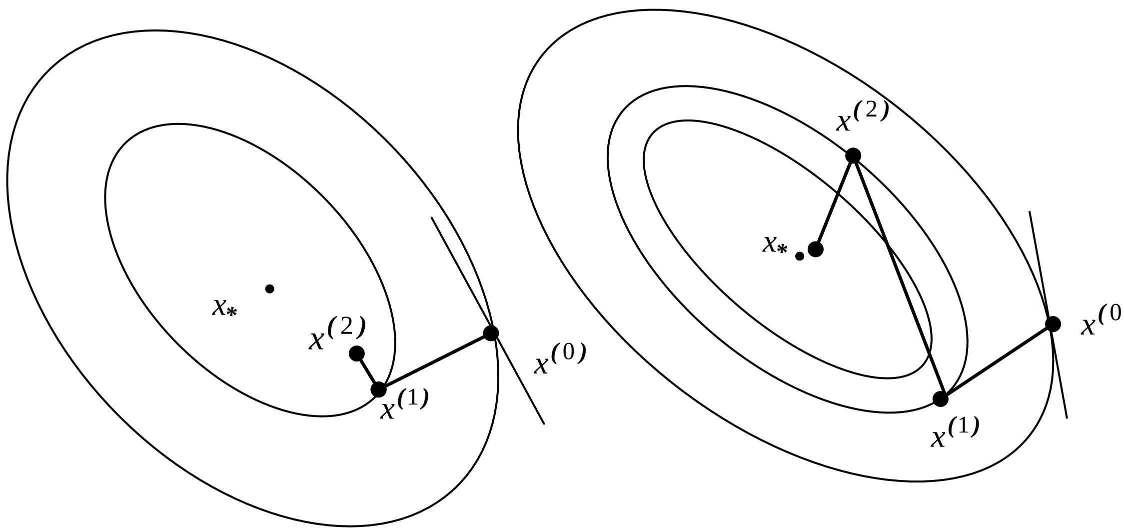


Рис.3.7

Обговорення градієнтних методів

Переваги градієнтних методів такі:

1) достатньо прості в реалізації;
2) вимоги до цільової функції не є дуже жорсткими (функція $f(x)$ повинна бути диференційовна в E^n);
3) є основою для розробки інших більш ефективних методів оптимізації.
4) часто використовуються на початковому етапі розв’язування задачі. Їхнє застосування доцільне у випадку, коли початкове наближення знаходиться далеко від точки мінімуму цільової функції, а кроки вздовж

антиградієнта дозволяють значно зменшити значення цільової функції. Потім застосовуються інші більш ефективні методи оптимізації.

Недоліки градієнтних методів такі:

1) невисока швидкість збіжності (тільки лінійна) при таких жорстких вимогах до цільової функції як двічі неперервно-диференційовність та сильна опуклість (теорема 3);
2) повільна збіжність для функцій, в яких поверхні (лінії рівня) сильно витягнуті та функція має так званий «яристий» характер (рис. 3.8).

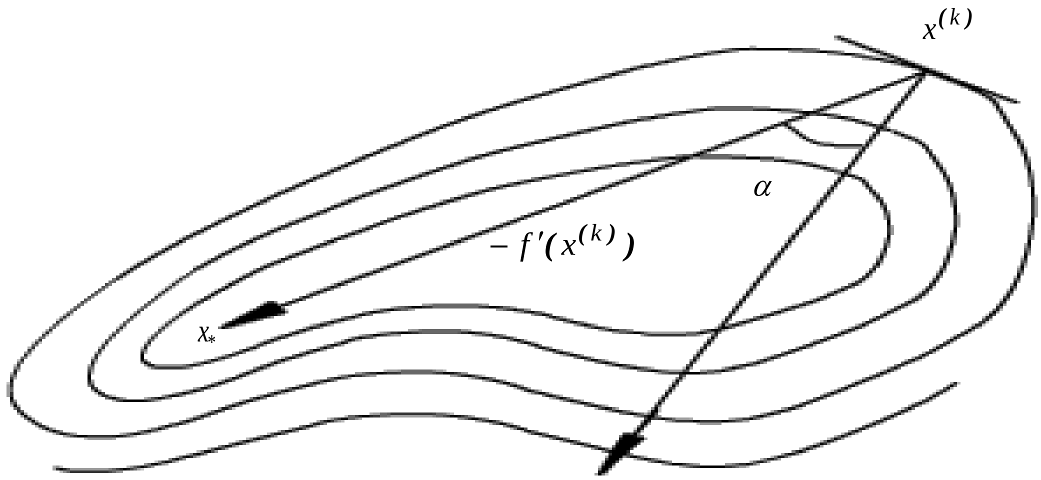


Рис. 3.8

Розглянемо другий недолік градієнтних методів більш докладно. З'ясуємо спочатку, якими характеристиками визначається структура кривизни ліній рівня, а потім подивимось, як залежить швидкість збіжності від кривизни ліній рівня.

Будемо вважати, що $f(x)$ – двічі неперервно-диференційовна і сильно опукла на E^n функція, тобто має місце нерівність (3.26)

$$m\|z\|^2 \le (f''(x)z,z) \le M\|z\|^2,\ 0 < m \le M,\ \forall x,z \in E^n.$$

Відомо, що в (3.26) за m, M можна взяти відповідно мінімальне та максимальне власні числа матриці $f''(x)$.

Матриця $f''(x)$ є добре обумовленою, якщо числа m, M мало відрізняються одне від одного, якщо ж числа m, M суттєво відрізняються одне від одного, тобто $m << M$ (або $\frac{m}{M} << 1$), то матриця $f''(x)$ є погано обумовленою.

Розглянемо залежність кривизни ліній рівня функції $f(x)$ від обумовленості матриці $f''(x)$ на прикладі квадратичної функції

$$f(x) = (A(x - x_*), x - x_*), \tag{3.28}$$

де A – додатно-визначена матриця порядку $n \times n$.

Неважко показати, що для цієї функції $f''(x) = A$. Для $m = M$ нерівність (3.26) буде мати вигляд:

$$(A(x - x_*), x - x_*) = M\|x - x_*\|^2.$$

Тоді рівняння лінії рівня $f(x) = const$ для функції (3.28) запишеться у вигляді:

$$f(x) = M\|x - x_*\|^2 = const,$$

але це і є рівняння кола з центром в точці x_*.

Таким чином, приходимо до такого висновку: для квадратичної функції (3.28) при $m = M$ лінії рівня є концентричними колами, і чим більше будуть відрізнятися одне від одного числа m і M, тим більше будуть витягнуті лінії рівня цільової функції.

Та ж залежність кривизни ліній рівня від чисел m і M залишається і для довільної функції $f(x)$, тобто, чим гірше обумовлена матриця других похідних $f''(x)$, тим сильніше будуть витягнуті лінії рівня функції $f(x)$.

З'ясуємо тепер, як залежить швидкість збіжності від кривизни ліній рівня на прикладі методу найшвидшого спуску (3.19), (3.20). Скористуємось оцінкою швидкості збіжності методу найшвидшого спуску (3.27), яка була отримана в теоремі 3:

$$\|x^{(k+1)} - x_*\| \le q\|x^{(k)} - x_*\|, \quad q = \frac{M - m}{M + m}.$$

Коли числа m і M мало відрізняються одне від одного (матриця $f''(x)$ добре обумовлена), то число q близьке до нуля і, відповідно, збіжність краще.

Якщо ж $\frac{m}{M} << 1$, то q близьке до одиниці і метод починає збігатися повільно. Геометричне тлумачення цього факту таке: зі зменшенням відношення $\frac{m}{M}$ поверхні рівня цільової функції $f(x)$ (тобто поверхні $f(x) = const$) стають більш витягнутими і напрямок спуску (антиградієнт) у більшості точок все більш істотно відхиляється від напрямку на точку мінімуму (див. рис. 3.8). Саме це і призводить до уповільнення швидкості збіжності.

Аналіз недоліків градієнтних методів дозволяє зробити висновок про необхідність застосування для мінімізації функцій з сильно витягнутими лініями рівня (а такі функції часто зустрічаються на практиці) такі методи, в яких вибір напрямку спуску $p^{(k)}$ та точки $x^{(k)}$ базується на інформації про кривизну ліній рівня. Урахування інформації про кривизну ліній рівня дозволить на кожній ітерації зменшити кут між напрямком спуску та напрямком на точку мінімуму (рис. 3.8), що призведе до покращення збіжності методу. Один з таких методів – метод Ньютона. Його буде розглянуто в наступному параграфі.

Явище зиґзаґа та погану збіжність алгоритму найшвидшого спуску на останніх ітераціях можна пояснити, якщо розглянути такий вираз для функції $f(x)$:

$$f\left(x^{(k)} - \alpha\, f'\left(x^{(k)}\right)\right) = f\left(x^{(k)}\right) - \alpha \left(f'\left(x^{(k)}\right), f'\left(x^{(k)}\right)\right) +$$
$$+ \alpha \left\| f'\left(x^{(k)}\right) \right\| \cdot d\left(x^{(k)}, \alpha\, f'\left(x^{(k)}\right)\right),$$

де $d\left(x^{(k)}, \alpha\, f'\left(x^{(k)}\right)\right) \to 0$ при $\alpha\, f'\left(x^{(k)}\right) \to 0$ і $\alpha \geq 0$. Якщо точка $x^{(k)}$ знаходиться близько до стаціонарної точки з нульовим градієнтом і якщо $f(x)$ неперервно-диференційовна, то величина $\left\| f'\left(x^{(k)}\right) \right\|$ є малою і, при цьому доданок $-\alpha \left\| f'\left(x^{(k)}\right) \right\|^2$ стає величиною меншого порядку. Так як метод найшвидшого спуску використовує лінійну апроксимацію функції $f(x)$, то напрямок, який генерується на останніх кроках, не буде достатньо ефективним.

Приклад. Розглянемо ту ж саму задачу: знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2, \; x = (x_1, x_2) \in E^2.$$

Оптимальним розв'язком цієї задачі є точка $x_* = (2.00, 1.00)$, $f(x_*) = 0$. У табл. 3.5 наведені результати обчислень за методом найшвидшого спуску для початкової точки $x^{(0)} = (0.00, 3.00)$. Процедура була зупинена при виконанні умови $\left\| f'\left(x^{(k)}\right) \right\| \le \varepsilon$ ($\varepsilon = 0.09$).

Відзначимо, що метод найшвидшого спуску, як і всі градієнтні методи, достатньо добре працює на початковій стадії процесу мінімізації (див. рис. 3.9), а в околі стаціонарної точки його збіжність уповільнюється.

Таблиця 3.5

k	$x^{(k)}$ $f(x^{(k)})$	$-f'(x^{(k)})$	α_j	$\left\| f'(x^{(k)}) \right\|$	$x^{(k+1)}$
1	(0.00,3.00) 52	(44.00,–24.00)	0.062	50.12	(2.70,1.51)
2	(2.70,1.51) 0.34	(–0.73,–1.28)	0.24	1.47	(2.52,1.20)
3	(2.52,1.20) 0.09	(–0.80,0.48)	0.11	0.93	(2.43,1.25)
4	(2.43,1.25) 0.04	(–0.18,–0.28)	0.31	0.33	(2.37,1.16)
5	(2.37,1.16)0 0.02	(–0.30,0.20)	0.12	0.36	(2.33,1.18)
6	(2.33,1.18) 0.01	(–0.08,–0.12)	0.36	0.14	(2.30,1.14)
7	(2.30,1.14) 0.009	(–0.15,0.08)	0.13	0.17	(2.28,1.15)
8	(2.28,1.15)	(–0.15,–0.08)		0.09	

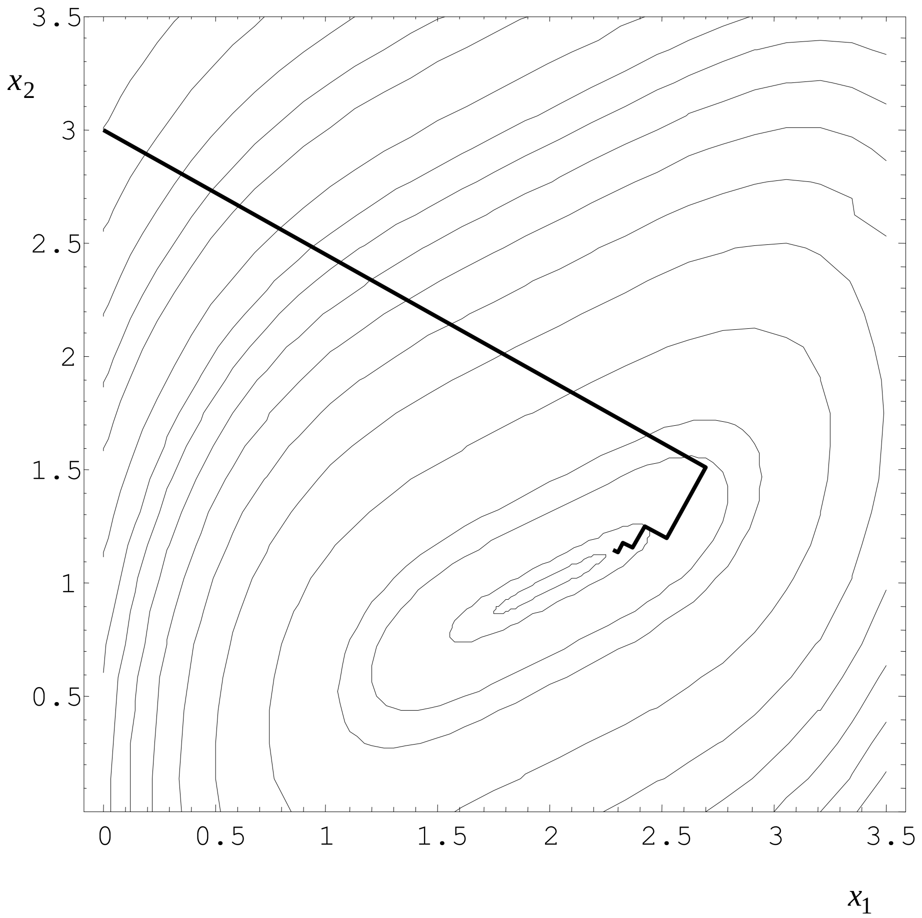


Рис.3.9

§5. Метод Ньютона

Постановка задачі

$$f(x) \to min, \; x \in E^n.$$

Функція $f(x)$ двічі неперервно-диференційовна на E^n.

Метод Ньютона є методом другого порядку, тобто використовує обчислення других похідних функції $f(x)$, яка мінімізується. В його модифікаціях (квазіньютонівських алгоритмах) матриця других похідних апроксимується за допомогою інформації про значення градієнтів функції $f(x)$, і тому ці модифікації є методами першого порядку.

У методі Ньютона послідовність $\left\{x^{(k)}\right\}_0^\infty$ генерується виходячи з квадратичної апроксимації цільової функції.

Згідно з визначенням двічі неперервно-диференційовної функції для чергової точки $x^{(k)}$ маємо

$$f(x)-f\left(x^{(k)}\right)=\left(f'\left(x^{(k)}\right),x-x^{(k)}\right)+$$
$$+\frac{1}{2}\left(f''\left(x^{(k)}\right)\left(x-x^{(k)}\right),\left(x-x^{(k)}\right)\right)+o\left(\left\|x-x^{(k)}\right\|^2\right).$$

Для визначення наступної точки $x^{(k+1)}$ мінімізується функція $f_k(x)$, яка є квадратичною частиною приросту $f(x)-f\left(x^{(k)}\right)$, тобто розв'язується задача

$$f_k(x)=\left(f'\left(x^{(k)}\right),x-x^{(k)}\right)+\frac{1}{2}\left(f''\left(x^{(k)}\right)\left(x-x^{(k)}\right),\left(x-x^{(k)}\right)\right)\to min, x\in E^n. \quad (3.29)$$

Виходячи із диференціальних умов оптимальності (див. розділ I, § 3) маємо, що в точці мінімуму функції $f_k(x)$ її градієнт $f_k'(x)$ перетворюється на нуль, тобто

$$f_k'(x)=f'\left(x^{(k)}\right)+f''\left(x^{(k)}\right)\left(x-x^{(k)}\right)=0. \quad (3.30)$$

Якщо матриця Гессе $f''\left(x^{(k)}\right)$ системи лінійних алгебраїчних рівнянь (3.30) невироджена , то отримуємо

$$x^{(k+1)}=x^{(k)}-\left[f''\left(x^{(k)}\right)\right]^{-1}f'\left(x^{(k)}\right), k=0,1,2,\ldots \quad (3.31)$$

Якщо, наприклад, вихідна функція $f(x)$, яка мінімізується, є опуклою на E^n, то з того, що $f_k''(x)=f''\left(x^{(k)}\right)$ витікає і опуклість функції $f_k(x)$. Дійсно, необхідною і достатньою умовами опуклості функції є невід'ємна визначеність матриці її других похідних (див. розділ I, § 2), а функція $f(x)$ опукла згідно умови, тоді і функція $f_k(x)$ є опуклою. При цих умовах точка $x^{(k+1)}$ з (3.31) є точкою мінімуму квадратичної апроксимації $f_k(x)$. Точку $x^{(k+1)}$, яка отримана за формулою (3.31), приймаємо за $(k+1)$-е наближення методу Ньютона. Коли ж функція $f(x)$ не є опуклою, то точка $x^{(k+1)}$, яка отримана за формулою (3.31), є стаціонарною і її необхідно додатково дослідити на оптимальність.

Неважко побачити, що метод Ньютона мінімізації функції $f(x)$, який визначається співвідношенням (3.31), співпадає з відомим з курсу методів обчислень методом Ньютона розв'язання системи нелінійних рівнянь $f'(x)=0$.

Збіжність методу Ньютона

Теорема. Нехай

1. Функція $f(x)$ є двічі неперервно-диференційовна на E^n.
2. Має місце співвідношення

$$m\,\|z\|^2 \le \left(f''(x)z, z\right) \le M\,\|z\|^2,\ 0 < m \le M,\ \forall x, z \in E^n,$$

(тобто $f(x)$ є сильно опуклою на E^n).

3. Матриця других похідних задовольняє умові Ліпшиця

$\|f''(x) - f''(z)\| \le L\,\|x - z\|$, для $\forall\, x, z \in E^n$, $\quad L > 0$.

4. Початкове наближення $x^{(0)}$ таке, що

$$\left\|f'\!\left(x^{(0)}\right)\right\| \le \frac{8M^2}{L} \text{ або } \left\|f'\!\left(x^{(0)}\right)\right\| = \frac{8M^2 q}{L}, \text{ де } q \in (0,1). \tag{3.32}$$

Тоді послідовність (3.31) збігається до точки мінімуму з квадратичною швидкістю і має місце така оцінка:

$$\left\|x^{(k)} - x_*\right\| \le \frac{4M}{L} q^{2^k}, \quad k = 0,1,\ldots. \tag{3.33}$$

Обговорення методу Ньютона

Збіжність методу Ньютона доведено лише для достатньо близького до точки мінімуму початкового наближення $x^{(0)}$. При розташованих далеко від точки мінімуму початкових наближень метод може розбігатися. При цьому умову (3.2), яка гарантує збіжність методу для даного початкового наближення, на практиці важко перевірити у силу того, що константи M і L, як правило, невідомі.

Залежність збіжності методу від початкового наближення є недоліком методу Ньютона. Ще одним недоліком є висока трудомісткість, яка обумовлена необхідністю обчислення і обертання на кожному кроці матриці Гессе.

У зв'язку з названими вище обставинами застосування класичного методу Ньютона далеко не завжди призводить до успіху. Модифікації його направлені на те, щоб зберігаючи основну перевагу методу Ньютона – швидку збіжність – зменшити трудомісткість та ослабити вимоги до вибору початкового наближення.

Модифікація методу Ньютона

(метод Ньютона з регулюванням кроку)

Розглянемо метод

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[f''\left(x^{(k)}\right)\right]^{-1} f'\left(x^{(k)}\right), \alpha_k > 0, k = 0,1,2,\ldots, \qquad (3.34)$$

який називається методом Ньютона з регулюванням кроку. При $\alpha_k = 1$ він співпадає з класичним методом Ньютона.

Вибір коефіцієнтів α_k в (3.34) здійснюється аналогічно градієнтним методам. Наприклад, адаптивний спосіб визначення α_k. Виберемо α_k з умови:

$$f\left(x^{(k)} + \alpha_k\, p^{(k)}\right) - f\left(x^{(k)}\right) \le \varepsilon\, \alpha_k \left(f'\left(x^{(k)}\right), p^{(k)}\right), \qquad (3.35)$$

де $p^{(k)} = -\left[f''\left(x^{(k)}\right)\right]^{-1} f'\left(x^{(k)}\right)$, $0 < \varepsilon < 1$ – довільна стала (однакова на всіх ітераціях), яка називається параметром методу.

Знаходження α_k з умови мінімізації цільової функції уздовж заданого напрямку $p^{(k)}$:

$$f\left(x^{(k)} - \alpha_k \left[f''\left(x^{(k)}\right)\right]^{-1} f'\left(x^{(k)}\right)\right) = \min_{\alpha \ge 0} f\left(x^{(k)} - \alpha \left[f''\left(x^{(k)}\right)\right]^{-1} f'\left(x^{(k)}\right)\right). \qquad (3.36)$$

Обговорення модифікованого методу Ньютона

Можна показати, що вказані варіанти методу (3.34) збігаються для будь якого початкового наближення $x^{(0)} \in E^n$. При цьому швидкість збіжності буде або надлінійною, або квадратичною залежно від вимог, яким буде задовольняти цільова функція $f(x)$.

Зменшити трудомісткість методу можна, якщо обчислювати матрицю Гессе не на кожному кроці, як це робиться в (3.34), а один раз через кожні s кроків, тобто

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[f''\left(x^{\left(s\left[\frac{k}{s}\right] \right)} \right) \right]^{-1} f'\left(x^{(k)}\right), \alpha_k > 0, k = 0,1,2,..., \qquad (3.37)$$

де $\left[\frac{k}{s}\right]$ – ціла частина числа $\frac{k}{s}$, тобто найбільше ціле число, яке менше або дорівнює $\frac{k}{s}$.

Підбираючи емпіричним шляхом s, іноді вдається отримати за допомогою цього методу непогані результати. Однак, істотного зменшення трудомісткості методу Ньютона цей прийом не дає.

Більш перспективним у цьому плані є інший підхід, при якому будується апроксимація матриці $\left[f''\left(x^{(k)}\right)\right]^{-1}$ на основі інформації про значення градієнтів $f'\left(x^{(k)}\right)$, $f'\left(x^{(k-1)}\right)$,... Такі методи називаються квазіньютоновськими.

Приклад. Розглянемо задачу мінімізації функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2, \ x = (x_1, x_2) \in E^2.$$

Оптимальним розв'язком цієї задачі є точка $x_* = (2.00, 1.00)$, $f(x_*) = 0$. Результати обчислень за методом Ньютона для початкової точки $x^{(0)} = (0.00, 3.00)$ наведені в таблиці 3.6.

Таблиця 3.6

k	$x^{(k)}$ $f(x^{(k)})$	$f'(x^{(k)})$	$f''(x^{(k)})$	$p^{(k)}$	$x^{(k+1)}$
0	(0.00,3.00) 52.0	$\begin{bmatrix} -44.0 \\ 24.0 \end{bmatrix}$	$\begin{bmatrix} 50.0 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\begin{bmatrix} 0.67 \\ -2.67 \end{bmatrix}$	(0.67,0.33)
1	(0.67,0.33) 3.13	$\begin{bmatrix} -9.4 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 23.2 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\begin{bmatrix} 0.44 \\ 0.23 \end{bmatrix}$	(1.11,0.56)
2	(1.11,0.56) 0.63	$\begin{bmatrix} -2.84 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 11.5 & -4.0 \\ -4.0 & 80.0 \end{bmatrix}$	$\begin{bmatrix} 0.30 \\ 0.14 \end{bmatrix}$	(1.41,0.75)

k	$x^{(k)}$ $f(x^{(k)})$	$f'(x^{(k)})$	$f''(x^{(k)})$	$p^{(k)}$	$x^{(k+1)}$
3	(1.41,0.75) 0.12	$\begin{bmatrix} -0.80 \\ 0.14 \end{bmatrix}$	$\begin{bmatrix} 6.2 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\begin{bmatrix} 0.20 \\ 0.10 \end{bmatrix}$	(1.61,0.80)
4	(1.61,0.80) 0.02	$\begin{bmatrix} -0.22 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 3.83 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\begin{bmatrix} 0.13 \\ 0.07 \end{bmatrix}$	(1.74,0.87)
5	(1.74,0.87) 0.005	$\begin{bmatrix} -0.07 \\ -0.0 \end{bmatrix}$	$\begin{bmatrix} 2.81 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\begin{bmatrix} 0.13 \\ 0.07 \end{bmatrix}$	(1.83,0.91)
6	(1.83,0.91) 0.0009	$\begin{bmatrix} 0.0003 \\ -0.04 \end{bmatrix}$			

Точки, отримані за допомогою цього методу показані на рис. 3.10.

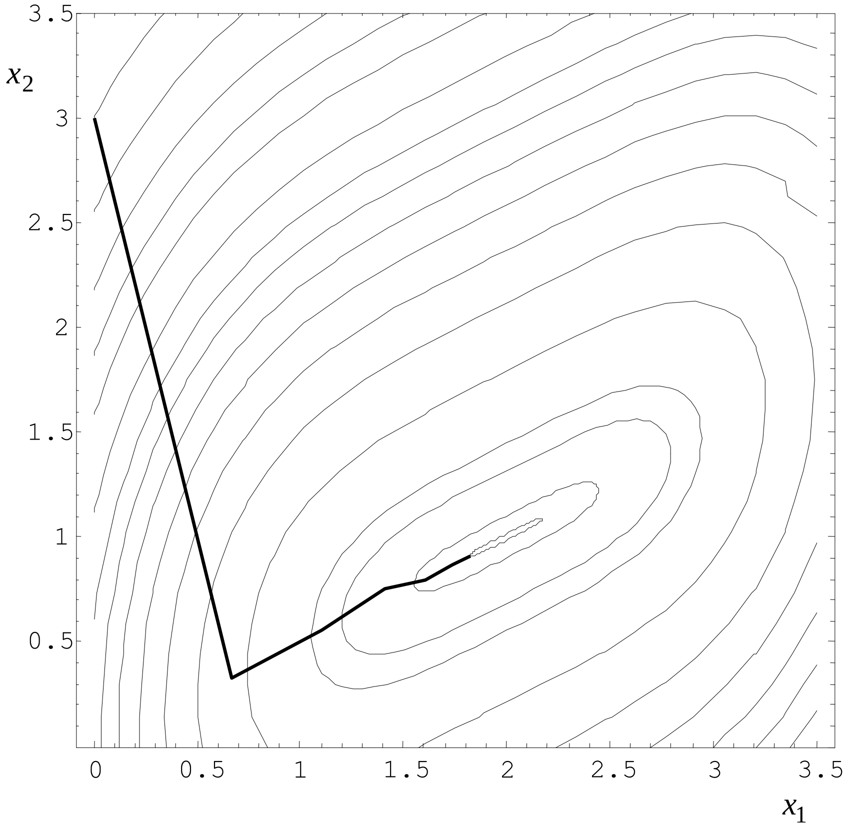


Рис.3.10

§6. Методи спряжених напрямків

1. A – спряжені вектори.

Нехай A – симетрична матриця, яка є додатно визначеною.

Два вектори x та y називають A – **спряженими** (спряженими відносно матриці A або A – ортогональними відносно один одного), якщо

$$x^T A\, y = 0.$$

Спряженість двох векторів є узагальненням поняття ортогональності, оскільки при $A = E$, де E – одинична матриця, вектори x та y ортогональні.

Якщо вектор x є спряженим з будь-яким вектором півпростору $M \subset E^n$, тобто якщо $x^T A\, z = 0$ для усіх $z \in M$, то вектор x називають спряженим з півпростором M.

Відзначимо, що спряжені напрямки визначаються неоднозначно.

Приклад. Розглянемо таку задачу:

$$f(x_1, x_2) = -12x_2 + 4x_1^2 + 4x_2^2 - 4x_1 x_2 \to min.$$

Матрицею Гессе заданої функції $f(x_1, x_2)$ є така матриця

$$A = f''(x) = \begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix}, \quad x = (x_1, x_2).$$

Побудуємо два A – спряжені напрямки. Припустимо, що $p_1^T = (1, 0)$. Тоді вектор $p_2^T = (a, b)$ повинен задовольняти рівності

$$p_1^T A\, p_2 = 0.$$

Отже, маємо $8a - 4b = 0$. Наприклад, можна вибрати $a = 1$, $b = 2$, так що $p_2^T = (1, 2)$.

2. Методи спряжених напрямків. Мінімізація квадратичних функцій

Розглянемо квадратичну функцію

$$f(x) = \frac{1}{2}(Ax, x) - (b, x),$$

де A – симетрична матриця, яка є додатно визначеною, b – вектор простору E^n.

Метод спряжених напрямків генерує послідовність таких напрямків $p^{(0)},...,p^{(n-1)}$, що

$$p^{(k)} A\, p^{(j)} = 0 \text{ для усіх } k,\, j, \quad k \neq j\,. \tag{3.38}$$

Система (ненульових) векторів $p^{(0)},...,p^{(n-1)}$, яка задовольняє умовам (3.38), є лінійно незалежною (як ортогональна система в метриці, яка задана невиродженою матрицею).

Ітераційна формула методу

$$x^{(k+1)} = x^{(k)} + \alpha_k\, p^{(k)}, \quad k = 0,1,2,... \tag{3.39}$$

Нехай крок α_k визначається з умови $\min\limits_{\alpha \geq 0} f\left(x^{(k)} + \alpha\; p^{(k)}\right)$. Тоді

$$\alpha_k = -\frac{\left(f'\left(x^{(k)}\right), p^{(k)}\right)}{\left(A\, p^{(k)}, p^{(k)}\right)}. \tag{3.40}$$

За допомогою спряжених напрямків задача мінімізації квадратичної функції може бути розв’язана за скінчену кількість кроків, яка не перебільшує n (n – розмір простору). Наведена властивість є важливішою з властивостей спряжених напрямків. Вона вказує на ефективність спряжених векторів для мінімізації квадратичної функції і є причиною широкого розповсюдження методів спряжених напрямків.

Розглянемо спосіб побудування A – ортогональних векторів. Кожний з таких способів буде визначати той або інший метод спряжених напрямків. Ефективність методів спряжених напрямків буде безпосередньо залежати від кількості обчислень, які потрібно виконати для побудови системи спряжених векторів. Сформулюємо загальні принципи, котрим повинен задовольняти будь-який з способів побудови спряжених векторів для того, щоб відповідний метод спряжених напрямків був достатньо ефективним.

По-перше, у процесі побудування спряжених векторів можуть бути використані лише обчислення функції та її градієнта і не повинні обчислюватися другі похідні функції.

По-друге, для побудування спряжених векторів використовується інформація про функцію лише в точках послідовності $x^{(k)}$. Для визначення вектора $p^{(i)}, 0 \le i \le n-1$ повинна використовуватися лише інформація про значення функції та її градієнта в точках $x^{(0)},\ldots,x^{(i)}$. З цієї вимоги витікає, що можна розглядати лише ті методи побудування спряжених векторів, для котрих умова

$$\left(f'\left(x^{(i)}\right), p^{(i)}\right) = 0, \quad 0 \le i \le n-1$$

виконується тільки тоді, коли $f'\left(x^{(i)}\right) = 0$. Інакше $\alpha_i = 0$ (тобто $x^{(i+1)} = x^{(i)}$) і процес буде вироджений, якщо $f'\left(x^{(i)}\right) \ne 0$.

Будемо позначати

$$r^{(i)} = x^{(i+1)} - x^{(i)} = \alpha_i p^{(i)},\; e^{(i)} = f'\left(x^{(i+1)}\right) - f'\left(x^{(i)}\right) = \alpha_i A p^{(i)}. \qquad (3.41)$$

В якості вектора $p^{(0)}$ можна вибрати довільний напрямок спуску цільової функції, тобто покласти $p^{(0)} = -H_0^T f'\left(x^{(0)}\right)$, де H_0 – симетрична, додатно визначена матриця. Визначимо вимоги до вектора $p^{(k)}, 1 \le k \le n-1$, для того, щоб виконувались умови A – ортогональності:

$$\left(p^{(k)}, A\, p^{(j)}\right) = 0, \quad 0 \le j \le k-1. \qquad (3.42)$$

Завдяки властивостям спряжених напрямків при виборі α_i в (3.39) за формулою (3.40) одночасно з виконанням умов (3.42) повинні виконуватись також рівності

$$\left(f'\left(x^{(k)}\right), p^{(j)}\right) = 0, \quad 0 \le j \le k-1. \qquad (3.43)$$

Якщо покласти

$$p^{(k)} = -H_k^T f'\left(x^{(k)}\right), \qquad (3.44)$$

де H_k – деяка квадратна матриця $n \times n$, то умову (3.42) можна переписати у вигляді

$$\left(f'\left(x^{(k)}\right), H_k A p^{(j)}\right) = 0, \quad 0 \le j \le k-1.$$

Таким чином умови (3.42), (3.44) будуть виконуватися, якщо

$$H_k A p^{(j)} = a\, p^{(j)}, \quad 0 \le j \le k-1,$$

де a – довільна стала.

З урахуванням (3.41) умови для визначення матриці H_k можна записати так:

$$H_k e^{(j)} = a\, r^{(j)}, \quad 0 \le j \le k-1 \tag{3.45}$$

При $k < n-1$ число векторних рівнянь (3.45) буде менше ніж n, звідки витікає, що матриця H_k визначається неєдиним чином. Окрім того, при різних значеннях константи a системи рівнянь для визначення матриці H_k також буде різними.

Представимо (3.45) у вигляді

$$\left(H_{k-1} + \Delta H_{k-1}\right) e^{(j)} = a\, r^{(j)}, \quad 0 \le j \le k-1. \tag{3.46}$$

Так як матриця H_{k-1} повинна задовольняти рівнянням

$$H_{k-1} e^{(j)} = a\, r^{(j)}, \quad 0 \le j \le k-2,$$

то з (3.46) витікає, що

$$\Delta H_{k-1}\, e^{(j)} = 0, \quad 0 \le j \le k-2, \; \Delta H_{k-1}\, e^{(k-1)} = a\, r^{(k-1)} - H_{k-1}\, e^{(k-1)}.$$

Остання рівність буде виконуватись, якщо

$$\Delta H_{k-1} = a\, \frac{r^{(k-1)}\left[u^{(k-1)}\right]^T}{\left(u^{(k-1)}, e^{(k-1)}\right)} - \frac{H_{k-1}\, e^{(k-1)}\left[u^{(k-1)}\right]^T}{\left(v^{(k-1)}, e^{(k-1)}\right)}, \tag{3.47}$$

де $u^{(k-1)}$, $v^{(k-1)}$ – невідомі вектори, які повинні задовольняти умовам:

$$\left(u^{(k-1)}, e^{(j)}\right) = 0, \; \left(v^{(k-1)}, e^{(j)}\right) = 0, \; 0 \le j \le k-2,$$

$$\left(u^{(k-1)}, e^{(k-1)}\right) \ne 0, \; \left(v^{(k-1)}, e^{(k-1)}\right) \ne 0.$$

Таким чином, вибираючи різні вектори $u^{(k)}, v^{(k)}$ та константу a, отримаємо різні алгоритми побудови спряжених векторів, тобто побудуємо різні методи спряжених напрямків. Виникає питання: чи будуть напрямки $p^{(j)}, \quad 0 \le j \le n-1$, які визначаються різними матрицями H_j, відрізнятися один від одного? Інакше, чи будуть точки $x^{(1)},...,x^{(n-1)}$ різними для різних алгоритмів (при умові, що точка $x^{(0)}$ та ж сама в кожному методі)?

Зауваження. Точки $x^{(0)}, x^{(1)},...,x^{(n-1)}$ не залежать від вибору векторів $u^{(k)}, v^{(k)}$, тобто способу побудови матриці H_k. Таким чином, послідовні наближення до розв'язку задачі мінімізації квадратичної функції будуть одні й ті ж самі для різних методів спряжених напрямків.

Конкретні алгоритми

1). Покладемо в (3.47): $a=1, u^{(k-1)}=r^{(k-1)}$, $v^{(k-1)}=H_{k-1}^T e^{(k-1)}$. Тоді

$$H_k = H_{k-1} + \frac{r^{(k-1)}\left[r^{(k-1)}\right]^T}{\left(r^{(k-1)}, e^{(k-1)}\right)} - \frac{H_{k-1} e^{(k-1)}\left[e^{(k-1)}\right]^T H_{k-1}}{\left(H_{k-1}^T e^{(k-1)}, e^{(k-1)}\right)}. \tag{3.48}$$

Матриця H_k є симетричною, додатно-визначеною. Формула для визначення напрямку $p^{(k)}$ має вигляд

$$p^{(k)} = -\gamma_k H_0 f'(x^{(k)}) + (1-\gamma_k) p^{(k-1)}, \tag{3.49}$$

де

$$\gamma_k = 1 - \frac{\left(H_0 f'(x^{(k)}),\ e^{(k-1)}\right)}{\left(H_0 f'(x^{(k)}), f'(x^{(k)})\right) - \left(p^{(k-1)}, f'(x^{(k)})\right)},$$

$$\gamma_k = -\frac{\left(p^{(k-1)}, f'(x^{(k)})\right)}{\left(H_0 f'(x^{(k)}), f'(x^{(k)})\right) - \left(p^{(k-1)}, f'(x^{(k)})\right)}.$$

Напрямок $p^{(k)}$ також можна визначити так:

$$p^{(k)} = -H_0 f'\left(x^{(k)}\right) + \tilde{\beta}_k \left(H_0 f'\left(x^{(k)}\right) + p^{(k-1)}\right), \tag{3.50}$$

де

$$\tilde{\beta}_k = 1 - \gamma_k = \frac{\left(H_0 f'\left(x^{(k)}\right), f'\left(x^{(k)}\right)\right)}{\left(H_0 f'\left(x^{(k)}\right), f'\left(x^{(k)}\right)\right) - \left(p^{(k-1)}, f'\left(x^{(k)}\right)\right)}.$$

Обчислюючи у виразах (3.49), (3.50) коефіцієнти $\tilde{\beta}_k, \gamma_k$ за різними формулами, ми отримаємо різні методи спряжених напрямків. Підкреслимо, що при мінімізації неквадратичних функцій різні формули для визначення $p^{(k)}$ будуть визначати різні (як за величиною, так і за напрямком) вектори.

Приклад. Побудуємо спряжені напрямки для функції $f(x)$, використовуючи формули (3.49):

$$f(x_1,x_2)=3x_1^2-2x_1x_2+x_2^2-6x_1-2x_2.$$

Нехай $x^{(0)}=(0,0)$, $H_0=\begin{pmatrix}1&0\\0&1\end{pmatrix}$. Початковий напрямок спуску

$$p^{(0)}=-H_0^T f'\left(x^{(0)}\right)=-\begin{pmatrix}1&0\\0&1\end{pmatrix}\cdot\begin{pmatrix}-6\\-2\end{pmatrix}=\begin{pmatrix}6\\2\end{pmatrix}.$$

Величина кроку α_0 визначається за формулою (3.40):

$$\alpha_0=-\frac{\left(f'\left(x^{(0)}\right),p^{(0)}\right)}{\left(Ap^{(0)},p^{(0)}\right)}=\frac{5}{22},\quad A=\begin{pmatrix}6&-2\\-2&2\end{pmatrix}.$$

Знайдемо $x^{(1)}=x^{(0)}+\alpha_0 p^{(0)}=\left(\frac{15}{11},\frac{5}{11}\right)^T$. Згідно з (3.49) $\gamma_1=\frac{121}{170}$,

$f'\left(x^{(1)}\right)=\begin{pmatrix}\frac{14}{17}\\\frac{56}{17}\end{pmatrix}$. Враховуючи, що $\alpha_1=\frac{17}{22}$, отримаємо $x^{(2)}=(2,3)^T$, $x^{(2)}=x_*$.

Таким чином, для функції двох змінних за два кроки знайдена точка мінімуму. На рис. 3.11 показані напрямки $p^{(0)}$ і $p^{(1)}$, а також для порівняння пунктирною лінією зображено напрямок $f'\left(x^{(1)}\right)$ та точка $\bar{x}^{(2)}$, яка обчислюється за методом найшвидшого спуску.

2). Покладемо в (3.47): $a=1, u^{(k-1)}=r^{(k-1)}=v^{(k-1)}$. Тоді

$$H_k=H_{k-1}+(r^{(k-1)}-H_{k-1}e^{(k-1)})\frac{\left[r^{(k-1)}\right]^T}{\left(r^{(k-1)},e^{(k-1)}\right)} \tag{3.51}$$

Для цього методу матриця H_k не є симетричною.

3. Виберемо: $a=0$ $v^{(k-1)}=r^{(k-1)}$. Тоді

$$H_k=H_{k-1}-\frac{H_{k-1}e^{(k-1)}\left[r^{(k-1)}\right]^T}{\left(r^{(k-1)},e^{(k-1)}\right)}. \quad (3.52)$$

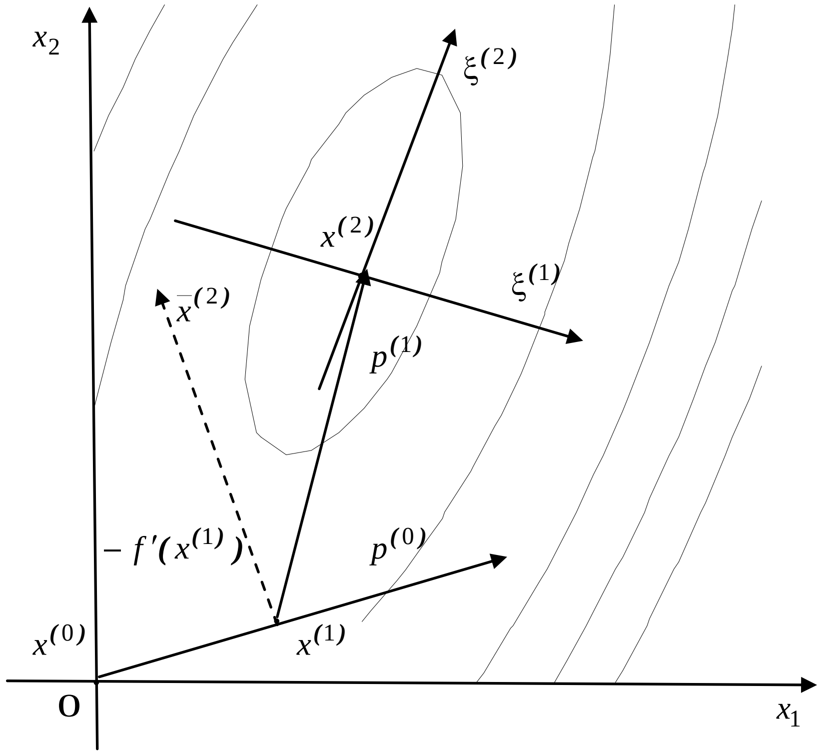


Рис. 3.11

Для другого та третього методів напрямок $p^{(k)}$ обчислюється за формулами

$$p^{(k)} = -H_0 f'\left(x^{(k)}\right) + \beta_k\, p^{(k-1)}, \qquad (3.53)$$

де

$$\beta_k = \frac{\left(H_0\, f'\left(x^{(k)}\right), e^{(k-1)}\right)}{\left(p^{(k-1)}, e^{(k-1)}\right)}.$$

3. Методи спряжених напрямків. Мінімізація довільних функцій

Нехай цільова функція $f(x) \in C^1\left(E^n\right)$, але не є квадратичною. У цьому випадку при побудові векторів $p^{(0)}, p^{(1)}, \ldots, p^{(k)}$ будь-яким методом, який розглянуто в другому пункті, ці вектори не будуть вже спряженими. Тільки коли початкова точка $x^{(0)}$ вибрана в достатньо малому околі точки мінімуму x_* і функція $f(x)$ є гладкою та опуклою , то в будь-якій точці цього околу матриця $f''(x)$ буде близькою до матриці $f''(x_*)$. Таким чином, у цьому випадку можна сподіватися, що властивості векторів $p^{(0)}, p^{(1)}, \ldots, p^{(k)}$ будуть близькими до властивостей спряжених векторів ($f''(x_*)$-ортогональних) і тому властивості ітераційного процесу

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \quad k = 0,1,\ldots,$$

в якому параметр α_k вибирається з умови мінімуму функції $f(x)$ у напрямку $p^{(k)}$, будуть близькі до властивостей вивчених методів спряжених напрямків. Відзначимо, що для неквадратичної функції побудовані методи не будуть збігатися за скінченну кількість кроків.

Теорема про збіжність методів. Нехай

1) $f(x) \in C^2(E^n)$,

2) $f(x)$ – сильно опукла функція,

3) H_0 – симетрична, додатно визначена матриця.

Тоді, якщо значення α_k вибирається з умови мінімуму функції в напрямку $p^{(k)}$, послідовність $\{x^{(k)}\}$ незалежно від вибору початкової точки $x^{(0)}$ збігається до розв'язку з надлінійною швидкістю.

§7. Методи спряжених градієнтів

1. Метод спряжених градієнтів Флетчера-Рівса для квадратичної функції

Найбільш просту формулу для обчислення A-ортогональних векторів можна отримати, якщо вибрати в (3.53) $H_0 = I$. При цьому

$$p^{(k)} = -f'\left(x^{(k)}\right) + \beta_k p^{(k-1)}, \tag{3.54}$$

де

$$\beta_k = \frac{\left(f'\left(x^{(k)}\right), f'\left(x^{(k-1)}\right) - f'\left(x^{(k)}\right)\right)}{\left\| f'\left(x^{(k-1)}\right) \right\|^2},$$

$$\beta_k = \frac{\left\| f'\left(x^{(k)}\right) \right\|^2}{\left\| f'\left(x^{(k-1)}\right) \right\|^2}, \tag{3.55}$$

$$\beta_k = \frac{\left(f'\left(x^{(k)}\right), A\, p^{(k-1)}\right)}{\left(A\, p^{(k-1)}, p^{(k-1)}\right)}.$$

Метод (3.54), (3.55) відомий як метод спряжених градієнтів. Реалізація кожної ітерації методу супроводжується неминучими похибками заокруглення. Як показує практика, ці похибки накопичуються і це призводить до того, що вектори $\left\{p^{(k)}\right\}$ перестають вказувати напрямок спадання цільової функції, і збіжність методу може порушитися. Для того, щоб боротися з цим явищем метод спряжених напрямків час від часу відновлюють, вибираючи у формулі (3.55) $\beta_k = 0$. Позначимо множину номерів $k \geq 1$, для яких приймається $\beta_k = 0$, через I_2. Номера $k \in I_2$ називаються **моментами оновлення методу**. Якщо метод використовується без оновлення, то $I_2 = \varnothing$. На практиці часто беруть $I_2 = \{n, 2n, 3n, \ldots\}$, де $n-$ розмір простору. Можливі й інші правила вибору моментів оновлення методу.

У випадку, коли $I_2 = \{1,2,3,\ldots\}$ метод спряжених градієнтів співпадає з методом найшвидшого спуску.

Алгоритм методу

Початковий етап. Вибрати число $\varepsilon > 0$ для зупинки алгоритму, початкову точку $x^{(0)}$, покласти $p^{(0)} = -f'\left(x^{(0)}\right)$, $j = k = 0$, $y^{(0)} = x^{(0)}$ і перейти до основного етапу.

Основний етап

Крок 1. Якщо $\left\| f'\left(y^{(k)}\right) \right\| < \varepsilon$, то зупинитися. У протилежному випадку взяти

$$\alpha_k = -\frac{\left(f'\left(y^{(k)}\right), p^{(k)}\right)}{\left(A\, p^{(k)}, p^{(k)}\right)}$$

Покласти $y^{(k+1)} = y^{(k)} + \alpha_k p^{(k)}$. Якщо $k < n$, то перейти до кроку 2, у протилежному випадку перейти до кроку 3.

Крок 2. Покласти $p^{(k+1)} = -f'\left(y^{(k+1)}\right) + \beta_k p^{(k)}$,

$$\text{де } \beta_k = \frac{\left\| f'\left(x^{(k)}\right)\right\|^2}{\left\| f'\left(x^{(k-1)}\right)\right\|^2}$$

замінити k на $k+1$ та перейти до кроку 1.

Крок 3. Покласти $y^{(0)} = x^{(j+1)} = y^{(n)}$, $p^{(0)} = -f'\left(y^{(0)}\right)$, $k = 0$, замінити j на $j+1$ та перейти до кроку 1.

Алгоритм описаний.

Теорема. Метод спряжених градієнтів для квадратичної функції збігається за скінченну кількість кроків, яка не перевищує вимірність простору.

2. Метод спряжених градієнтів Глетчера-Рівса (випадок довільної функції

Нехай $f(x) \in C^1\left(E^n\right)$. Ітераційна формула методу

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \quad k = 0, 1, \ldots,$$

де

$$p^{(k)} = \begin{cases} -f'\left(x^{(0)}\right), & k = 0, \\ -f'\left(x^{(k)}\right) + \beta_k p^{(k-1)}, & k = 1, 2, \ldots, \end{cases} \tag{3.54}$$

кроковий множник α_k знаходиться з умови:

$$g_k(\alpha_k) = \min_{\alpha \geq 0} g_k(\alpha), \quad g_k(\alpha) = f\left(x^{(k)} - \alpha f'\left(x^{(k)}\right)\right),$$

$$\beta_k = \begin{cases} \dfrac{\left(f'\left(x^{(k)}\right), f'\left(x^{(k)}\right) - f'\left(x^{(k-1)}\right)\right)}{\left\| f'\left(x^{(k-1)}\right)\right\|^2}, & k \in I_1, \\ 0, & k \in I_2. \end{cases}$$

У формулах для визначення параметра β_k: I_1 – множина індексів, $I_1 \cup I_2 = \{0,1,2,\ldots\}$, I_2 – множина індексів, яка відповідає моментам оновлення методу.

Приклад. Розглянемо таку задачу: знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x = (x_1, x_2) \in E^2.$$

Оптимальним розв’язком цієї задачі є точка $x_* = (2.00, 1.00)$, $f(x_*) = 0$. Результати обчислень за методом Флетчера-Рівса наведені в табл. 3.7 для початкової точки $x^{(0)} = (0.00, 3.00)$. На четвертій ітерації отримана точка $x^{(4)} = (2.185,\ 1.094)$. На рис. 3.12 показано траєкторію спуску до точки мінімуму.

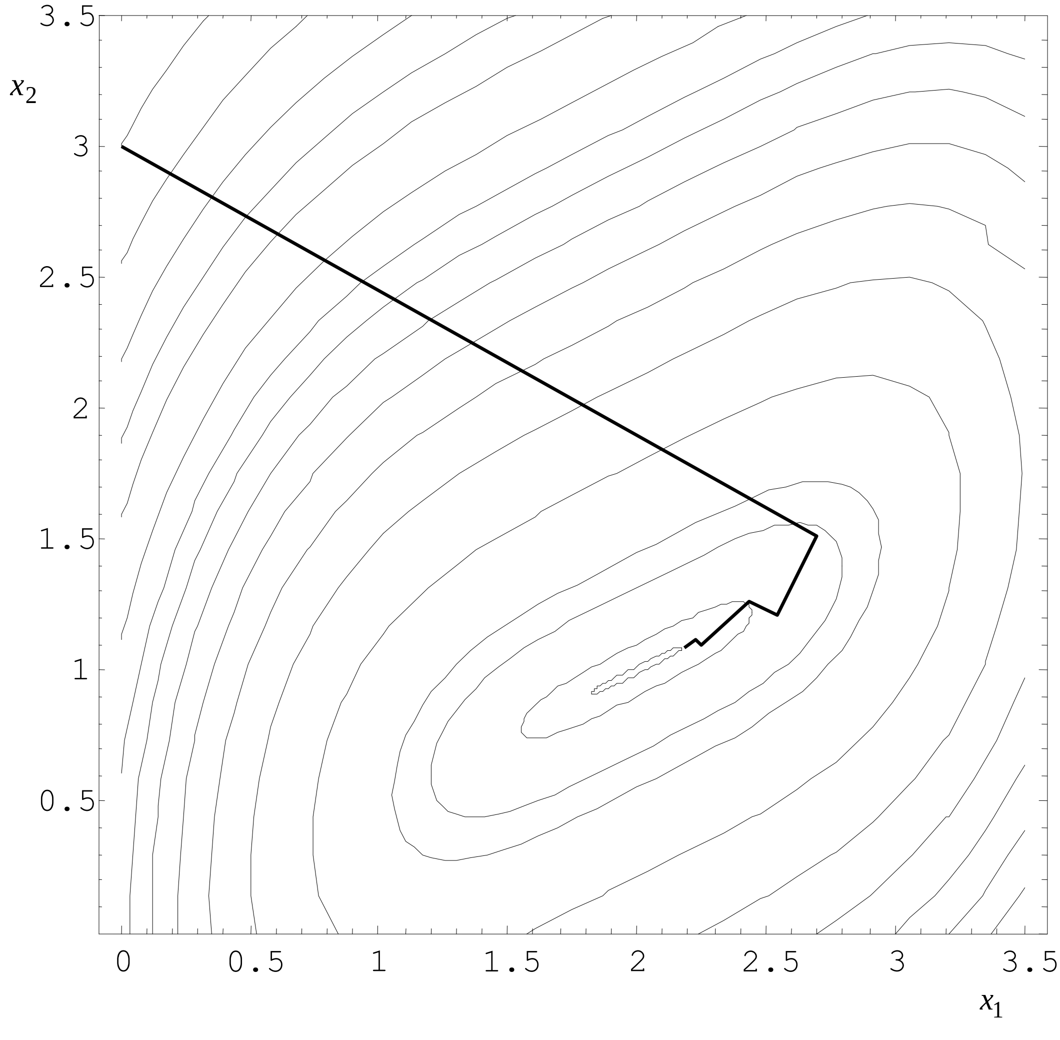


Рис.3.12

Таблиця 3.7

k	$x^{(k)}$ $f(x^{(k)})$	j	$y^{(j)}$ $f(y^{(j)})$	β_j	$p^{(j)}$	α_j	$y^{(j+1)}$
1	(0.00,3.00) 52	1	(0.00,3.00) 52		$\begin{bmatrix} 44.0 \\ -24.0 \end{bmatrix}$	0.062	(2.70,1.51)
		2	(2.7,1.51) 0.34	0.0009	$\begin{bmatrix} -0.69 \\ -1.3 \end{bmatrix}$	0.23	(2.54,1.21)
2	(2.54,1.21) 0.10	1	(2.54,1.21) 0.10		$\begin{bmatrix} -0.87 \\ 0.48 \end{bmatrix}$	0.11	(2.44,1.26)
		2	(2.44,1.26) 0.04	0.14	$\begin{bmatrix} -0.30 \\ -0.25 \end{bmatrix}$	0.63	(2.25,1.10)
3	(2.25,1.10) 0.008	1	(2.25,1.1) 0.008		$\begin{bmatrix} -0.16 \\ -0.20 \end{bmatrix}$	0.10	(2.23,1.12)
		2	(2.23,1.12) 0.003	0.04	$\begin{bmatrix} -0.036 \\ -0.03 \end{bmatrix}$	1.02	(2.19,1.09)
4	(2.19,1.09) 0.0017	1	(2.19,1.09) 0.0017		$\begin{bmatrix} -0.05 \\ 0.04 \end{bmatrix}$	0.11	(2.19,1.09)
		2			$\begin{bmatrix} -0.036 \\ -0.03 \end{bmatrix}$		

§8. Методи змінної метрики (квазіньютонові методи)

Для застосування методу Ньютона необхідно знати матрицю Гессе і виконувати її обернення. Але в багатьох випадках може статися, що матриця Гессе невідома або її обчислення зв'язано з великими труднощами. Окрім того, необхідне число множень при оберненні матриці розміру $n \times n$ пропорційно n^3, що потребує великих обчислювальних затрат при невеликих n. У зв'язку з цим розроблена група методів, в яких матриця обернена до матриці Гессе, будується поступово у процесі спуску.

Нехай функція $f(x)$ є двічі неперервно-диференційовною. Розглянемо метод

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \ k = 0,1,2,\ldots, \quad (3.56)$$

де $p^{(k)} = -H_k f'\left(x^{(k)}\right)$.

Матрицю H_k будемо вибирати таким чином, щоб вона, у певному сенсі, апроксимувала матрицю $\left[f''\left(x^{(k)}\right)\right]^{-1}$. Зауважимо, що

$$f'\left(x^{(k)}\right) - f'\left(x^{(k+1)}\right) = f''\left(x^{(k+1)}\right)\left(x^{(k)} - x^{(k+1)}\right) + o\left(\left\| x^{(k)} - x^{(k+1)} \right\|\right).$$

Припускаючи, що матриця $f''\left(x^{(k+1)}\right)$ невироджена, з точністю до членів більш високого порядку малості порівняно з $\left\| x^{(k)} - x^{(k+1)} \right\|$, маємо

$$\left[f''\left(x^{(k)}\right)\right]^{-1}\left(f'\left(x^{(k)}\right) - f'\left(x^{(k+1)}\right)\right) \approx x^{(k+1)} - x^{(k)}. \quad (3.57)$$

При цьому, якщо $f(x)$ – квадратична функція, то наближена рівність (3.57) стає точною

$$\left[f''\left(x^{(k)}\right)\right]^{-1} \Delta y_k = \Delta x_k,$$

де $\Delta y_k = f'\left(x^{(k+1)}\right) - f'\left(x^{(k)}\right)$, $\Delta x_k = x^{(k+1)} - x^{(k)}$.

Тому природно вимагати, щоб для матриці H_{k+1}, яка наближує $\left[f''\left(x^{(k)}\right)\right]^{-1}$, виконувалась умова

$$H_{k+1} \Delta y_k = \Delta x_k \quad (3.58)$$

Ця умова носить назву **квазіньютонової.** Вона лежить в основі цілого ряду методів апроксимації $\left[f''(x^{(k)})\right]^{-1}$. Відповідні методи мінімізації, для яких на кожному кроці виконується квазіньютонова умова, також називають квазіньютоновими.

Нехай наближення до $\left[f''\left(x^{(k)}\right)\right]^{-1}$ перераховуються від кроку до кроку за формулою

$$H_{k+1} = H_k + \Delta H_k.$$

У залежності від того, чи має матриця ΔH_k ранг 1 чи 2, ми будемо казати про корекцію рангу 1 або 2.

Спробуємо вказати яку-небудь матрицю ΔH_k, яка забезпечує виконання умови (3.58). Для цього перепишемо (3.58) у вигляді

$$\Delta H_k \Delta y_k = \Delta x_k - H_k \Delta y_k . \tag{3.59}$$

Ясно, що цій рівності задовольняє матриця рангу 1, яка визначається за формулою

$$\Delta H_k = \frac{1}{(z_k, \Delta y_k)} \left(x^{(k)} - H_k \Delta y_k \right) z_k^T ,$$

де z_k – довільний вектор такий, що $(z_k, \Delta y_k) \neq 0$. Вибравши $z_k = \Delta x^{(k)} - H_k \Delta y_k$ отримуємо з (3.59) при $\left(\Delta x^{(k)} - H_k \Delta y_k ,\, \Delta y_k \right) \neq 0$, таку ітераційну формулу

$$H_{k+1} = H_k + \frac{\left(\Delta x^{(k)} - H_k \Delta y_k \right)\left(\Delta x^{(k)} - H_k \Delta y_k \right)^T}{\left(\Delta x^{(k)} - H_k \Delta y_k , \Delta y_k \right)} . \tag{3.60}$$

Розглянемо методи, які використовують формули корекції рангу 2.

У методі **Давидона-Флетчера-Пауела (ДФП)** ітераційна формула має вигляд

$$H_{k+1} = H_k + \frac{\Delta x^{(k)} \left[\Delta x^{(k)} \right]^T}{\left(\Delta x^{(k)}, \Delta y_k \right)} - \frac{H_k \left(\Delta y_k , \Delta y_k \right)^T}{\left(H_k \Delta y_k , \Delta y_k \right)} . \tag{3.61}$$

У квадратичному випадку напрямки спуску, знайдені за алгоритмом ДФП, є послідовно спряженими відносно матриці A квадратичної форми, а формула (3.61) співпадає з формулою (3.48) методів спряжених напрямків.

У методі **Пройдена-Флетчера-Гольдфарба-Шанно (БФГШ)** для знаходження матриці H_{k+1} використовується формула

$$H_{k+1} = H_k + \left(1 + - \frac{\left[\Delta y_k \right]^T H_k \Delta y_k}{\left[\Delta x_k \right]^T \Delta y_k} \right) \frac{\Delta x^{(k)} \left[\Delta x^{(k)} \right]^T}{\left[\Delta x^{(k)} \right]^T \Delta y_k} -$$

$$- \frac{\Delta x^{(k)} \left[\Delta y^{(k)} \right]^T H_k + H_k \Delta y^{(k)} \left[\Delta x^{(k)} \right]^T}{\left[\Delta x^{(k)} \right]^T \Delta y_k} . \tag{3.62}$$

Відзначимо, що алгоритм БФГШ є менш чутливим до неточностей у процедурі одновимірного пошуку ніж алгоритм ДФП. Це дозволяє використовувати «економічні» одновимірні методи оптимізації (типу Голдстейна, Вольфа–Пауела, див. розділ 2, § 5), котрі потребують дуже мало оцінок цільової функції на кожній ітерації, але їхнє застосування не погіршує збіжності алгоритму.

За матрицю H_0 в квазіньютонових методах можна вибрати будь-яку додатно визначену симетричну матрицю. На практиці, часто вибирається одинична матриця I. Довжина кроку частіше усього вибирається з умови мінімуму цільової функції уздовж заданого напрямку

$$f\left(x^{(k)} + \alpha_k p^{(k)}\right) = \min_{\alpha \ge 0} f\left(x^{(k)} + \alpha\, p^{(k)}\right). \qquad (3.63)$$

Можливі також інші способи вибору параметра α_k.

Зауважимо, що для квадратичної функції $f(x) = \frac{1}{2}(Ax, x) - (b, x)$, де A – симетрична, додатно визначена матриця, b – вектор простору E^n, всі три методи (3.56), (3.60)–(3.62), (3.63) для будь-якого початкового наближення $x^{(0)} \in E^n$ генерують одну й ту ж послідовність точок $x^{(0)}, x^{(1)}, \ldots, x^{(n)}$, причому $H_n = \left[f''\left(x^{(n)}\right)\right]^{-1} = A^{-1}$, $x^{(n)} = x_* = A^{-1} b$, тобто квазіньютонові методи дозволяють знайти мінімум квадратичної функції за n кроків.

Для неквадратичних функцій це, звичайно, не так. Але можна показати, що при відповідних припущеннях

$$H_k - \left[f''\left(x^{(k)}\right)\right]^{-1} \to 0, \; x^{(k)} \to x_*,$$

причому швидкість збіжності є надлінійною.

Теорема. Якщо $f(x)$ – двічі неперервно-диференційовна, сильно опукла на E^n функція, то при будь-якому початковому наближенні $x^{(0)} \in E^n$ послідовність точок $\left\{x^{(k)}\right\}$, яка визначається формулами (3.56), (3.61), (3.63) збігається до точки мінімуму x_*. Якщо при цьому для всіх x таких що $f(x) \le f\left(x^{(0)}\right)$, справедливі нерівності

$$\left| \frac{\partial^2 f(x)}{\partial x_i \partial x_j} - \frac{\partial^2 f(x_*)}{\partial x_i \partial x_j} \right| \le L \, \| x - x_* \| \text{ для будь–яких } i, j = \overline{1,n},$$

то $x^{(k)}$ збігається до x_* надлінійно.

Квазіньютонові методи називають також методами **змінної метрики**. Ця назва пояснюється тим, що будь-яка симетрична, додатно визначена матриця H_k визначає скалярний добуток $(u,v)_k = (H_k\, u, v)$ і пов'язану з ним метрику. Лінійна частина приросту $f\left(x^{(k)} + \Delta x_k\right) - f\left(x^{(k)}\right)$ має вигляд

$$\left(f'\left(x^{(k)}\right), \Delta x_k\right) = \left(H_k\, H_k^{-1}\, f'\left(x^{(k)}\right), \Delta x_k\right) = \left(H_k^{-1}\, f'\left(x^{(k)}\right), \Delta x_k\right)_k$$

і тому вектор $H_k^{-1}\, f'\left(x^{(k)}\right)$ можна розглядати як градієнт функції $f(x)$ у точці $x^{(k)}$ у просторі з скалярним добутком $(\cdot)_k$. Таким чином, метод (3.56) є узагальненням градієнтного методу на випадок простору зі змінною метрикою.

Переваги та недоліки методів змінної метрики. Квазіньютонові методи є ефективним засобом розв'язання задач безумовної оптимізації. Вони характеризуються високою швидкістю збіжності і, в той же час, при реалізації квазіньютонових алгоритмів не потрібно виконувати такі трудомісткі операції, як обчислення матриці других похідних або обернення матриці. Але при великій вимірності простору необхідність зберігання та перерахування на кожному кроці матриць H_k спричиняє високі вимоги до об'єму пам'яті комп'ютера. Цей недолік не властивий методу спряжених градієнтів.

Наведемо алгоритм одного з указаних методів.

Алгоритм методу Давидона-Флетчера-Пауела.

Початковий етап. Вибрати число $\varepsilon > 0$ для зупинки алгоритму, початкову точку $x^{(1)}$ і початкову симетричну додатно визначену матрицю H_1, покласти $k = j = 1$, $y^{(1)} = x^{(1)}$ і перейти до основного етапу.

Основний етап

Крок 1. Якщо $\left\| f'\left(y^{(j)}\right)\right\| < \varepsilon$, то зупинитися. У протилежному випадку взяти $p^{(j)} = -H_j f'\left(y^{(j)}\right)$ і за α_j вибрати оптимальний розв'язок задачі мінімізації $f(y^{(j)} + \alpha\, p^{(j)})$ при $\alpha_j \ge 0$. Покласти $y^{(j+1)} = y^{(j)} + \alpha_j p^{(j)}$. Якщо $j < n$, то перейти до кроку 2. Якщо $j = n$, то покласти $y^{(1)} = x^{(k+1)} = y^{(n+1)}$, замінити k на $k+1$, покласти $j = 1$ та перейти до кроку 1.

Крок 2. Побудувати матрицю H_{j+1} за формулою

$$H_{j+1} = H_j + \frac{\Delta x^{(j)}\left[\Delta x^{(j)}\right]^T}{\left(\Delta x^{(j)}, \Delta y^{(j)}\right)} - \frac{H_j\left(\Delta y^{(j)}, \left[(\Delta y^{(j)})^T\right]\right)}{(H_j\, \Delta y^{(j)}, \Delta y^{(j)})},$$

замінити j на $j+1$ та перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо задачу мінімізації функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x = (x_1, x_2) \in E^2.$$

Результати обчислень за методом Давидона-Флетчера-Пауела наведені в табл. 3.8 для початкової точки $x^{(0)} = (0.00, 3.00)$. Траєкторія руху, яку отримано за методом Давидона-Флетчера-Пауела показано на рис. 3.12.

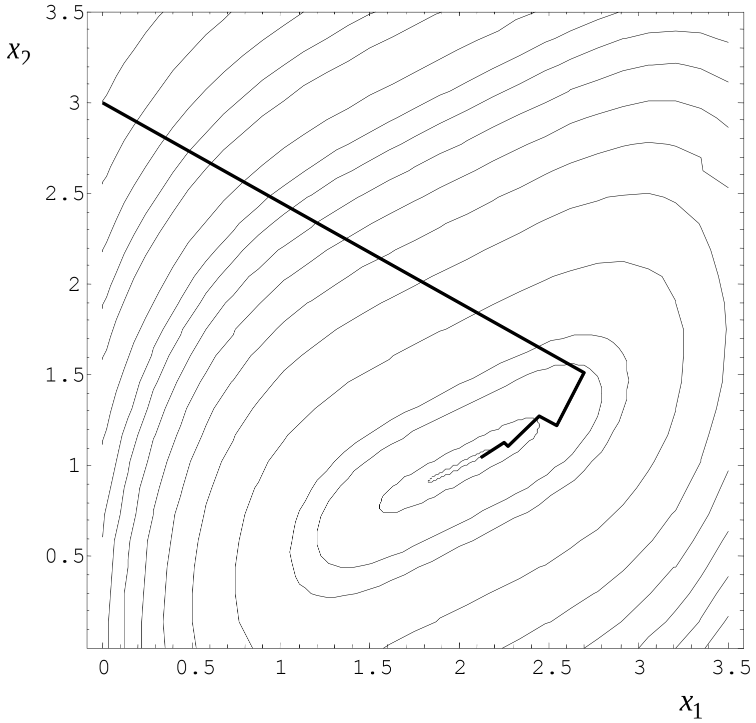


Рис.3.13

Таблиця 3.8

k	$x^{(k)}$ $f(x^{(k)})$	j	$y^{(j)}$ $f(y^{(j)})$	$f'(y^{(j)})$ $\lVert f'(y^{(j)}) \rVert$	H_j $p^{(j)}$	α_j	$y^{(j+1)}$
1	(0.00,3.00) 52	1	(0.00,3.00) 52	$\begin{bmatrix} 44.0 \\ -24.0 \end{bmatrix}$ 50.12	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (44,–24)	0.062	(2.70,1.51)
		2	(2.7,1.51) 0.34	$\begin{bmatrix} 0.73 \\ 1.28 \end{bmatrix}$ 1.47	$\begin{bmatrix} 0.25 & 0.38 \\ 0.38 & 0.25 \end{bmatrix}$ (–0.67,1.31)	0.22	(2.55,1.22)
2	(2.55,1.22) 0.1036	1	(2.55,1.22) 0.1036	$\begin{bmatrix} 0.89 \\ -0.44 \end{bmatrix}$ 0.99	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (–0.89,0.44)	0.11	(2.45,1.27)
		2	(2.45,1.27) 0.05	$\begin{bmatrix} 0.18 \\ 0.36 \end{bmatrix}$ 0.40	$\begin{bmatrix} 0.65 & 0.45 \\ 0.45 & 0.65 \end{bmatrix}$ (–0.28,–0.25)	0.64	(2.27,1.11)
	(2.27,1.11) 0.008	1	(2.27,1.11) 0.008	$\begin{bmatrix} 0.18 \\ -0.20 \end{bmatrix}$ 0.27	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (–0.18,0.20)	0.10	(2.25,1.13)
4	(2.12,1.05) 0.0005	1	(2.12,1.05) 0.0005	$\begin{bmatrix} 0.05 \\ -0.08 \end{bmatrix}$ 0.09	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (–0.05,0.08)	0.10	(2.115,1.06)

§9. Тестові задачі

Кожну обчислювальну програму для розв’язання задачі безумовної оптимізації потрібно протестувати з метою з’ясування чи правильно вона працює і яка її надійність та ефективність.

З’ясуємо питання правильності реалізації алгоритму.

Якщо всі модулі програми працюють правильно, то потрібно протестувати програму для низки нелінійних задач. По-перше, програму тестують на самих простих задачах. Такими задачами можуть бути задачі мінімізації квадратичної функції двох або трьох змінних з додатно визначеною матрицею других похідних. Якщо програма працює правильно, тоді можна протестувати її на деяких стандартних задачах, які зарекомендували себе як добрі тести. Часто є корисним запускати ці тестові задачі не тільки з стандартної початкової точки $x^{(0)}$, а також з точок, які розташовані від неї в 10–100 разів далі і знаходяться на променях, які виходить з розв'язку x_* у напрямку до стандартної точки $x^{(0)}$.

Наведемо ряд тестових функцій, для яких указані вимірність простору n, початкова точка $x^{(0)}$, оптимальний розв'язок x_*:

1. $f_1(x_1,x_2)=(x_1-2)^4+(x_1-2x_2)^2,\ n=2,\ x^{(0)}=(0.00,3.00),\ x_*=(2.0,1.0);$
2. $f_2(x_1,x_2)=\alpha\left(x_2-x_1^2\right)^2+(1-x_1)^2,\quad \alpha=0{,}1;\ 1{,}0;\ 10;\ 100,\quad n=2,$ $x^{(0)}=(-1.2,\,1.0),\ x_*=(1.0,\,1.0);$

   При $\alpha=100$ цю функцію називають функцією Розенброка.
3. $f_3(x_1,x_2)=\left(x_2-x_1^2\right)^2+100(1-x_1)^2,\ n=2,\ x^{(0)}=(-1.2,1.0),\ x_*=(1.0,1.0);$
4. $f_4(x_1,x_2)=100\left(x_2-x_1^3\right)^2+(1-x_1)^2,\ n=2,\ x^{(0)}=(-1.2,1.0),\ x_*=(1.0,1.0);$
5. $f_5(x_1,x_2)=\sum_{i=1}^{3}\left\{c_i-x_1(1-x_2)^i\right\}^2,\quad n=2,\quad c_1=1.5;\ c_2=2.25;\ c_3=2.625;$ $x^{(0)}=(-1.2,1.0),\ x_*=(3.0,0.5);$
6. $f_6(x_1,x_2)=(x_1-2)^4+(x_1-2)^2{x_2}^2+(x_2+1)^2,\quad n=2,\quad x^{(0)}=(1.0,1.0),$ $x_*=(2.0,-1.0);$
7. $f_7(x_1,x_2)={x_1}^4+(x_1+x_2)^2+(exp\,x_2-1)^2,\quad n=2,\quad x^{(0)}=(1.0,1.0),$ $x^{(0)}=(-1.0,3.0),\ x_*=(0.0,0.0);$

При формулюванні наступних тестових задач будемо використовувати цільову функцію вигляду

$$f(x)=\sum_{i=1}^{n}{f_i}^2(x).$$

8. Розширена функція Розенброку. Число n – будь-яке додатне ціле, кратне 2.

$$f_{2i-1}(x)=10\left(x_{2i}-x_{2i-1}{}^{2}\right),$$

$$f_{2i}(x)=1-x_{2i-1},$$

$$i=1,\ldots,n/2,$$

$$x^{(0)}=(-1.2,1.0,\ldots,-1.2,1.0),\ x_{*}=(1,\ 1,\ldots,1\ ).$$

9. Розширена узагальнена функція Пауела. Число n – будь-яке додатне ціле, кратне 4.

$$f_{4i-3}(x)=x_{4i-3}+10x_{4i-2},$$

$$f_{4i-2}(x)=\sqrt{5}(x_{4i-1}-x_{4i}),$$

$$f_{4i-1}(x)=(x_{4i-2}-2x_{4i-1})^{2},$$

$$f_{4i}(x)=\sqrt{10}(x_{4i-3}-x_{4i})^{2},$$

$$i=1,\ldots,n/4,$$

$$x^{(0)}=(3,-1,0,1,\ldots 3,-1,0,1),\ x_{*}=(0,0,\ldots,0).$$

10. Тригонометрична функція. Число n – будь-яке додатне ціле.

$$f_k(x)=n-\sum_{j=1}^{n}\left\{\cos x_j+k(1-\cos x_k)-\sin x_k\right\},$$

$$x^{(0)}=\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right).$$

11. Функція типу спиралевидного жолоба: $n=3$.

$$f_1(x)=10(x_3-10\,\theta(x_1,x_2)),$$

$$f_2(x)=10\left(\sqrt{x_1^2+x_2^2}-1\right),$$

$$f_3(x)=x_3,$$

де

$$\theta(x_1,x_2)=\begin{cases}\dfrac{1}{2\pi}arctg\,\dfrac{x_2}{x_1}, & x_1>0\\ \dfrac{1}{2\pi}arctg\,\dfrac{x_2}{x_1}+0.5, & x_1<0\end{cases}$$

$x^{(0)}=(-1,0,0)$, $x_*=(1,0,0)$.

12. Функція Вуда: $n=4$.

$$f(x)=100\left(x_1^2-x_2\right)^2+(1-x_1)^2+90\left(x_3^2-x_4\right)^2+(1-x_3)^2+$$

$$+10.1\left[(1-x_2)^2+(1-x_4)^2\right]+19.8(1-x_2)(1-x_4),$$

$x^{(0)}=(-3,-1,-3,-1)$, $x_*=(1,1,\ldots,1)$.

13. $f(x)=\sum\limits_{t=0.1}^{1}\left[(exp(-x_1t)-exp(-x_2t))-(exp(-t)-exp(-10t))\right]^2$, параметр t змінюється з кроком 0.1.

14. $f(x)=-x_1^2\,exp\left(1-x_1^2-20.25(x_1-x_2)^2\right),$

$x^{(0)}=(1,1)$, $x_*=(0.8,0.7)$.

Вправи до розділу 3

1. Знайти в явному вигляді кроковий множник α_k в методі найшвидшого спуску для квадратичної функції.

2. Знайти точку мінімуму функції $f(x,y)=\frac{1}{2}\left(x^2-y\right)^2+\frac{1}{2}(1-x)^2$. Виконати одну ітерацію класичного методу Ньютона з початкової точки $x^{(0)}=(2,2)$. Обчислити значення $f\left(x^{(0)}\right)$, $f\left(x^{(1)}\right)$.
3. Сформулювати алгоритм БФГШ.
4. Розглянемо задачу мінімізації функції $f(x,y)=x^2+y^2-xy-x+2y$. Починаючи з початку координат, розв'яжіть цю задачу методом Давидона-Флетчера-Пауела при $H_1=I$. Розв'яжіть цю задачу методом спряжених градієнтів. Переконайтеся, що обидві ці процедури породжують однакову множину напрямків. Покажіть, що в загальному випадку, якщо $H_1=I$, то обидва методи ідентичні для квадратичних функцій.
5. Розв'язати задачу мінімізації функції $f(x,y)=x+2y^2+exp\left(x^2+y^2\right)$, починаючи з точки $x^{(0)}=(0,1)$ і використовуючи метод спряжених градієнтів та метод Розенброка.
6. Розв'язати такі задачі безумовної оптимізації для квадратичної функції:

$$f(x)=a\,x_1^2+b\,x_1\,x_2+c\,x_2^2+d\,x_1+e\,x_2\rightarrow min,$$

запропонованими методами безумовної оптимізації. Коефіцієнти a,b,c,d,e задані в таблиці 3.9.

Таблиця 3.9

№	a	b	c	d	e	№	a	b	c	d	e
1.	1	2	2	–2	–3	39.	2	2	4	–2	–3
2.	7	1	1	–16	–3	40.	8	3	3	1	1
3.	2	2	1	–2	–6	41.	7	–2	1	6	6
4.	1	2	3	–2	–3	42.	2	2	5	–2	–3
5.	9	0	6	–90	–128	43.	4	–5	3	–1	–4
6.	1	–1	8	2	–1	44.	3	2	2	–2	–3
7.	4	2	5	–2	–3	45.	1	2	3	1	1
8.	6	2	1	6	6	46.	3	2	5	–2	–3

9.	1	–1	1	–2	1	47.	2	1	1	–2	–5
10.	3	2	1	–2	–3	48.	3	2	4	–2	–3
11.	5	4	1	–16	–12	49.	1	1	4	2	1
12.	8	2	1	–3	–6	50.	2	2	2	–2	–3
13.	3	2	3	–2	–3	51.	1	–4	1	2	2
14.	9	5	1	6	2	52.	4	2	4	–2	–3
15.	2	1	6	–5	–13	53.	5	2	1	–2	–10
16.	7	–1	1	7	–4	54.	5	2	2	–4	–2
17.	3	1	1	1	5	55.	6	2	1	–2	–3
18.	7	5	1	6	3	56.	2	2	1	–2	–3
19.	4	2	3	–2	–3	57.	5	4	6	–2	–6
20.	9	1	1	2	–1	58.	6	2	3	–2	–6
21.	5	4	1	6	4	59.	5	–6	4	–2	–3
22.	1	2	4	–2	–3	60.	6	2	4	–2	15
23.	3	3	1	6	5	61.	6	4	5	–2	–3
№	*a*	*b*	*c*	*d*	*e*	№	*a*	*b*	*c*	*d*	*e*
24.	6	–1	1	–3	5	62.	5	2	3	–6	–4
25.	4	2	1	–2	–3	63.	7	2	2	–2	–17
26.	3	2	1	12	–6	64.	1	2	6	–2	–3
27.	3	4	2	–2	4	65.	9	3	2	0	3
28.	8	–2	1	–1	1	66.	2	3	3	3	6
29.	2	2	3	–2	–3	67.	4	1	1	5	0
30.	5	–2	1	–2	3	68.	3	–2	1	–4	0
31.	1	2	5	–2	–3	69.	1	2	4	–3	1
32.	5	–2	2	2	3	70.	6	4	3	–2	4
33.	1	3	3	–1	–2	71.	1	5	8	–3	–4
34.	6	–3	2	–9	–2	72.	2	–2	4	–2	–5
35.	4	2	1	–3	–5	73.	1	4	5	1	–1
36.	4	2	2	–2	–3	74.	7	4	2	10	0
37.	1	3	3	–4	5	75.	8	5	1	2	3
38.	8	–4	1	0	1	76.	1	2	4	–1	–3

Розділ 4. Чисельні методи умовної оптимізації

У цьому розділі розглядаються чисельні методи для задачі умовної оптимізації:

$$f(x) \to min, \; x \in X \subset E^n.$$

Розробка чисельних методів для розв'язання задач оптимізації з обмеженнями є більш важкою проблемою, ніж побудова методів безумовної оптимізації. Ефективні алгоритми вдається побудувати лише для спеціальних класів умовних задач, до яких, у першу чергу, потрібно віднести задачі лінійного, квадратичного та опуклого програмування.

У цьому розділі розглядаються деякі найбільш відомі методи оптимізації, які часто використовуються на практиці. Зроблено короткий опис кожного з методів, досліджено питання збіжності, сформульовані алгоритми методів, наведені приклади.

§1. Метод проекції градієнта

Метод проекції градієнта є узагальненням градієнтного методу безумовної оптимізації на випадок задач умовної оптимізації.

Розглянемо задачу

$$f(x) \to min, \tag{4.1}$$

$$x \in X, \tag{4.2}$$

де X – замкнена, опукла множина в E^n, $f(x)$ – диференційовна функція на X.

Проекцією точки a на множину X називається точка $P_X(a)$, яка є найближчою до a серед усіх точок множини X. Тобто, $P_X(a)$ є розв'язком задачі проектування

$$\varphi(x) = \| x - a \|^2 \to min, \; x \in X. \tag{4.3}$$

Лема. Нехай X – замкнена, опукла множина в E^n. Тоді

1) проекція $P_X(a)$ будь-якої точки $a \in E^n$ існує і єдина;

2) точка $\bar{x}$ є проекцією точки a на множині X ($\bar{x} = P_X(a)$) тоді і тільки тоді, якщо $(\bar{x} - a, x - \bar{x}) \ge 0$ для всіх $x \in X$;

3) для будь-яких точок $a_1, a_2 \in E^n$ справедлива нерівність

$$\| P_X(a_1) - P_X(a_2) \| \le \| a_1 - a_2 \|,$$

тобто оператор проектування має властивість нерозтягнення відстаней.

У методі проекції градієнта за наступну точку наближення до розв'язку задачі (4.1), (4.2) вибирається точка

$$x^{(k+1)} = P_X\left(x^{(k)} - \alpha_k f'\left(x^{(k)}\right)\right), \ k = 0,1,2,\ldots, \tag{4.4}$$

Коефіцієнти $\alpha_k \ge 0$ можна вибирати за тими ж правилами, які сформульовані в розділі 3, § 4. Наведемо теорему про збіжність варіанта методу з апріорним вибором α_k.

Теорема. Нехай множина X є замкненою і опуклою множиною в E^n. Функція $f(x)$ – сильно опукла з константою $M > 0$ і диференційовна на X, її градієнт задовольняє умові Ліпшиця

$$\| f'(x') - f'(x'') \| \le L \| x' - x'' \|, \text{ для будь–яких } x', x'' \in X.$$

Тоді послідовність $\{x^{(k)}\}$, яка генерується за правилом (4.4), де $x^{(0)}$ – довільна точка з X, а $\alpha_k \equiv \alpha \in \left(0, \frac{4M}{L^2}\right)$, збігається до розв'язку x_* задачі (4.1), (4.2) зі швидкістю геометричної прогресії

$$\| x^{(k+1)} - x_* \| \le q \| x^{(k)} - x_* \|,$$

де $q = \sqrt{1 - 4M\alpha + \alpha^2 L^2} \in (0,1)$.

В описаному методі на кожній k-ій ітерації потрібно проводити операцію проектування точки на множину X, тобто розв'язувати задачу вигляду (4.3) при $a = x^{(k)} - \alpha_k f'\left(x^{(k)}\right)$. У деяких випадках вдається вказати явний вигляд проекції:

1) якщо $X = \{ x \in E^n : \| x - c \| \le R \}$ – куля, то

$$P_X(a) = c + \frac{a-c}{\|a-c\|} R;$$

2) якщо $X = \left\{ x \in E^n : b_j \le x_j \le c_j ,\ j = \overline{1,n} \right\}$ – координатний паралелепіпед, то

$$\left(P_X(a)\right)_j = \begin{cases} b_j, & a_j < b_j, \\ a_j, & b_j \le a_j \le c_j, \\ c_j, & a_j > c_j, \end{cases}$$

3) якщо $X = \left\{ x \in E^n : x_j \ge 0,\ j = \overline{1,n} \right\}$ – невід'ємний октант, то

$$P_X(a) = \left(max(0, a_1), \ldots, max(0, a_n) \right);$$

4) якщо $X = \left\{ x \in E^n : (p, x) \ge \beta \right\},\ p \ne 0$ – півпростір, то

$$P_X(a) = a + max(0, \beta - (p, a)) \frac{p}{\|p\|^2};$$

5) якщо $X = \left\{ x \in E^n : Ax = b \right\}$ – афінна множина, при цьому рядки матриці A лінійно незалежні, то

$$P_X(a) = a - A^T \left(A \cdot A^T \right)^{-1} (Aa - b),$$

де A^T – транспонована матриця.

Але, якщо множина X задається за допомогою більш-менш складної системи рівностей та нерівностей, то метод проекції градієнта застосувати практично неможливо, тому що задача (4.3) є такої ж складності, як і початкова.

Відзначимо, що використовуючи ідею проектування, можна модифікувати стосовно до задач умовної оптимізації та інші методи безумовної оптимізації, в тому числі метод Ньютона, методи спряжених напрямків.

1. Метод проекції градієнта Розена

Як відомо, напрямком найшвидшого спуску є антиградієнт цільової функції. Але при наявності обмежень рух уздовж напрямку найшвидшого спуску може привести до недопустимих точок. У методі проекції градієнта

Розена антиградієнт проектується таким чином, що значення цільової функції покращується і, в той же час, зберігається допустимість точок траєкторії.

Припустимо, що всі обмеження $g_i(x) \geq 0, i = \overline{1,n}$ – лінійні, та перепишемо їх у вигляді

$$a_i x \leq b_i \ , i \in I_1, \tag{4.5}$$

$$a_i x = b_i \ , i \in I_2, \tag{4.6}$$

де $a_i = \{a_{i1}, a_{i2} \ldots, a_{in}\}$ – вектори-стовпці, I_1, I_2 – множини індексів.

Нехай допустимий розв'язок x є відомим і $I_0(x) = \{ i : \ \ a_i x = b_i, i \in I_1 \} \cup I_2$ є множиною індексів обмежень, які є активними в точці x. Знайдемо напрямок p $(\| p \| = 1)$, який дозволяє зменшити значення цільової функції $f(x)$ (мінімізувати $f'(x) \cdot p$) й залишитися в множині допустимих розв'язків. Це призводить до того, що напрямок p повинен задовольняти співвідношенням:

$$a_i \ p = 0, \quad \forall i \in I_0(x). \tag{4.7}$$

Позначимо через A_0 підматрицю матриці обмежень A, яка складається з рядків $i \in I_0(x)$ матриці A. Припустимо, що ранг матриці A_0 дорівнює q і дорівнює кількості елементів множини $I_0(x)$.

Теорема. Якщо A_0 є матриця вимірності $q \times n$ і рангу $q \leq n$, то оптимальним розв'язком задачі

$$f'(x) \cdot p \to min,$$

$$A_0 \ p = 0, \quad \| p \| = 1$$

є вектор $p = \dfrac{\bar{y}}{\| \bar{y} \|}$, де $\bar{y}$ – проекція вектора $(- f'(x))$ на $S_0 = \{ y : A_0 \, y = 0 \}$, яка визначається за формулою

$$\bar{y} = -P_0 \, f'(x) = -\left(I - A_0^T \left[A_0 A_0^T \right]^{-1} A_0 \right) f'(x).$$

Матриця P_0 називається **матрицею проектування**.

Зауважимо, що при $\bar{y} \neq 0$ напрямок $\bar{y}$ є можливим напрямком спуску, тому що

$$f'(x) \cdot y = -\bar{y}^T \bar{y} < 0 .$$

Якщо $\bar{y} = -P_0 f'(x) = 0$ у поточній точці x, то

$$f'(x) + A_0^T u = 0, \tag{4.8}$$

$$u = -\left[A_0 A_0^T\right]^{-1} A_0 f'(x) \tag{4.9}$$

Так як стовпці матриці A_0^T є градієнтами активних обмежень у точці x, то ясно, що при $u \geq 0$ співвідношення (4.8) виражає умови Куна-Такера у поточній точці x, а це – рівносильне твердженню, що точка x є точкою локального оптимуму задачі.

У випадку, коли $\bar{y} = 0$, але вектор u, який визначається співвідношенням (4.9) має строго від'ємні компоненти, тоді поточна точка може не бути оптимальною і необхідно шукати інший напрямок переміщення. Для цього виключимо з $I_0(x)$ одне з i обмежень, для яких $u_i < 0$ (наприклад те, для якого u_i має найбільше значення за модулем). Отримаємо нову матрицю проектування $\overline{P_0}$, яка дозволить знайти новий напрямок переміщення $\bar{y} = -\overline{P_0} f'(x)$.

Позначимо через $\overline{I_0}(x)$ множину обмежень неактивних у допустимій точці x.

$$\overline{I_0}(x) = \{\, i : \quad a_i x = b_i, i \in I_1 \,\},\ A_1 = \{a_i\},\ b_1 = \{b_i\},\ i \in \overline{I_0}(x).$$

Для кращого розуміння методу проекції градієнта Розена розглянемо геометричну ілюстрацію методу (рис. 4.1). Припустимо, що отримано допустимий розв'язок $x^{(k)}$. Ми не можемо рухатися в напрямку градієнта, не порушуючи обмеження 1. Матриця A_0, яка використовується в методі, є просто $A_0 = a_1$. Як показано на рис. 4.1, вектор $p^{(k)}$ є ортогональною проекцією вектора $\left(-f'\left(x^{(k)}\right)\right)$ на границю множини допустимих розв'язків. Таким чином, на наступному кроці, потрібно рухатися вздовж границі, доти не буде досягнута крайня точка A, в якій ітераційний процес закінчиться.

Розглянемо тепер ситуацію, яку зображено на рис. 4.2. У цьому прикладі будуть порушуватись обидва обмеження, якщо рухатися в напрямку градієнта. Ми можемо знайти напрямок руху, якщо припустимо, що обмеження $a_2 p \le 0$ виконувалося би як нерівність. Щоб отримати матрицю $\overline{A_0}$, викреслимо рядок, який відповідає a_2 з A_0, та прийдемо до попереднього випадку.

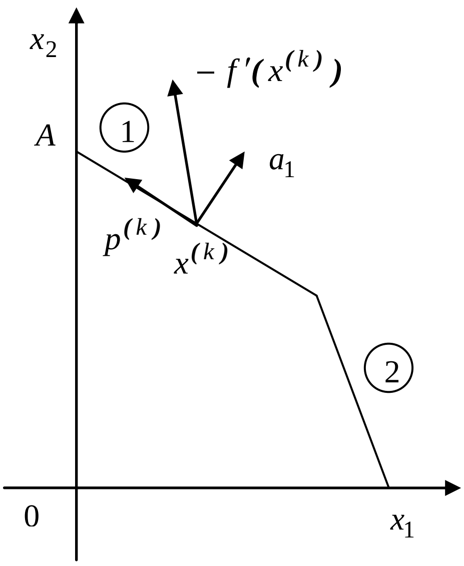


Рис. 4.1

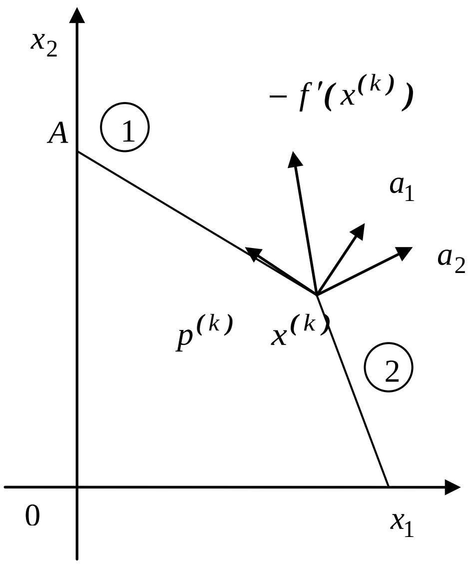


Рис. 4.2

Алгоритм методу проекції градієнту Розена

Початковий етап. Задати допустиму точку $x^{(1)}$, покласти $k=1$ та перейти до основного етапу.

Основний етап

Крок 1. Визначити множину $I_0\left(x^{(k)}\right)$ індексів активних обмежень та множину індексів неактивних обмежень $\overline{I_0}\left(x^{(k)}\right)$.

Крок 2. Нехай A_0 – матриця, рядки якої відповідають обмеженням $i \in I_0\left(x^{(k)}\right)$, а A_1 – матриця, рядки якої відповідають обмеженням $i \in \overline{I_0}(x^{(k)})$. Якщо A_0 порожня, то покласти $P_0 = I$. У протилежному випадку обчислити матрицю проектування

$$P_0 = I - A_0^T\left[A_0 A_0^T\right]^{-1} A_0 \text{ та вектор } p^{(k)} = -P_0 f'\left(x^{(k)}\right).$$

Якщо $p^{(k)} = 0$, то перейти на крок 4.

Крок 3. Якщо $p^{(k)} \neq 0$, то знайти

$$\alpha_{max} = \begin{cases} \min\limits_i \left(\dfrac{e_i}{d_i}, d_i > 0 \right), d > 0, \\ \infty, \quad d \leq 0 \end{cases} \tag{4.10}$$

де $e = b_1 - A_1 x^{(k)}$, $d = A_1 p^{(k)}$. Потім знайти таке $x^{(k+1)}$, що

$$f\left(x^{(k+1)}\right) = \min_{0 \leq \alpha \leq \overline{\alpha}} f\left(x^{(k)} + \alpha\ p^{(k)}\right),\ \overline{\alpha} = \alpha_{max}.$$

Покласти $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, $k = k+1$ та перейти до кроку 1.

Крок 4. Якщо $p^{(k)} = 0$ і матриця A_0 порожня, то зупинитися. Нехай

$$u = -\left[A_0 A_0^T\right]^{-1} A_0 f'\left(x^{(k)}\right).$$

Якщо $u \geq 0$, то зупинитися: $x^{(k)}$ задовольняє умовам Куна-Такера. У протилежному випадку нехай u_i – найбільша за модулем з від'ємних компонент вектора u. Покласти $I_0\left(x^{(k)}\right) = I_0\left(x^{(k)}\right) - \{i\}$ та перейти до кроку 1.

Алгоритм описаний.

При реалізації методу Розена на ЕОМ потрібно мати на увазі можливість виникнення додаткових труднощів. Наприклад, навіть, якщо $p^{(k)} = -P_0 f'\left(x^{(k)}\right) = 0$, то при наявності помилок округлення вектор $p^{(k)}$, який обчислений на комп'ютері, може відрізнятися від нульового. Отже, потрібно використовувати інше правило переходу до викреслювання рядку з A_0. Одне з правил таке: завжди потрібно обчислювати $f'\left(x^{(k)}\right) \cdot p^{(k)}$, $p^{(k)} = -P_0 f'\left(x^{(k)}\right)$ та $f'\left(x^{(k)}\right) \cdot \overline{p}^{(k)}$, $\overline{p}^{(k)} = -\overline{P_0} f'\left(x^{(k)}\right)$, де $\overline{A_0}$ може бути отримана з A_0 шляхом викреслювання рядка, який відповідає найбільшому за модулем від'ємному u_i, коли є хоча б одне від'ємне u_i. Потім потрібно вибрати напрямок, який відповідає більшому з цих двох чисел.

Приклад. Розглянемо таку задачу

$$f(x) = 2x_1^2 + 2x_2^2 - 2x_1 x_2 - 4x_1 - 6x_2 \to min$$

при обмеженнях

$$x_1 + x_2 \le 2,$$

$$x_1 + 5x_2 \le 5,$$

$$-x_1 \qquad \le 0,$$

$$-x_2 \le 0.$$

Візьмемо за початкову точку $x^{(1)} = (0,0)$. Відзначимо, що

$$f'(x) = \begin{pmatrix} 4x_1 - 2x_2 - 4 \\ -2x_1 + 4x_2 - 6 \end{pmatrix}.$$

Ітерація 1. Пошук напрямку. У точці $x^{(1)}$ маємо $f'\left(x^{(1)}\right) = (-4,-6)$. Множина індексів активних обмежень $I_0\left(x^{(1)}\right) = \{3,4\}$, а $\overline{I_0}\left(x^{(1)}\right) = \{1,2\}$. Матриця A_0 має вигляд

$$A_0 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \; A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}.$$

Тоді

$$p^{(0)} = I - A_0^T \left[A_0 A_0^T\right]^{-1} A_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

і $p^{(0)} = (0,0)$. Так як матриця A_0 є не порожньою, то обчислимо $u = -\left[A_0 A_0^T\right]^{-1} A_0 f'\left(x^{(k)}\right) = (-4,-6)$. Виберемо $u_4 = -6$ та видалимо градієнт, який відповідає четвертому обмеженню задачі. Множина $I_0\left(x^{(1)}\right) = \{3\}$, а $\overline{A_0} = (-1,0)$. Нова матриця проектування буде мати вигляд

$$\overline{p}^{(0)} = I - \overline{A_0}^T \left[\overline{A_0}\, \overline{A_0}^T\right]^{-1} \overline{A_0} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

а напрямок руху $p^{(1)}$ визначається вектором

$$p^{(1)} = -\overline{P_0} f'\left(x^{(1)}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -4 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

Лінійний пошук. Будь-яка точка $x^{(2)}$, яка отримана з точки $x^{(1)}$ за напрямком $p^{(1)}$, може бути представлена у вигляді $x^{(2)} = x^{(1)} + \alpha\, p^{(1)} = (0,\ 6\alpha)$, а відповідне цій точці значення цільової функції дорівнює

$f\left(x^{(2)}\right)=72\alpha^2-36\alpha$.

Максимальне значення α, для якого точка $x^{(1)}+\alpha\, p^{(1)}$ є допустимою, знаходиться у відповідності з (4.10) і дорівнює $\alpha_{max}=min\left\{\frac{2}{6},\frac{5}{30}\right\}=\frac{1}{6}$.

Отже, α_1 є оптимальним розв'язком такої задачі:

$$72\alpha^2-36\alpha\to min,\ 0\le\alpha\le\frac{1}{6}.$$

Оптимальний розв'язок дорівнює $\alpha_1=\frac{1}{6}$ та $x^{(2)}=x^{(1)}+\alpha_1 p^{(1)}=(0,1)$.

Ітерація 2. Пошук напрямку. У точці $x^{(2)}=(0,1)$ маємо $f'\left(x^{(2)}\right)=(-6,-2)$. У цій точці активними є друге та третє обмеження, так що маємо

$$A_0=\begin{pmatrix}1 & 5\\ -1 & 0\end{pmatrix},\ A_1=\begin{pmatrix}1 & 1\\ 0 & -1\end{pmatrix}.$$

Далі маємо

$$p^{(0)}=I-A_0^T\left[A_0A_0^T\right]^{-1}A_0=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\ \text{і}\ p^{(0)}=(0,0).$$

Обчислимо $u=-\left[A_0A_0^T\right]^{-1}A_0\ f'\left(x^{(2)}\right)=\left(\frac{2}{5},\frac{-28}{5}\right)$. Так як $u_3<0$, то рядок $(-1,0)$ видалимо з матриці A_0 і отримаємо матрицю $\overline{A_0}=(1,5)$. Матриця проектування і відповідний напрямок визначається так

$$\overline{p}^{(0)}=I-\overline{A}_0^T\left[\overline{A}_0\overline{A}_0^T\right]^{-1}\overline{A}_0=\begin{pmatrix}\frac{25}{26} & \frac{-5}{26}\\ \frac{-5}{26} & \frac{1}{26}\end{pmatrix},$$

$$p^{(2)}=-\overline{P_0}\,f'\left(x^{(2)}\right)=\left(\frac{40}{13}\ ,\frac{-14}{13}\right).$$

Так як довжина вектора $p^{(2)}$ несуттєва, то вектор $\left(\frac{40}{13}\quad,\frac{-14}{13}\right)$ замінимо на вектор $p^{(2)}=(5,-1)$.

Лінійний пошук. Розглянемо точку $x^{(3)}=x^{(2)}+\alpha\, p^{(2)}=(5\alpha,1-\alpha)$, в якій значення цільової функції дорівнює $f\left(x^{(3)}\right)=62\alpha^2-28\alpha-4$.

Максимальне значення α, для якого точка $x^{(2)}+\alpha\, p^{(2)}$ є допустимою, дорівнює
$$\alpha_{max}=min\left\{\frac{1}{4},1\right\}=\frac{1}{4}.$$

Визначимо α_2 з розв'язання задачі
$$62\alpha^2-28\alpha-4\to min,\ 0\le\alpha\le\frac{1}{4}.$$

Оптимальним розв'язком є $\alpha_2=\frac{7}{31}$ та $x^{(3)}=\left(\frac{35}{31},\frac{24}{31}\right)$.

Ітерація 3. Пошук напрямку. У точці $x^{(3)}=\left(\frac{35}{31},\frac{24}{31}\right)$ маємо $f'\left(x^{(3)}\right)=\left(-\frac{32}{31},-\frac{160}{31}\right)$. У цій точці активними є тільки друге обмеження, так що маємо $A_0=(1,5)$.

Далі маємо
$$p^{(0)}=I-A_0^T\left[A_0A_0^T\right]^{-1}A_0=\frac{1}{26}\begin{pmatrix}25 & -5\\ -5 & 1\end{pmatrix}\text{ і } p^{(3)}=(0,0).$$

Обчислимо $u=-\left[A_0A_0^T\right]^{-1}A_0\,f'\left(x^{(3)}\right)=\frac{32}{31}>0.$ Отже, точка $x^{(3)}$ є оптимальною. Так як функція f у цьому прикладі є опуклою, то точка $x^{(3)}$ є точкою глобального мінімуму задачі.

2. Метод зведеного градієнта Вольфа

Розглянемо іншу процедуру побудови можливого напрямку спуску. Метод базується на скороченні вимірності задачі за допомогою представлення всіх змінних через підмножину незалежних змінних.

Знайдемо розв’язок такої задачі:

$$f(x) \to min, \tag{4.11}$$

$$Ax = b, \quad x \geq 0, \tag{4.12}$$

де A – матриця порядку $m \times n$ і рангу m, b – вектор простору E^m, цільова функція $f(x) \in C^1(E^n)$.

Зробимо таке припущення про невиродженість матриці A: будь-які m стовпців матриці A лінійно незалежні та кожна екстремальна точка допустимої області має рівно m додатних змінних і саме більше $n-m$ нульових компонент.

Нехай B є базис, тобто неособлива матриця порядку $m \times n$, яка вилучена з матриці A. Покладемо $A = [B, N]$, $x^T = \lfloor x_B^T, x_N^T \rfloor$. Обмеження (4.12) можна записати у вигляді

$$B x_B + N x_N = b.$$

Це представлення дозволяє записати базисні змінні через небазисні:

$$x_B = \bar{b} - \overline{N} x_N,$$

де $\bar{b} = B^{-1} b$, $\overline{N} = B^{-1} N$.

Нехай $f'(x)^T = \lfloor f_B'(x)^T, f_N'(x)^T \rfloor$, де $f_B'(x)$ – градієнт цільової функції $f(x)$ за базисним змінним x_B, а $f_N'(x)$ – градієнт цільової функції $f(x)$ за небазисним змінним x_N. Нагадаємо, що напрямок p є напрямком спуску і можливим напрямком для цільової функції $f(x)$ у точці x, якщо

$$f'(x)^T \cdot p < 0, \quad A p = 0 \text{ і } p_j \geq 0, \text{ якщо } x_j = 0.$$

Тепер точно визначимо вектор p, який має вказані властивості. Насамперед представимо вектор p у вигляді $\left[p_B^T, p_N^T\right]$. Відзначимо, що рівність

$$B\,p_B + N\,p_N = 0 = A\,p$$

автоматично виконується, якщо для будь–якого p_N покласти $p_B = -B^{-1}N\,p_N$.

Нехай вектор

$$r^T = \left[r_B{}^T, r_N{}^T\right] = f'(x)^T - f_B'(x)^T B^{-1}A = \left[0,\, f_N'(x)^T - f_B'(x)^T B^{-1}N\right]$$

є **зведений градієнт**.

Розглянемо добуток $f'(x)^T \cdot p$:

$$f'(x)^T \cdot p = f_B'(x)^T\, p_B + f_N'(x)^T\, p_N = \left[f_N'(x)^T - f_B'(x)^T B^{-1}N\right] p_N = r_N^T\; p_N .$$

Ми повинні вибрати p_N так, щоб $r_N^T\; p_N < 0$ і $p_j \geq 0$, якщо $x_j = 0$.

Для кожної небазисної змінної покладемо

$$p_j = \begin{cases} -r_j, \text{якщо } r_j \leq 0, \\ -x_j\, r_j, \text{якщо } r_j > 0. \end{cases}$$

Якщо $p_N = 0$, то умови Куна-Такера виконуються. Для опуклої цільової функції $f(x)$ це означає, що досягається глобальний оптимум. Для неопуклої цільової функції $f(x)$ маємо локальний оптимум.

Алгоритм методу зведеного градієнта Вольфа

Початковий етап. Задати допустиму точку $x^{(1)}$, покласти $k=1$ та перейти до основного етапу.

Основний етап

Крок 1. Визначити множину I_k як множину індексів m найбільших компонент вектора $x^{(k)}$:

$$B = \left\{A_j,\, j \in I_k\right\}, \quad N = \left\{A_j,\, j \notin I_k\right\}, \tag{4.13}$$

$$r^T = f'\left(x^{(k)}\right)^T - f_B'\left(x^{(k)}\right)^T B^{-1}A, \tag{4.14}$$

$$\left(p_j\right)_N = \begin{cases} -r_j, \text{якщо } j \notin I_k,\ r_j \leq 0, \\ -x_j\, r_j, \text{якщо } j \in I_k,\ r_j > 0, \end{cases} \tag{4.15}$$

$$p_B = -B^{-1}N\, p_N\,. \tag{4.16}$$

Покласти $\left(p^{(k)}\right)^T = \left[p_B^T, p_N^T\right]$. Якщо $p^{(k)} = 0$, то зупинитися: точка $x^{(k)}$ є точкою Куна-Такера. У протилежному випадку перейти до кроку 2.

Крок 2. Розв'язати таку задачу одновимірної мінімізації

$$f\left(x^{(k)} + \alpha\, p^{(k)}\right) \to min, \quad 0 \le \alpha \le \alpha_{max}, \tag{4.17}$$

де

$$\alpha_{max} = \begin{cases} \min\limits_{1 \le j \le n}\left(-\dfrac{x_j^{(k)}}{p_j^{(k)}}, p_j^{(k)} < 0\right), p^{(k)} < 0. \\ \infty, \quad p^{(k)} \ge 0. \end{cases} \tag{4.18}$$

Покласти α_k таким, що дорівнює оптимальному розв'язку задачі (4.17), (4.18), $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Теорема про збіжність. Нехай функція $f(x) \in C^1\left(E^n\right)$, A – матриця порядку $m \times n$, а b – m – вимірний вектор. Припустимо, що всі екстремальні точки допустимої області мають m додатних компонент і будь-які m стовбців матриці A лінійно незалежні. Тоді будь-яка гранична точка послідовності $\left\{x^{(k)}\right\}$, яку знайдено у відповідності до алгоритму методу зведеного градієнта, є точкою Куна-Такера.

Приклад. Розглянемо задачу

$$f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \to min$$

при обмеженнях

$$x_1 + x_2 \le 2,$$

$$x_1 + 5x_2 \le 5,$$

$$-x_1 \le 0,$$

$$-x_2 \le 0.$$

Запишемо обмеження задачі у вигляді (4.12):

$$x_1 + x_2 + x_3 = 2,$$

$$x_1 + 5x_2 + x_4 = 5,$$

$$x_j \qquad \geq 0, \quad j = \overline{1,4}.$$

Розв'яжемо цю задачу методом зведеного градієнта Вольфа, взявши за початкову точку $x^{(1)} = (0, 0, 2, 5)$. Відзначимо, що

$$f'(x)^T = (4x_1 - 2x_2 - 4; -2x_1 + 4x_2 - 6;\ 0;\ 0).$$

Ітерація 1. Пошук напрямку. Матриця обмежень має вигляд:

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{pmatrix}.$$

Множина $I_1 = \{3, 4\}$, так що

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ N = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}.$$

Результати обчислень для точки $x^{(1)}$ наведені в таблиці 4.1.

Таблиця 4.1

<table>
<tr><td colspan="2" rowspan="2">Розв'язок $x^{(1)}$</td><td>x_1</td><td>x_2</td><td>x_3</td><td>x_4</td></tr>
<tr><td>0</td><td>0</td><td>2</td><td>5</td></tr>
<tr><td colspan="2">$f'\left(x^{(1)}\right)$</td><td>–4</td><td>–6</td><td>0</td><td>0</td></tr>
<tr><td rowspan="2">$f'_B\left(x^{(1)}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$</td><td>$x_3$</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>x_4</td><td>1</td><td>5</td><td>0</td><td>1</td></tr>
<tr><td colspan="2">r</td><td>–4</td><td>–6</td><td>0</td><td>0</td></tr>
</table>

Відповідно (4.15) маємо $p_N^T = (4, 6)$. Обчислюючи p_B за формулою (4.16), маємо

$$p_B = -B^{-1}N\ p_N = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -10 \\ -34 \end{pmatrix},\ p_B = (p_3, p_4).$$

Відзначимо, що $B^{-1}N$ записана під змінними, які відповідають матриці N (у даному випадку під x_1 і x_2). Вектор напрямку $p^{(1)} = (4, 6, -10, -34)^T$.

Лінійний пошук. При початковій точці $x^{(1)} = (0, 0, 2, 5)$ мінімізуємо цільову функцію за напрямком $p^{(1)} = (4, 6, -10, -34)^T$. Максимальне значення α, для якого точка $x^{(1)} + \alpha\, p^{(1)}$ є допустимою, обчислюється за формулою (4.18) і дорівнює

$$\alpha_{max} = min\left\{\frac{2}{10}, \frac{5}{34}\right\} = \frac{5}{34}.$$

Задача лінійного пошуку (4.17) має вигляд

$$56\alpha^2 - 52\alpha \to min,\ 0 \le \alpha \le \frac{5}{34}.$$

Оптимальний розв'язок дорівнює $\alpha_1 = \dfrac{5}{34}$ та

$$x^{(2)} = x^{(1)} + \alpha_1 p^{(1)} = \left(\frac{10}{17}, \frac{15}{17}, \frac{9}{17}, 0\right).$$

Ітерація 2. Пошук напрямку. У точці $x^{(2)}$ маємо $I_2 = \{1, 2\}$,

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix},\ N = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Ця інформація зведена до таблиці 4.2.

Таблиця 4.2

		x_1	x_2	x_3	x_4
Розв'язок $x^{(2)}$		$\frac{10}{17}$	$\frac{15}{17}$	$\frac{9}{17}$	0
$f'\left(x^{(2)}\right)$		$-\frac{58}{17}$	$-\frac{62}{17}$	0	0
$f'_B\left(x^{(2)}\right) = \begin{pmatrix} -\frac{58}{17} \\ -\frac{62}{17} \end{pmatrix}$	x_1	1	0	$\frac{5}{4}$	$-\frac{1}{4}$
	x_2	1	5	$-\frac{1}{4}$	$\frac{1}{4}$
r		0	0	$\frac{57}{17}$	$\frac{4}{17}$

У відповідності з (4.14) маємо

$$r^T = \left(-\frac{58}{17}, -\frac{62}{17}, 0, 0\right) - \left(-\frac{58}{17}, -\frac{62}{17}\right)^T \begin{pmatrix} 1 & 0 & 5/4 & -1/4 \\ 0 & 1 & -1/4 & 1/4 \end{pmatrix} = \left(0, 0, \frac{57}{17}, \frac{4}{17}\right).$$ Тоді,

згідно з (4.15) маємо $p_N = \left(-\frac{513}{289}, 0\right)$. З формули (4.16) знайдемо

$$p_B = -\begin{pmatrix} \frac{5}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} -\frac{513}{289} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2565}{1156} \\ -\frac{513}{1156} \end{pmatrix}, \ p_B = (p_1, p_2).$$

Таким чином, етап пошуку напрямку привів до вектора $p^{(2)} = \left(\frac{2565}{1156}, \frac{-513}{1156}, \frac{-513}{289}, 0\right)^T$.

Лінійний пошук. Починаючи процедуру з точки $x^{(2)}$ мінімізуємо цільову функцію в напрямку $p^{(2)}$. Максимальне значення α, для якого точка $x^{(2)} + \alpha\ p^{(2)}$ є допустимою, обчислюється за формулою (4.18) і дорівнює

$\alpha_{max} = \frac{17}{57}$. Значення α_2 отримаємо із розв’язання такої задачі:

$$12{,}21\alpha^2 - 5{,}95\alpha - 6{,}436 \to min, \ 0 \le \alpha \le \frac{17}{57}.$$

Оптимальний розв’язок дорівнює $\alpha_2 = \frac{68}{279}$. та

$$x^{(3)} = x^{(2)} + \alpha_2 p^{(2)} = \left(\frac{35}{31}, \frac{24}{31}, \frac{3}{31}, 0\right).$$

Ітерація 3. Пошук напрямку. Тепер $I_3 = \{1, 2\}$. Так як $I_3 = I_2$, то таблицю попередньої ітерації можна залишити

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}, \ N = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Ця інформація зведена до таблиці 4.3.

Таблиця 4.3

<table>
<tr><td colspan="2" rowspan="2">Розв'язок $x^{(3)}$</td><td>x_1</td><td>x_2</td><td>x_3</td><td>x_4</td></tr>
<tr><td>$\frac{35}{31}$</td><td>$\frac{24}{31}$</td><td>$\frac{3}{31}$</td><td>0</td></tr>
<tr><td colspan="2">$f'\left(x^{(3)}\right)$</td><td>$-\frac{32}{31}$</td><td>$-\frac{160}{31}$</td><td>0</td><td>0</td></tr>
<tr><td rowspan="2">$f'_B\left(x^{(3)}\right)=\begin{pmatrix} -\frac{32}{31} \\ -\frac{160}{31} \end{pmatrix}$</td><td>$x_1$</td><td>1</td><td>0</td><td>$\frac{5}{4}$</td><td>$-\frac{1}{4}$</td></tr>
<tr><td>x_2</td><td>1</td><td>5</td><td>$-\frac{1}{4}$</td><td>$\frac{1}{4}$</td></tr>
<tr><td colspan="2">r</td><td>0</td><td>0</td><td>$\frac{57}{17}$</td><td>$\frac{4}{17}$</td></tr>
</table>

Із формул (4.15), (4.16) маємо $p_N = (0,0)^T$, $p_B = (0,0)$. Отже, $p^{(3)} = (0,0,0,0)^T$ і розв'язок $x^{(3)}$ є оптимальним.

§2. Метод умовного градієнта

У цьому параграфі розглядається метод умовного градієнта або метод лінійної апроксимації (лінеаризації) цільової функції.

Розглянемо задачу

$$f(x) \to min,\ x \in X \subset E^n, \qquad (4.19)$$

де X – опукла, замкнена, обмежена множина в E^n, $f(x) \in C^1(X)$.

Ітераційна формула методу умовного градієнта має вигляд:

$$x^{(k+1)} = x^{(k)} + \alpha_k h^{(k)},\ k = 0,1,2,\ldots, \qquad (4.20)$$

де $h^{(k)}$ – напрямок спуску цільової функції $f(x)$ у точці $x^{(k)}$, α_k – параметр, який регулює довжину кроку вздовж $h^{(k)}$.

Для вибору $h^{(k)}$ на k-тій ітерації розв'язується задача мінімізації на множині X лінійної апроксимації цільової функції $f(x)$ в точці $x^{(k)}$, тобто такої функції:

$$f_k(x) = f\left(x^{(k)}\right) + \left(f'\left(x^{(k)}\right), x - x^{(k)}\right),$$

де $\left(f'\left(x^{(k)}\right), x - x^{(k)}\right)$ – головна лінійна частина приросту цільової функції $f(x)$ у точці $x^{(k)}$. Відкидаючи константу $f\left(x^{(k)}\right)$, цю задачу можна записати у вигляді

$$\left(f'\left(x^{(k)}\right), x - x^{(k)}\right) \to min, \; x \in X. \qquad (4.21)$$

Нехай $\bar{x}^{(k)}$ – оптимальний розв'язок задачі (4.21), а $f_k\left(\bar{x}^{(k)}\right) = \left(f'\left(x^{(k)}\right), \bar{x}^{(k)} - x^{(k)}\right)$ – оптимальне значення цільової функції задачі (4.21). Згідно теоремі Вейерштраса розв'язок $\bar{x}^{(k)}$ задачі (4.21) завжди існує. Якщо задача (4.21) має декілька розв'язків, то обирається один з них.

Враховуючи, що $x^{(k)} \in X$, маємо

$$\min_x f_k(x) = f_k\left(\bar{x}^{(k)}\right) \le f_k\left(x^{(k)}\right) = 0.$$

Тому можливі лише два випадки: $f_k\left(\bar{x}^{(k)}\right) = 0$ або $f_k\left(\bar{x}^{(k)}\right) < 0$.

Якщо $f_k\left(\bar{x}^{(k)}\right) = 0$, то $\left(f'\left(x^{(k)}\right), \bar{x}^{(k)} - x^{(k)}\right) \ge f_k\left(\bar{x}^{(k)}\right) = 0$ для будь-яких $x \in X$, тобто точка $x^{(k)}$ – стаціонарна точка задачі (4.19) згідно теоремі 2 § 3 розділу 1. Робота алгоритму завершується, точку $x^{(k)}$ необхідно дослідити на оптимальність. Якщо функція $f(x)$ – опукла функція на множині X, то згідно другій частині теореми 2 § 3 розділу 1, точка $x^{(k)}$ – розв'язок задачі (4.19).

Нехай тепер $f_k\left(\bar{x}^{(k)}\right) < 0$. У цьому випадку $\bar{x}^{(k)} \ne x^{(k)}$. Тоді у формулі (4.20) покладемо $h^{(k)} = \bar{x}^{(k)} - x^{(k)}$ і тоді ітераційна формула (4.20) запишеться у вигляді

$$x^{(k+1)} = x^{(k)} + \alpha_k\left(\bar{x}^{(k)} - x^{(k)}\right), \; k = 0,1,2,\ldots, \qquad (4.20)$$

Вектор $h^{(k)}$ прийнято називати **умовним антиградієнтом** цільової функції $f(x)$ у точці $x^{(k)}$.

Так як множина допустимих розв'язків X є опуклою, то для будь-якого α_k з відрізка $[0,1]$ точка $x^{(k+1)} = x^{(k)} + \alpha_k\left(\bar{x}^{(k)} - x^{(k)}\right) \in X$.

Способи вибору крокового множника α_k

1. Визначимо α_k з умови:

$$g_k(\alpha_k) = \min_{0 \le \alpha \le 1} g_k(\alpha), \tag{4.23}$$

$$g_k(\alpha) = f\left(x^{(k)} + \alpha h^{(k)}\right).$$

2. Параметр α_k виберемо апріорно:

$$0 < \alpha_k \le 1,\ k = 0,1,\ldots,\ \sum_{k=0}^{\infty} \alpha_k = \infty,\ \sum_{k=0}^{\infty} \alpha_k^2 < \infty .$$

Наприклад, за α_k можна узяти $\alpha_k = \dfrac{1}{k+1}$.

3. Адаптивний (автоматичний) вибір крокового множника. Параметр α_k вибирається за правилом дроблення до тих пір, доки не виконається нерівність

$$f\left(x^{(k)} + \alpha\left(\bar{x}^{(k)} - x^{(k)}\right)\right) - f\left(x^{(k)}\right) \le \delta\, \alpha\, f_k\left(\bar{x}^{(k)}\right),$$

де δ – параметр методу, $0 < \delta < 1$.

4. Дроблення кроку. Виберемо $\alpha_k = 1$ і перевіримо виконання умови монотонності $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$, а потім при необхідності дробимо кроковий множник α_k ($\alpha_k = \lambda\, \alpha_k$, $\lambda \in (0,1)$) доки не буде виконана умова монотонності.

Умови збіжності методу умовного градієнта

Теорема. Нехай X – замкнена, обмежена, опукла множина в E^n, $f(x) \in C^1(X)$, причому її градієнт задовольняє умові Ліпшиця

$$\left\| f'(x) - f'(z) \right\| \le L \left\| x - z \right\|, \text{ для } \forall\, x, z \in X .$$

Тоді будь-яка гранична точка x_* послідовності $\left\{x^{(k)}\right\}_0^{\infty}$, що визначається умовами (4.21) – (4.23), є стаціонарною в задачі (4.19), тобто

$$\left(f'(x_*), x - x_* \right) \ge 0 \text{ для } \forall\, x \in X .$$

Якщо при цьому цільова функція $f(x)$ опукла на множині X, то x_* – розв’язок задачі (4.19) і $\lim\limits_{k\to\infty} f\left(x^{(k)}\right) = f_*$, де $f_* = \min\limits_{x\in X} f(x)$.

Зауваження. Аналогічні теореми можна сформулювати і для інших способів вибору α_k.

Критерії завершення ітераційного процесу:

$$\left| f_k\left(\bar{x}^{(k)}\right) \right| \le \varepsilon,$$

$$\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon.$$

Алгоритм методу умовного градієнта

Початковий етап. Вибрати початкову точку $x^{(0)} \in X$. Задати константу δ $(0 < \delta < 1)$, константу зупинки $\varepsilon > 0$, покласти $k = 0$ та перейти до основного етапу.

Основний етап

Крок 1. Обчислити $f'\left(x^{(k)}\right)$. Знайти оптимальний розв’язок $\bar{x}^{(k)}$ задачі

$$\left(f'\left(x^{(k)}\right), x - x^{(k)}\right) \to \min,\ x \in X.$$

Крок 2. Якщо $f_k\left(\bar{x}^{(k)}\right) = \left(f'\left(x^{(k)}\right), x - x^{(k)}\right) = 0$,то зупинитися: $x^{(k)}$ – стаціонарна точка, інакше обчислити умовний антиградієнт $h^{(k)} = \bar{x}^{(k)} - x^{(k)}$, покласти $\alpha_k = 1$.

Крок 3. Перевірити виконання умови

$$f\left(x^{(k)} + \alpha\left(\bar{x}^{(k)} - x^{(k)}\right)\right) - f\left(x^{(k)}\right) \le \delta\, \alpha\, f_k\left(\bar{x}^{(k)}\right). \qquad (4.24)$$

Якщо нерівність виконується, то перейти до кроку 4, інакше провести дроблення крокового множника α_k доки не буде виконано нерівність (4.24).

Крок 4. Обчислити наступне наближення $x^{(k+1)} = x^{(k)} + \alpha_k h^{(k)}$, замінити k на $k+1$. Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon$, то зупинитися, у протилежному випадку – перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо задачу

$$f(x)=2x_1^2+2x_2^2-2x_1x_2-4x_1-6x_2 \to min$$

при обмеженнях

$$x_1+x_2 \le 2,$$

$$x_1+5x_2 \le 5,$$

$$-x_1 \le 0,$$

$$-x_2 \le 0.$$

Градієнт цільової функції дорівнює $f'(x)=(4x_1-2x_2-4,\ -2x_1+4x_2-6)^T$.

Розв'яжемо цю задачу методом умовного градієнта, взявши за початкову точку $x^{(0)}=(0,0)^T$.

Ітерація 1. Пошук напрямку. У початковій точці $x^{(0)}=(0,0)^T$ маємо $f'\left(x^{(0)}\right)=(-4;-6)^T$. Задача для знаходження напрямку $h^{(0)}$ має вигляд:

$$f_0(x)=-4x_1-6x_2 \to min,$$

$$x_1+x_2 \le 2,$$

$$x_1+5x_2 \le 5,$$

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

Сформульована задача є задачею лінійного програмування і може бути розв'язана симплекс–методом. Оптимальним розв'язком цієї задачі є точка $\bar{x}^{(0)}=\left(\frac{5}{4},\frac{3}{4}\right)^T$, $f_0\left(\bar{x}^{(0)}\right)=-9.5$. Так як $f_0\left(\bar{x}^{(0)}\right)<0$, обчислюємо вектор $h^{(0)}=\bar{x}^{(0)}-x^{(0)}=\left(\frac{5}{4},\frac{3}{4}\right)^T$.

Лінійний пошук. Будь-яка точка $x^{(1)}$, яку знайдено з точки $x^{(0)}$ у напрямку $h^{(0)}$, може бути представлена у вигляді $x^{(1)}=x^{(0)}+\alpha\, h^{(0)}$, де $0 \le \alpha \le 1$, а відповідне їй значення цільової функції дорівнює $f\left(x^{(0)}+\alpha\, h^{(0)}\right)=\frac{19}{8}\alpha^2-\frac{19}{2}\alpha$.

Значення α_0 знаходиться з розв'язання такої задачі одновимірної оптимізації:

$$\frac{19}{8}\alpha^2 - \frac{19}{2}\alpha \to min, \; 0 \le \alpha \le 1.$$

Оптимальний розв'язок дорівнює $\alpha_0 = 1$. Отже, $x^{(1)} = \left(\frac{5}{4}, \frac{3}{4}\right)^T$.

Ітерація 2. Пошук напрямку. У точці $x^{(1)} = \left(\frac{5}{4}, \frac{3}{4}\right)^T$ маємо

$f'\left(x^{(1)}\right) = \left(-\frac{1}{2}, -\frac{11}{2}\right)^T$. Для знаходження напрямку $h^{(1)}$ розв'яжемо таку задачу:

$$f_1(x) = -\frac{1}{2}x_1 - \frac{11}{2}x_2 + \frac{19}{4} \to min,$$

$$x_1 + x_2 \le 2,$$

$$x_1 + 5x_2 \le 5,$$

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

Оптимальним розв'язком цієї задачі є точка $\bar{x}^{(1)} = (0,1)^T$, $f_1(\bar{x}^{(1)}) = -\frac{3}{4}$. Так як $f_1\left(\bar{x}^{(1)}\right) < 0$, обчислюємо вектор $h^{(1)} = \bar{x}^{(1)} - x^{(1)} = \left(-\frac{5}{4}, \frac{1}{4}\right)^T$.

Лінійний пошук. Значення кроку α_1 отримаємо мінімізацією функції $f\left(x^{(1)} + \alpha\, h^{(1)}\right) = \frac{31}{8}\alpha^2 - \frac{3}{4}\alpha - \frac{63}{8}$ за умови $0 \le \alpha \le 1$. Оптимальний розв'язок дорівнює $\alpha_1 = \frac{3}{31}$. Отже, $x^{(2)} = \left(\frac{35}{31}, \frac{24}{31}\right)^T$.

Ітерація 3. Пошук напрямку. У точці $x^{(2)} = \left(\frac{35}{31}, \frac{24}{31}\right)^T$ маємо

$f'\left(x^{(1)}\right) = \left(-\frac{32}{31}; -\frac{160}{31}\right)^T$. Задача для знаходження напрямку $h^{(2)}$ має вигляд:

$$f_2(x) = -\frac{32}{31}x_1 - \frac{160}{31}x_2 + \frac{4960}{961} \to min,$$

$$x_1 + x_2 \le 2, \qquad x_1 + 5x_2 \le 5, \qquad x_1 \ge 0, \; x_2 \ge 0.$$

Оптимальним розв’язком цієї задачі є точка $\overline{x}^{(2)} = \left(\frac{5}{4}, \frac{3}{4}\right)^T$, $f_2(\overline{x}^{(2)}) = 0$. Отже,

точка $x^{(2)} = \left(\frac{35}{31}, \frac{24}{31}\right)^T$ є оптимальним розв’язком задачі.

§3. Метод лінеаризації

Даний метод базується на ідеї лінійної апроксимації цільової функції та обмежень задачі. У цьому сенсі він є узагальненням методу умовного градієнта. Разом з тим метод лінеаризації містить і якісно нові моменти: до лінійної апроксимації цільової функції додається квадратичний член і тому в якості допоміжних виникають задачі квадратичного (а не лінійного, як можна було б очікувати) програмування.

Розглянемо задачу математичного програмування вигляду:

$$f(x) \to min,$$

$$g_i(x) \le 0, \quad i = \overline{1,m}, \tag{4.25}$$

де функції $f(x), g_1(x), \dots, g_m(x)$ – диференційовні , а функції $g_1(x), \dots, g_m(x)$, окрім того, опуклі на E^n.

Нехай X – допустима множина задачі (4.25). Будемо вважати, що X – непорожня множина.

Довільній точці $x \in E^n$ поставимо у відповідність таку задачу квадратичного програмування відносно p:

$$(f'(x), p) + \frac{1}{2}\| p \|^2 \to min,$$

$$(g_i'(x), p) + g_i(x) \le 0, \quad i = \overline{1,m}. \tag{4.26}$$

Ця задача сформульована на основі лінійних частин розкладень

$$f(x+p) = f(x) + (f'(x), p) + o(\|p\|),$$

$$g_i(x+p) = g_i(x) + (g_i'(x), p) + o(\|p\|), \quad i = \overline{1,m},$$

причому в цільову функцію добавлено квадратичний член $\frac{1}{2}\| p \|^2$, а константа $f(x)$ відкинута.

Зауважимо, що у силу опуклості функцій $g_1(x),\ldots,g_m(x)$ і умови $X \neq \varnothing$, допустима множина задачі (4.26) завжди непорожня. Завдяки квадратичному члену цільова функція задачі (4.26) є сильно опуклою. Тому ця задача має єдиний розв'язок. Позначимо його через $p(x)$.

У методі лінеаризації послідовність наближень $\{x^{(k)}\}$ генерується за формулою

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)},\ k = 0,1,2,\ldots, \tag{4.27}$$

де $p^{(k)} = p\left(x^{(k)}\right)$ – розв'язок задачі (4.26) при $x = x^{(k)}$, α_k – параметр, який регулює довжину кроку вздовж $p^{(k)}$. При цьому, на відміну від методу умовного градієнта, точки послідовності $\{x^{(k)}\}$ не обов'язково повинні належати множині X, а послідовність $\left\{f\left(x^{(k)}\right)\right\}$ – спадати.

Сформулюємо теорему про збіжність варіанта методу з сталим кроком α_k.

Теорема. Нехай функції $f(x), g_1(x),\ldots,g_m(x)$ є диференційовними у просторі E^n, їхні градієнти задовольняють в E^n умові Ліпшиця, функції $g_1(x),\ldots,g_m(x)$ є опуклими в E^n та існує точка $\bar{x} \in E^n$, така, що $g_i(\bar{x}) < 0$ для всіх $i = \overline{1,m}$ (умова Слейтера). Нехай $x^{(0)} \in X$, причому множина $N_0 = \left\{x \in E^n : f(x) \le f\left(x^{(0)}\right)\right\}$ є обмеженою. Тоді існує число $\bar{\alpha} > 0$ таке, що будь-яка гранична точка x_* послідовності $\{x^{(k)}\}$, яка визначається за формулою (4.27), де $\alpha_k = \alpha \in (0, \bar{\alpha}]$, є стаціонарною в задачі (4.25), тобто

$$f'(x_*) + \sum_{i=1}^{m} \lambda_i^* g_i(x_*) = 0, \tag{4.28}$$

$$\lambda_i^* g_i(x_*) = 0, \quad i = \overline{1,m}, \tag{4.29}$$

де $\lambda_1^* \geq 0, \ldots \lambda_m^* \geq 0$ – певні числа (множники Лагранжа). Якщо при цьому цільова функція $f(x)$ є опуклою у просторі E^n, то x_* – розв'язок задачі (4.25).

Відзначимо, що умова обмеженості множини N_0 забезпечується, наприклад, умовою сильної опуклості функції або умовою, $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$. При доведенні теореми використовуються такі дві леми.

Лема 1. При виконанні припущень теореми множники Лагранжа сім'ї задач (4.26) рівномірно обмежені на N_0, тобто

$$\sum_{i=1}^{m} \lambda_i(x) \leq C \text{ для будь–яких } x \in N_0, \tag{4.30}$$

де C деяка константа.

Уведемо функції

$$\Phi(x, C) = f(x) + C\, g(x),\ \varphi(x, C) = C\, g(x), \tag{4.31}$$

де $g(x) = max\{0, g_1(x), \ldots, g_m(x)\}$.

Корисно відзначити, що функція $\varphi(x, C)$ є штрафною для задачі (4.25) (дивись § 6).

Лема 2. Нехай виконуються припущення теореми. При будь-якому $\varepsilon \in (0,1)$ покладемо

$$\bar{\alpha} = min\left(1,\ \frac{1-\varepsilon}{(1+C)L}\right), \tag{4.32}$$

де C константа з (4.31), а L – константа Ліпшиця функцій $f'(x), g_1'(x), \ldots, g_m'(x)$. Тоді для будь-яких $x \in N_0,\ \alpha \in (0, \bar{\alpha}]$ справедлива нерівність

$$\Phi(x + \alpha\, p, C) \leq \Phi(x, C) - \alpha\varepsilon \| p \|^2, \tag{4.33}$$

де $p = p(x)$ – розв'язок задачі (4.26).

Алгоритм методу лінеаризації

Початковий етап. Вибрати початкову точку $x^{(0)}$, константу зупинки $\varepsilon > 0$, покласти $k = 0$ та перейти до основного етапу.

Основний етап

Крок 1. Обчислити $f'\left(x^{(k)}\right),\ g_i'\left(x^{(k)}\right),\ g_i'\left(x^{(k)}\right)\ i=\overline{1,m}$.

Крок 2. Знайти оптимальний розв'язок $p^{(k)}$ задачі:

$$\left(f'\left(x^{(k)}\right), p\right)+\frac{1}{2}\| p \|^2 \to min,$$

$$\left(g_i'\left(x^{(k)}\right),\ p\right)+g_i\left(x^{(k)}\right)\le 0,\quad i=\overline{1,m}.$$

Крок 3. Взяти за α_k оптимальний розв'язок такої задачі лінійного пошуку:

$$f\left(x^{(k)}+\alpha\ p^{(k)}\right)\to min,\ 0\le\alpha\le 1.$$

Покласти $x^{(k+1)}=x^{(k)}+\alpha_k\ p^{(k)}$, замінити k на $k+1$.

Крок 4. Якщо $\left\|x^{(k+1)}-x^{(k)}\right\|\le\varepsilon$, то зупинитися: $x^{(k)}$ – наближення до точки мінімуму. У протилежному випадку – перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо задачу

$$f(x)=2x_1^2+2x_2^2-2x_1x_2-4x_1-6x_2\to min$$

при обмеженнях

$$x_1+x_2\le 2,$$

$$x_1+5x_2\le 5,$$

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

Розв'яжемо цю задачу методом лінеаризації, взявши за початкову точку $x^{(0)}=(0,0)^T$. Кожна ітерація алгоритму містить розв'язання підзадачі знаходження напрямку, яка визначена в описі кроку 2, а потім лінійний пошук уздовж цього напрямку. Відзначимо, що загальна формула для градієнта має вигляд: $f'(x)=(4x_1-2x_2-4,\ -2x_1+4x_2-6)^T$, $g_4'(x)=(0;-1)$.

Ітерація 1. Пошук напрямку. У точці $x^{(0)}=(0,0)^T$ маємо $f'\left(x^{(0)}\right)=(-4;-6)^T$, $g_1\left(x^{(0)}\right)=-2$, $g_2\left(x^{(0)}\right)=-5$, $g_3\left(x^{(0)}\right)=0$, $g_4\left(x^{(0)}\right)=0$. Напрямок спуску $p^{(0)}$ знайдемо з розв'язання такої задачі квадратичного програмування:

$$\frac{1}{2}\left(p_1^2 + p_2^2\right) - 4p_1 - 6p_2 \to min$$

при обмеженнях

$p_1 + p_2 - 2 \le 0,$

$p_1 + 5p_2 - 5 \le 0,$

$-p_1 \le 0,\ -p_2 \le 0.$

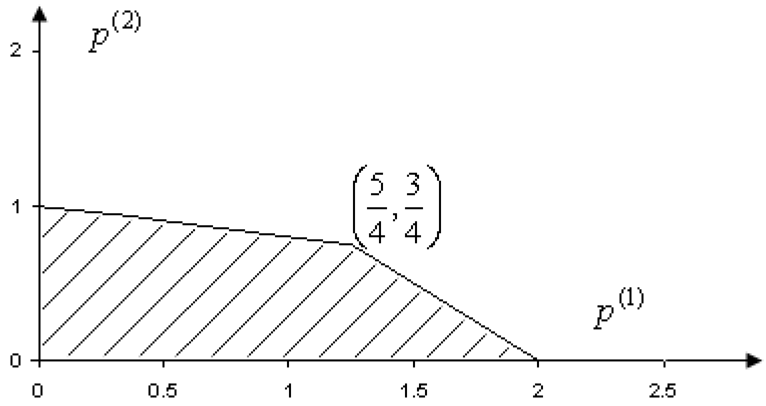


Рис. 4.3

Оптимальним розв'язком цієї задачі є вектор $p^{(0)} = \left(\frac{5}{4},\frac{3}{4}\right)^T$. На рис.4.3 зображена допустима область задачі.

Лінійний пошук. При початковій точці $x^{(0)}$ будь-яка точка в напрямку $p^{(0)}$ може бути представлена у вигляді $x^{(1)} = x^{(0)} + \alpha\ p^{(0)} = \left(\frac{5}{4}\alpha,\ \frac{3}{4}\alpha\right)$, а відповідне їй значення цільової функції дорівнює $f\left(x^{(0)} + \alpha\ p^{(0)}\right) = \frac{19}{8}\alpha^2 - \frac{19}{2}\alpha$.

Отже, довжину кроку α_0 знайдемо з розв'язання такої задачі одновимірної оптимізації:

$$\frac{19}{8}\alpha^2 - \frac{19}{2}\alpha \to min,\ 0 \le \alpha \le 1.$$

Оптимальний розв'язок дорівнює $\alpha_0 = 1$. Отже, $x^{(1)} = \left(\frac{5}{4},\frac{3}{4}\right)^T$.

Ітерація 2. **Пошук напрямку**. У точці $x^{(1)} = \left(\frac{5}{4},\frac{3}{4}\right)^T$ маємо

$$f'\left(x^{(1)}\right) = \left(-\frac{1}{2};-\frac{11}{2}\right)^T,\quad g_1\left(x^{(1)}\right) = 0,\quad g_2\left(x^{(1)}\right) = 0,\quad g_3\left(x^{(1)}\right) = -\frac{5}{4},\quad g_4\left(x^{(1)}\right) = -\frac{3}{4}.$$

Напрямок $p^{(1)}$ знайдемо з розв'язання такої задачі:

$$\frac{1}{2}\left(p_1^2 + p_2^2\right) - \frac{1}{2}p_1 - \frac{11}{2}p_2 \to min$$

при обмеженнях

$p_1 + p_2 \le 0$,

$p_1 + 5p_2 \le 0$,

$-p_1 - \frac{5}{4} \le 0$,

$-p_2 - \frac{3}{4} \le 0$.

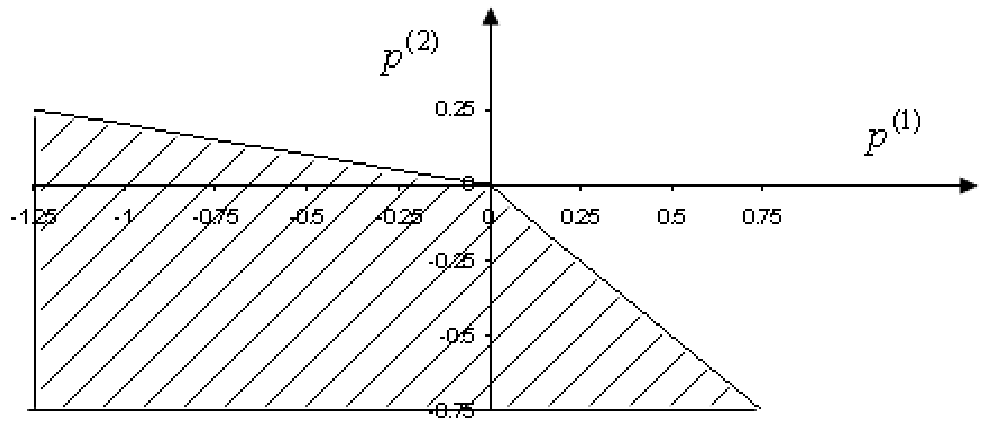


Рис.4.4

Допустима область задачі зображена на рис.4.4. Оптимальним розв'язком цієї задачі є вектор $p^{(1)} = \left(-\frac{15}{26}, \frac{3}{26}\right)^T$.

Лінійний пошук. Знайдемо положення точки $x^{(2)}$, рухаючись з точки $x^{(1)}$ у напрямку $p^{(1)}$. Враховуючи, що

$$x^{(2)} = x^{(1)} + \alpha\ p^{(1)} = \left(\frac{5}{4} - \frac{15}{26}\alpha,\ \frac{3}{4} + \frac{3}{26}\alpha\right),$$

маємо таку задачу для визначення довжини кроку α_1:

$$\frac{279}{338}\alpha^2 - \frac{9}{26}\alpha - \frac{57}{8} \to min,\ 0 \le \alpha \le 1.$$

Оптимальний розв'язок дорівнює $\alpha_1 = \frac{39}{186}$. Отже, $x^{(2)} = \left(\frac{35}{31}, \frac{24}{31}\right)^T$.

Ітерація 3. Пошук напрямку. У точці $x^{(2)} = \left(\frac{35}{31}, \frac{24}{31}\right)^T$ маємо

$$f'\left(x^{(2)}\right) = \left(-\frac{32}{31}; -\frac{160}{31}\right)^T,\ g_1\left(x^{(2)}\right) = -\frac{3}{31},\ g_2\left(x^{(2)}\right) = 0,\ g_3\left(x^{(2)}\right) = -\frac{35}{31},$$

$g_4\left(x^{(2)}\right) = -\frac{24}{31}$. Задача для знаходження напрямку $p^{(2)}$ має вигляд:

$$\frac{1}{2}\left(p_1^2+p_2^2\right)-\frac{32}{31}p_1-\frac{160}{31}p_2 \to min$$

при обмеженнях

$$p_1+p_2-\frac{3}{31}\le 0,$$

$$p_1+5p_2\le 0,$$

$$-p_1-\frac{35}{31}\le 0,$$

$$-p_2-\frac{24}{31}\le 0$$

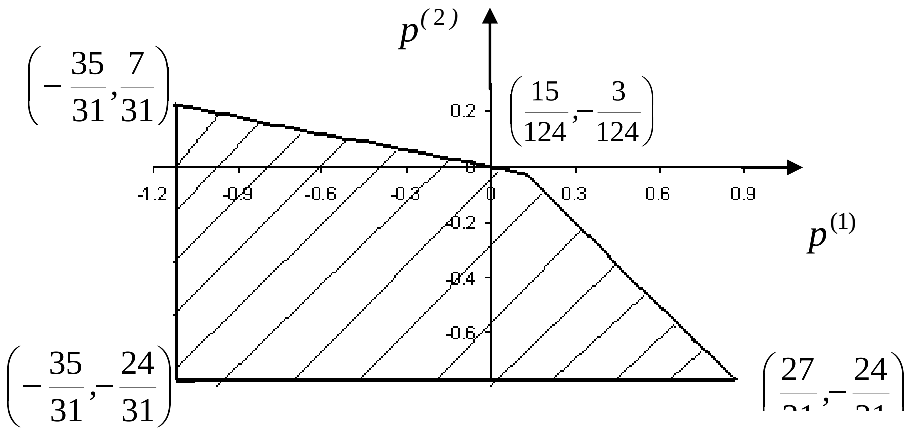


Рис.4.5

Оптимальним розв’язком допоміжної задачі є вектор $p^{(2)}=(0,0)^T$. Отже, точка $x^{(2)}=\left(\frac{35}{31},\frac{24}{31}\right)^T$ є оптимальним розв’язком задачі.

§4. Методи можливих напрямків

Цей клас методів розв’язання задачі нелінійного програмування базується на переміщенні з однієї допустимої точки в іншу допустиму точку з кращим значенням цільової функції. Типова стратегія в алгоритмах можливих напрямків така: береться допустима точка $x^{(k)}$ і знаходиться такий напрямок $p^{(k)}$, що для достатньо малих $\alpha>0$ виконуються такі дві вимоги:

1) точка $x^{(k)}+\alpha\, p^{(k)}$ – допустима;
2) значення цільової функції в точці $x^{(k)}+\alpha\, p^{(k)}$ краще, ніж у точці $x^{(k)}$.

Після знаходження такого напрямку розв’язується задача одновимірної мінімізації для того, щоб визначити як далеко потрібно рухатись уздовж напрямку $p^{(k)}$. Це приводить до нової точки $x^{(k+1)}$ і процес повторюється.

Нагадаємо визначення можливого напрямку, яке наведено у § 3 розділу 1. Вектор $p \in E^n$ задає можливий напрямок відносно множини X у точці $x \in X$, якщо $x + \alpha\, p \in X$ при всіх достатньо малих $\alpha > 0$. Вектор p називається **можливим напрямком спуску** в точці $x \in X$, якщо $f(x + \alpha\, p) < f(x)$ і $x + \alpha\, p \in X$ при всіх достатньо малих $\alpha > 0$.

1. Метод Зойтендейка. Випадок лінійних обмежень

Знайдемо розв’язок такої задачі:

$$
\begin{gathered}
f(x) \to min, \\
A\,x = b, \\
D\,x = e.
\end{gathered}
\tag{4.34}
$$

Тут A – матриця порядку $m \times n$, D – матриця порядку $l \times n$, b – вектор простору E^m, e – вектор простору E^l, цільова функція $f(x)$ є неперервно-диференційовною.

Нехай $A^T = \left[A_1^T, A_2^T\right]$ і $b^T = \left[b_1^T, b_2^T\right]$, де $A_1 x = b_1$, $A_2 x < b_2$. Вектор p є можливим напрямком спуску для задачі (4.34), якщо $A_1\, p \le 0$, $D\, p = 0$, $f'(x)^T \cdot p < 0$.

Побудова можливого напрямку спуску полягає в мінімізації $f'(x)^T \cdot p$ при умовах $A_1 p \le 0$ і $D\,p = 0$. Але, якщо існує вектор p_* такий, що $f'(x)^T \cdot p_* < 0$, $A_1 p_* \le 0$, $D\,p_* = 0$, то оптимальне значення цільової функції у сформульованій задачі дорівнює $-\infty$, тому що обмеженням цієї задачі задовольняє будь-який вектор $\lambda\, p_*$, де λ – як завгодно велике число. Тому, до задачи повинна бути додана умова, яка обмежує вектор p. Таке обмеження називають **нормуючим**. З урахуванням вищесказаного, задача для знаходження можливого напрямку спуску має вигляд

$$
\begin{gathered}
f'(x)^T \cdot p \to min, \\
A_1 p \le 0,
\end{gathered}
$$

$$D\,p = 0, \qquad (4.35)$$
$$-1 \le p_j \le 1, \quad j = \overline{1,n}.$$

Задача (4.35) є задачею лінійного програмування і може бути розв’язана симплекс-методом. Якщо оптимальне значення цільової функції в точці x дорівнює нулю, тоді точка x є точкою Куна-Такера. Якщо ж це значення від’ємне, то це означає, що побудовано можливий напрямок спуску.

Алгоритм методу Зойтендейка у випадку лінійних обмежень

Початковий етап. Задати допустиму точку $x^{(1)}$, для якої $Ax^{(1)} = b$, $Dx^{(1)} = e$. Покласти $k = 1$ та перейти до основного етапу.

Основний етап

Крок 1. Нехай $A_1 x^{(k)} = b_1$, $A_2 x^{(k)} < b_2$. Взяти за напрямок $p^{(k)}$ оптимальний розв’язок такої задачі:

$$f'\left(x^{(k)}\right)^T \cdot p \to min,$$
$$A_1 p \le 0,$$
$$D\,p = 0,$$
$$-1 \le p_j \le 1, \quad j = \overline{1,n}.$$

Якщо $f'\left(x^{(k)}\right)^T \cdot p^{(k)} = 0$, то зупинитися: точка $x^{(k)}$ є точкою Куна-Такера. У протилежному випадку перейти до кроку 2.

Крок 2. Покласти α_k рівним оптимальному розв’язку задачі лінійного пошуку

$$f\left(x^{(k)} + \alpha\, p^{(k)}\right) \to min, \quad 0 \le \alpha \le \alpha_{max},$$

де

$$\alpha_{max} = \begin{cases} \min\limits_{1 \le j \le n} \left(\dfrac{\bar{b}_i^{(k)}}{\bar{p}_i^{(k)}} : \ \bar{p}_i^{(k)} < 0 \right), & \text{якщо } \bar{p}^{(k)} \ge 0, \\ \infty, \quad \bar{p}^{(k)} \le 0. & \end{cases} \qquad (4.36)$$

$$\bar{b} = b_2 - A_2 x^{(k)}, \ \bar{p} = A_2 p^{(k)}.$$

Покласти $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, перевизначити матриці A_1 та A_2, замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо задачу

$$f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \to min$$

при обмеженнях

$$x_1 + x_2 \le 2,$$

$$x_1 + 5x_2 \le 5,$$

$$-x_1 \qquad \le 0,$$

$$-x_2 \le 0.$$

Розв'яжемо цю задачу методом Зойтендейка (випадок лінійних обмежень), узявши за початкову точку $x^{(1)} = (0,0)^T$. Відзначимо, що

$f'(x) = (4x_1 - 2x_2 - 4;\ -2x_1 + 4x_2 - 6;\ 0;\ 0)^T$.

Ітерація 1. Пошук напрямку. У точці $x^{(1)} = (0,0)^T$ маємо $f'\left(x^{(1)}\right) = (-4,-6)^T$. Активними обмеженнями в цій точці є тільки обмеження невід'ємності змінних, так що $I = \{3,4\}$. Запишемо матриці A_1 та A_2:

$$A_1 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},\ A_2 = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}.$$

Задача для знаходження напрямку має вигляд:

$$-4p_1 - 6p_2 \to min,$$

$$-p_1 \le 0,\ -p_2 \le 0,$$

$$-1 \le p_1 \le 1,\ -1 \le p_2 \le 1.$$

Цю задачу можна розв'язати симплекс-методом розв'язання задач лінійного програмування. Оптимальним розв'язком цієї задачі є вектор $p^{(1)} = (1,1)^T$, а оптимальне значення цільової функції дорівнює –10. На рис. 4.6 показано допустиму область цієї задачі.

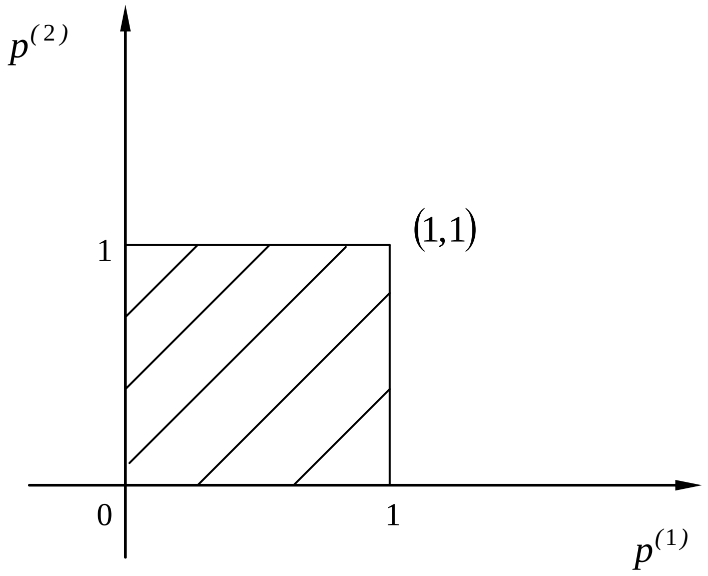


Рис. 4.6

Лінійний пошук. Рухаючись з точки $x^{(1)} = (0, 0, 2, 5)$ вздовж напрямку $p^{(1)} = (1,1)^T$, потрібно знайти точку, в якій значення цільової функції $f(x)$ є мінімальним. Будь-яка точка променя, який виходить з $x^{(1)}$ у напрямку $p^{(1)}$, може бути записана у вигляді $x^{(1)} + \alpha\, p^{(1)} = (\alpha, \alpha)^T$, а цільова функція в цій точці має вигляд $f\left(x^{(1)} + \alpha\, p^{(1)}\right) = -10\alpha + 2\,\alpha^2$. Максимальне значення кроку α, для якого точка $x^{(1)} + \alpha\, p^{(1)}$ є допустимою, обчислюється за формулою (4.36) і дорівнює $\alpha_{max} = min\left\{\frac{2}{10}, \frac{5}{34}\right\} = \frac{5}{34}$.

Значення α_1 отримаємо із розв'язання такої задачі одновимірного пошуку:

$$2\alpha^2 - 10\alpha \to min,\ 0 \le \alpha \le \frac{5}{6}.$$

Оптимальний розв'язок дорівнює $\alpha_1 = \frac{5}{6}$, так що $x^{(2)} = x^{(1)} + \alpha_1 p^{(1)} = \left(\frac{5}{6}, \frac{5}{6}\right)^T$.

Ітерація 2. Пошук напрямку. У точці $x^{(2)} = \left(\frac{5}{6}, \frac{5}{6}\right)^T$ маємо $f'\left(x^{(2)}\right) = \left(-\frac{7}{3}, -\frac{13}{3}\right)^T$. Активним обмеженням у цій точці є тільки друге обмеження, так що напрямок спуску знаходиться з розв'язання такої задачі:

$$-\frac{7}{3}p_1 - \frac{13}{3}p_2 \to min,$$

$$p_1 + 5p_2 \le 0,$$

$$-1 \le p_1 \le 1,\ -1 \le p_2 \le 1.$$

Оптимальним розв'язком цієї задачі лінійного програмування є точка $p^{(2)} = \left(1, -\frac{1}{5}\right)^T$ (рис. 4.7), а відповідне оптимальне значення цільової функції дорівнює $-\frac{22}{15}$.

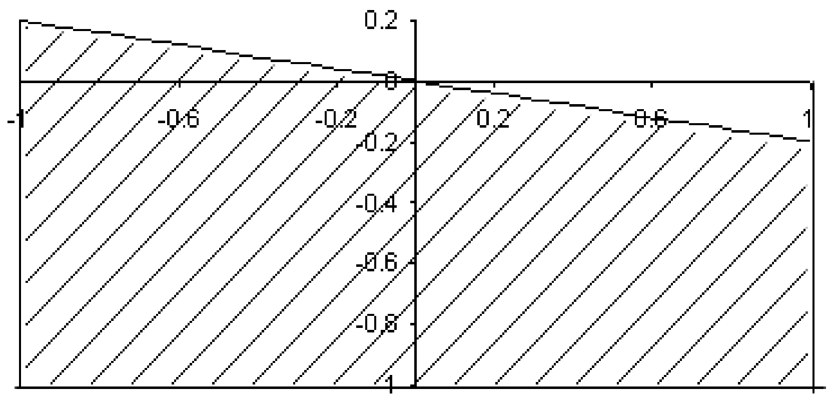


Рис. 4.7

Лінійний пошук. При початковій точці $x^{(2)}$ будь-яка точка вздовж напрямку може бути записана у вигляді $x^{(2)} + \alpha\ p^{(2)} = \left(\frac{5}{6} + \alpha,\ \frac{5}{6} - \frac{1}{5}\alpha\right)^T$, а відповідне значення цільової функції дорівнює $f\left(x^{(2)} + \alpha\ p^{(2)}\right) = -\frac{125}{8} - \frac{22}{15}\alpha + \frac{62}{25}\alpha^2$. Максимальне значення кроку α, для якого точка $x^{(2)}+\alpha\, p^{(2)}$ залишається припустимою, обчислюється за формулою (4.36) і дорівнює $\alpha_{max}=min\left\{\frac{5}{12},\frac{25}{6}\right\}=\frac{5}{12}$.

Таким чином, за α_2 береться оптимальний розв'язок такої задачі:

$$-\frac{125}{8}-\frac{22}{15}\alpha+\frac{62}{25}\alpha^2\to min\ \ 0\le\alpha\le\frac{5}{12}.$$

Оптимальний розв'язок цієї задачі дорівнює $\alpha_2=\frac{55}{186}$, так що $x^{(3)}=x^{(2)}+\alpha_2 p^{(2)}=\left(\frac{35}{31},\frac{24}{31}\right)^T$.

Ітерація 3. Пошук напрямку. У точці $x^{(3)}=\left(\frac{35}{31},\frac{24}{31}\right)^T$ маємо $f'\left(x^{(3)}\right)=\left(-\frac{32}{31},-\frac{160}{31}\right)^T$. Активним обмеженням у цій точці є тільки друге обмеження, так що напрямок спуску знаходиться з розв'язання такої задачі:

$$-\frac{32}{31}p_1-\frac{160}{31}p_2\to min,$$

$$p_1+5p_2\le 0,$$

$$-1\le p_1\le 1,\ -1\le p_2\le 1.$$

Оптимальним розв'язком цієї задачі лінійного програмування є точка $p^{(3)}=\left(1,-\frac{1}{5}\right)^T$, а відповідне оптимальне значення цільової функції дорівнює нулю, і процедура закінчується. Точка $x^{(3)}=\left(\frac{35}{31},\frac{24}{31}\right)^T$ є точкою Куна-Таккера і, враховуючи опуклість цільової функції $f(x)$, оптимальним розв'язком задачі.

2. Метод Зойтендейка. Випадок нелінійних обмежень-нерівностей

Розглянемо задачу, в якій допустима область задається системою обмежень-нерівностей не обов'язково лінійних:

$$f(x) \to min,\ g_i(x) \le 0,\ i = \overline{1,m}. \qquad (4.37)$$

Теорема. Нехай x – допустима точка задачі (4.37), а $I = \{\, i : g_i(x) = 0 \,\}$ – множина індексів обмежень, активних у точці x. Припустимо, що функції $f(x),\ g_i(x),\ i \in I$ неперервно-диференційовні в точці x, а функції $g_i(x),\ i \notin I$ неперервні в цій точці. Якщо $f'(x)^T \cdot p < 0$ та $g_i'(x)^T \cdot p < 0$ при $i \in I$, то вектор p є можливим напрямком спуску.

Для того, щоб знайти вектор p, який задовольняє нерівностям $f'(x)^T \cdot p < 0$ і $g_i'(x)^T \cdot p < 0$ при $i \in I$, природно мінімізувати максимум $f'(x)^T \cdot p$ та $g_i'(x)^T \cdot p$ для $i \in I$. Позначимо цей максимум через z. Вводячи нормуючі обмеження для кожного j, отримаємо таку задачу для знаходження напрямку:

$$z \to min,$$

$$f'(x)^T \cdot p - z \le 0, \qquad (4.38)$$

$$g_i'(x)^T \cdot p - z \le 0,\ i \in I,\ -1 \le p_j \le 1, \quad j = \overline{1,n}.$$

Нехай $(\bar{z}, \bar{p})$ – оптимальний розв’язок цієї задачі лінійного програмування. Якщо $\bar{z} < 0$, то $\bar{p}$ є можливим напрямком спуску. Якщо ж $\bar{z} = 0$, то x є стаціонарною точкою задачі (4.37). Якщо функції $f(x),\ g_i(x)$ опуклі на E^n, а обмеження $g_i(x) \le 0,\ i \in I$ є регулярними, то точка x є точкою мінімуму задачі (4.37).

Алгоритм методу Зойтендейка у випадку нелінійних обмежень

Початковий етап. Вибрати початкову допустиму точку $x^{(1)}$. Покласти $k = 1$ та перейти до основного етапу.

Основний етап

Крок 1. Покласти $I = \{ i : g_i(x^{(k)}) = 0 \}$ і розв’язати таку задачу:

$$z \to min,$$

$$f'\left(x^{(k)}\right)^T \cdot p - z \le 0,$$

$$g_i'\left(x^{(k)}\right)^T \cdot p - z \le 0,\ i \in I,$$

$$-1 \le p_j \le 1, \quad j = \overline{1,n}.$$

Нехай $\left(z^{(k)}, p^{(k)}\right)$ – оптимальний розв'язок. Якщо $z^{(k)} = 0$, то зупинитися: $x^{(k)}$ є стаціонарною точкою. Якщо $z^{(k)} < 0$, то перейти до кроку 2.

Крок 2. Взяти за α_k оптимальний розв'язок задачі одновимірної мінімізації:

$$f\left(x^{(k)} + \alpha\ p^{(k)}\right) \to min, \quad 0 \le \alpha \le \alpha_{max},$$

де $\alpha_{max} = sup\left\{\alpha : g_i\left(x^{(k)} + \alpha\ p^{(k)}\right) \le 0, i = \overline{1,m}\right\}$.

Покласти $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Збіжність методу Зойтендейка в загальному випадку не гарантується. Складність полягає в тому, що кроки уздовж напрямків, які генеруються і не виводять за границі допустимої області, є такими, що прямують до нуля, створюючи «заїдання» процесу в неоптимальній точці.

3. Модифікація Топкіса-Вейнотта алгоритму методу можливих напрямків

У цьому методі при визначенні можливого напрямку спуску враховуються всі обмеження задачі (4.37), а не тільки активні обмеження. На відмінність від методу можливих напрямків, який було описано вище, знайдений напрямок не має різкого розриву, коли обмеження перестає бути активним або коли стає активним нове обмеження.

Алгоритм методу можливих напрямків Топкіса-Вейнотта

Початковий етап. Вибрати початкову допустиму точку $x^{(1)}$. Покласти $k = 1$ та перейти до основного етапу.

Основний етап

Крок 1. Покласти $\left(z^{(k)}, p^{(k)}\right)$ оптимальному розв'язку такої задачі лінійного програмування:

$$z \to min,$$
$$f'\left(x^{(k)}\right)^T \cdot p - z \le 0, \qquad (4.39)$$
$$g_i'\left(x^{(k)}\right)^T \cdot p - z \le 0,\ i \in I,$$
$$-1 \le p_j \le 1, \quad j = \overline{1,n}.$$

Якщо $z^{(k)} = 0$, то зупинитися: $x^{(k)}$ є стаціонарною точкою. Якщо $z^{(k)} < 0$, то перейти до кроку 2.

Крок 2. Взяти за α_k оптимальний розв'язок задачі одновимірної мінімізації:

$$f\left(x^{(k)} + \alpha\ p^{(k)}\right) \to min, \quad 0 \le \alpha \le \alpha_{max},$$

де

$$\alpha_{max} = sup\left\{\alpha : g_i\left(x^{(k)} + \alpha\ p^{(k)}\right) \le 0, i = \overline{1,m}\right\}.$$

Покласти $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Теорема про збіжність. Нехай функції $f(x),\ g_i(x),\ i = \overline{1,m}$ є неперервно-диференційованими. Тоді будь-яка гранична точка послідовності $\left\{x^{(k)}\right\}$, яка одержана за алгоритмом Топкіса-Вейнотта для розв'язання задачі (4.37), є точкою Куна-Танкера.

Приклад. Розглянемо задачу

$$f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \to min$$

при обмеженнях

$$2x_1^2 - x_2 \le 0,$$
$$x_1 + 5x_2 \le 5,$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Виконаємо п'ять ітерацій алгоритму Топкіса і Вейнота, починаючи з точки $x^{(1)} = (0,\ 0.75)^T$. Відзначимо, що градієнт цільової функції дорівнює $f'(x) = (4x_1 - 2x_2 - 4;\ -2x_1 + 4x_2 - 6;\ 0;\ 0)^T$, а градієнти обмежень відповідно дорівнюють

$$g_1'(x)=(4x_1;-1)^T,\ g_2'(x)=(1;5)^T,\ g_3'(x)=(-1;0)^T,\ g_4'(x)=(0;-1)^T.$$

Усі ці градієнти використовуються в задачі пошуку напрямку на кожній ітерації методу.

Ітерація 1. Пошук напрямку. У точці $x^{(1)}=(0;0.75)^T$ маємо $f'\left(x^{(1)}\right)=(-5.5;-3)^T$. Задача для знаходження напрямку має вигляд:

$$z \to min,$$

$$-5.5p_1-3p_2-z\le 0,$$

$$p_1+5p_2-z\le 1.25,$$

$$-p_1-z\le 0,$$

$$-p_2-z\le 0.75,$$

$$-1\le p_1\le 1,\ -1\le p_2\le 1.$$

Оптимальним розв'язком цієї задачі є вектор $p^{(1)}=(0.7143;-0.0357)^T$, а $z_1=-0.7143$.

Лінійний пошук. Максимальне значення кроку α, для якого точка $x^{(1)}+\alpha\ p^{(1)}$ є допустимою дорівнює $\alpha_{max}=0.84$. Значення $\alpha_1=0.84$ є розв'язком задачі одновимірного пошуку:

$$0.972\alpha^2-4.036\alpha-3.375\to min,\ 0\le\alpha\le 0.84.$$

Таким чином, $x^{(2)}=x^{(1)}+\alpha_1 p^{(1)}=(0.6;0.72)^T$.

Ітерація 2. Пошук напрямку. У точці $x^{(2)}=(0.6;0.72)^T$ маємо $f'\left(x^{(2)}\right)=(-3.03;-4.32)^T$. Напрямок спуску $p^{(2)}$ знаходиться з розв'язання такої задачі:

$$z \to min,$$

$$-3.04p_1-4.31p_2-z\le 0,$$

$$p_1+5p_2-z\le 0.8,$$

$$2.4p_1-p_2-z\le 0$$

$$-p_1-z\le 0.6,$$

$$-p_2 - z \le 0.72,$$
$$-1 \le p_1 \le 1,\ -1 \le p_2 \le 1.$$

Оптимальним розв’язком цієї задачі лінійного програмування є вектор $p^{(2)} = (-0.07123, 0.1167)^T$ і $z^{(2)} = -0.2877$.

Лінійний пошук. Максимальне значення кроку α, для якого точка $x^{(2)} + \alpha\, p^{(2)}$ залишається допустимою дорівнює $\alpha_{max} = 1.5617$,

$$f\left(x^{(2)} + \alpha\, p^{(2)}\right) = 0.054\alpha^2 - 0.2876\alpha - 5.8272.$$

Значення $\alpha_2 = 1.5617$ і $x^{(3)} = x^{(2)} + \alpha_2 p^{(2)} = (0.4888, 0.9022)^T$.

Далі цей процес повторюється.

Таблиця 4.4

k	$x^{(k)}$	$f(x^{(k)})$	Пошук напрямку			Лінійний пошук		
			$f'(x^{(k)})$	$p^{(k)}$	$z^{(k)}$	α_{max}	α_k	$x^{(k+1)}$
1	$\begin{pmatrix}0.0\\0.75\end{pmatrix}$	–3.375	$\begin{pmatrix}-5.5\\-3.0\end{pmatrix}$	$\begin{pmatrix}0.714\\-0.0357\end{pmatrix}$	–0.714	0.84	0.84	$\begin{pmatrix}0.6\\0.72\end{pmatrix}$
2	$\begin{pmatrix}0.6\\0.72\end{pmatrix}$	–5.827	$\begin{pmatrix}-3.04\\-4.32\end{pmatrix}$	$\begin{pmatrix}-0.0712\\-0.117\end{pmatrix}$	–0.288	1.562	1.562	$\begin{pmatrix}0.489\\0.902\end{pmatrix}$
3	$\begin{pmatrix}0.489\\0.902\end{pmatrix}$	–6.145	$\begin{pmatrix}-3.849\\-3.369\end{pmatrix}$	$\begin{pmatrix}0.0957\\-0.0555\end{pmatrix}$	–0.182	1.564	1.564	$\begin{pmatrix}0.6385\\0.8154\end{pmatrix}$
4	$\begin{pmatrix}0.6385\\0.8154\end{pmatrix}$	–6.343	$\begin{pmatrix}-5.63\\-4.02\end{pmatrix}$	$\begin{pmatrix}-0.016\\-0.0433\end{pmatrix}$	–0.084	1.419	1.419	$\begin{pmatrix}0.616\\0.877\end{pmatrix}$
5	$\begin{pmatrix}0.616\\0.877\end{pmatrix}$	–6.508	$\begin{pmatrix}-3.29\\-3.725\end{pmatrix}$	$\begin{pmatrix}0.0268\\-0.0132\end{pmatrix}$	–0.030	1.455	1.455	$\begin{pmatrix}0.655\\0.858\end{pmatrix}$

Таблиця 4.4 наводить результати обчислень на п'яти ітераціях. Зауважимо, що в оптимальній точці $(0.658872, 0.868226)^T$ значення цільової функції дорівнює -6.613086.

§5. Методи відокремлюючої гіперплощини

Основна ідея методу полягає в тому, що допустима область апроксимується деяким багатогранником, який зменшується від одного ітераційного кроку до наступного, при цьому все краще апроксимуючи допустиму область в околі розв'язку. Метод відокремлюючої гіперплощини (або **метод Келлі січних площин**) може застосовуватися до загальних опуклих задач з нелінійною цільовою функцією та нелінійними обмеженнями:

$$f(x) \to min, \tag{4.40}$$
$$g_i(x) \le 0, \quad i = \overline{1,m}.$$

Тут $f(x)$ і $g_i(x)$ є опуклими та неперервно-диференційовними функціями.

Введемо додаткову змінну y і перепишемо задачу (4.40) у вигляді:

$$y \to min,$$
$$f(x) - y \le 0, \tag{4.41}$$
$$g_i(x) \le 0, \quad i = \overline{1,m}.$$

Задача (4.41) є опуклою задачею з лінійною цільовою функцією та нелінійними обмеженнями у просторі E^{n+1}

Оскільки показано, як здійснити перехід від задачі (4.40) з нелінійною цільовою функцією до задачі (4.41) з лінійною цільовою функцією, подальше викладення проведемо для такої опуклої задачі:

$$(c, x) \to min,$$
$$g_i(x) \le 0, \quad i = \overline{1,m}, \tag{4.42}$$

де $g_i(x)$ є опуклими та неперервно-диференційовними функціями, $x \in E^n$.

Нехай множина $X = \left\{x \in E^n : g_i(x) \le 0, \quad i = \overline{1,m}\right\}$ є допустимою множиною розв'язків задачі (4.42). Апроксимуємо множину X багатогранною множиною

Z. У випадку, коли X – обмежена, замкнена множина, то Z – багатогранник. Ідея методу полягає у відтинанні на кожному кроці частини багатогранної множини, яка не містить точку мінімуму початкової задачі.

Алгоритм методу

Початковий етап. Апроксимувати множину X багатогранною множиною Z_1. Покласти $k=1$, задати $\varepsilon > 0$ та перейти до основного етапу.

Основний етап

Крок 1. Розв'язати задачу лінійного програмування:

$$(c, x) \to min,,$$

$$x \in Z_k. \tag{4.43}$$

Нехай точка $x^{(k)}$ є оптимальним розв'язком задачі (4.43). Якщо $g_i(x) \le 0$ для всіх $i=\overline{1,m}$, то зупинитися: $x^{(k)}$ є оптимальним розв'язком початкової задачі. У протилежному випадку перейти до кроку 2.

Крок 2. Нехай $g_j\left(x^{(k)}\right) = \max\limits_{1 \le i \le m} g_i\left(x^{(k)}\right)$. Додати до обмежень, які визначають множину Z_k, додаткове обмеження вигляду

$$g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) \le 0.$$

Таким чином, множина Z_{k+1} має вигляд

$$Z_{k+1} = Z_k \cap \left\{ g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) \le 0 \right\}.$$

Замінити k на $k+1$та, якщо перейти до кроку 1.

Крок 3. Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon$, то зупинитися: $x^{(k)}$ – наближення до точки мінімуму. У протилежному випадку перейти до кроку 1.

Алгоритм описаний.

Гіперплощина

$$g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) = 0$$

строго відокремлює точку $x^{(k)}$ від множини X і на цій підставі вона називається **січною гіперплощиною**.

Збіжність методу відокремлюючої гіперплощини стверджує така теорема.

Теорема. Нехай $g_i(x)$, $i=\overline{1,m}$ є опуклими та неперервно-диференційовними функціями. Якщо задача (4.42) має мінімум на скінченній відстані, то будь-яка точка згущення послідовності $\{x^{(k)}\}$, що генерується методом відокремлюючої гіперплощини, є оптимальним розв'язком задачі (4.42).

Приклад. Розглянемо таку задачу:

$$f(x)=x^2 \to min, \quad -1 \le x \le 2\,.$$

Введемо додаткову змінну y і перейдемо до задачі з лінійною цільовою функцією:

$$\begin{aligned} &y \to min\,, \\ &x^2 - y \le 0\,, \\ &-1-x \le 0\,, \\ &x-2 \le 0\,. \end{aligned} \tag{4.44}$$

Від одновимірної початкової задачі перейшли до двовимірної задачі.

Апроксимуємо допустиму множину многокутником Z_1 (у даному випадку трикутником ABC): $A(-1{,}1)$, $B(2{,}4)$, $C(0.5,-2)$. Отримаємо задачу:

$$\begin{aligned} &y \to min\,, \\ &y-x-2 \le 0\,, \\ &4x-y-4 \le 0\,, \\ &-2x-y-1 \le 0\,. \end{aligned} \tag{4.45}$$

Оптимальним розв'язком цієї задачі буде точка $x^{(1)}=(0.5,2)$. Перевіримо, чи задовольняє ця точка обмеженням задачі (4.44):

$$g_1\left(x^{(1)}\right)=\left[x^{(1)}\right]^2 - y^{(1)} = \frac{1}{4}+2>0\,,$$

$$g_2\left(x^{(1)}\right)=-1-\frac{1}{2}<0\,,$$

$$g_3\left(x^{(1)}\right)=\frac{1}{2}-2<0\,.$$

Добавляємо до обмежень задачі (4.45) нове обмеження

$$g_1\left(x^{(1)}\right)+g_1'\left(x^{(1)}\right)\left(x-x^{(1)}\right)\le 0 \text{ або } x-y-0.25\le 0.$$

Новий многокутник Z_2 буде чотирикутником з вершинами: $A(-1,1)$, $B(2,4)$, $D(-0.25,-0.5)$, $E(1.25,1)$. Оптимальним розв'язком задачі

$$y\to min,$$

$$y-x-2\le 0,$$

$$4x-y-4\le 0,$$

$$-2x-y-1\le 0,$$

$$x-y-0.25\le 0.$$

буде точка $x^{(2)}=(-0.25,-0.5)$. Не виконується перша умова задачі (4.44). Тобто, $j=1$. До обмежень задачі лінійного програмування добавляємо обмеження:

$$-0.5x-y-0.0625\le 0.$$

Отримуємо п'ятикутник Z_3 з вершинами: $A(-1,1)$, $B(2,4)$, $E(1.25,1)$, $F(-0.625,-0.25)$, $G(0.125,-0.125)$. Перевіряємо критерій закінчення: $\left\|x^{(2)}-x^{(1)}\right\|>\varepsilon$. Оптимальним розв'язком задачі

$$y\to min,$$

$$y-x-2\le 0,$$

$$4x-y-4\le 0,$$

$$-2x-y-1\le 0,$$

$$x-y-0.25\le 0,$$

$$-0.5x-y-0.0625\le 0$$

буде точка $x^{(3)}=(0.125,-0.125)$. Не виконується знову перша умова задачі (4.44). Тобто, $j=1$. Додамо обмеження

$$-0.25x-y-0.015625\le 0.$$

Отримуємо новий многокутник Z_4. Обчислення закінчуються, якщо виконана умова зупинки.

§6 Методи штрафних функцій

Здійснення мінімізації без обмежень є більш простою задачею, ніж мінімізація з обмеженнями. Тому природною є спроба зведення задачі з обмеженнями до задачі без обмежень.

Методи штрафних функцій дозволяють звести задачу нелінійного програмування

$$f(x) \to min, \qquad (4.46)$$

$$g_i(x) \le 0,\ i = \overline{1,m}$$

до задачі або послідовності задач безумовної мінімізації певних допоміжних штрафних функцій. При цьому допоміжні функції підбираються так, що

1) на більшій частині допустимої множини задачі ці функції є близькими до нуля;
2) кожна з них достатньо швидко збільшується або при наближенні зсередини до границі допустимої множини (**внутрішні** або **бар'єрні штрафні функції**), або при виході за його границі (**зовнішні штрафні функції**);
3) ступінь близькості штрафу до нуля і швидкість його зростання залежать від значення штрафного параметру і збільшується із зростанням параметра.

У комбінованих методах штрафних функцій, які потрібно використовувати при обмеженнях–рівностях, у процесі мінімізації частина обмежень задовольняється, а частина – ні. Але при досягненні шуканого розв'язку всі умови в границях заданої точності задовольняються.

1. Метод зовнішньої точки

Основна ідея методу зовнішньої точки полягає в такому перетворенні цільової функції $f(x)$, при якому значення перетвореної цільової функції в допустимій області точно або наближено дорівнюють значенням функції $f(x)$, у той час коли значення зовні допустимої області $X = \left\{ x \in E^n : g_i(x) \le 0, i = \overline{1,m} \right\}$ є дуже великими у порівнянні із значеннями функції $f(x)$. При виконанні цих умов мінімум вихідної функції $f(x)$ не буде істотно відрізнятися від мінімуму перетвореної функції.

Послідовність функцій $\{P(x, r_k)\}$, $k = 1, 2, \ldots$, які визначені і невід'ємні на множині X, називається **штрафною функцією** множини X (зовнішньою штрафною функцією), якщо

$$\lim_{k \to \infty} P(x, r_k) = \begin{cases} 0, & \text{якщо } x \in X, \\ +\infty, & \text{якщо } x \notin X. \end{cases} \tag{4.47}$$

Метод штрафних функцій розв'язання задачі (4.46) полягає в заміні цієї задачі послідовністю задач безумовної мінімізації функцій

$$f_k(x) = f(x) + P(x, r_k).$$

При побудові допоміжних функцій бажано забезпечити потрібну гладкість, опуклість, зручність обчислення значень функції та потрібних похідних. Якщо функції $g_i(x)$ не є опуклими, то штрафна функція також не буде опуклою на множині X. Тоді вона може мати локальні мінімуми, в той час, як передбачається, що знаходиться глобальний мінімум задачі (4.48). Це порушує збіжність і є істотним недоліком методу штрафних функцій при застосуванні його до неопуклих задач. У випадку, коли розглядається задача опуклого програмування, штрафна функція також буде опуклою. Але в цьому випадку виникає інша складність. Для отримання доброго наближення потрібно параметр r_k вибрати достатньо великим. При цьому всі похідні від $P(x, r_k)$ по змінній x також будуть великими, тому що вони пропорційні r_k. Але встановлено, що розмір околу, в якому збіжність надлінійна, є обернено пропорційним константі Ліпшиця других похідних, тобто в розглядуваному випадку цей окіл буде також малим, і навіть теоретично добре збіжний метод може стати неефективним.

Задача (4.48) розв'язується одним з методів безумовної оптимізації, які описані в розділі 3. Якщо функція $f_k(x)$ є неперервно-диференційовною і обчислення її похідних не викликає істотних труднощів, то для безумовної мінімізації можна застосовувати методи першого порядку. Якщо функція $f_k(x)$ є двічі неперервно-диференційовною, то можна застосувати метод Ньютона. Негладкі функції штрафу призводять до необхідності мінімізації негладкої функції $f_k(x)$. У цьому випадку можна застосовувати методи нульового порядку (покоординатного спуску, конфігурацій, Розенброка та інші).

Теорема про збіжність. Нехай цільова функція $f(x)$ є неперервною, а штрафна функція $P(x,r_k)$ має такі властивості:

1) $P(x,r_k)$ неперервна по змінним x та r_k і $P(x,r_k)=0$ при $x \in X$;
2) $P(x,r_k)$ монотонно зростає з зростанням r_k

Нехай, окрім того, для довільної сталої C множина $X_C(k)=\{x \in E^n : f_k(x) \le C\}$ є обмеженою. Тоді:

1) функція $f_k(x)$ досягає на E^n свого мінімуму f_k^* в деякій точці x_k^*, при цьому $f_k^* \le f_*$, де $f_* = \min\limits_{x \in X} f(x)$ і $f_k^* \to f_*$;
2) будь-яка гранична точка x_* послідовності $\{x_{r_k}^*\}$ при $r_k \to \infty$, $k \to \infty$ є точкою мінімуму x_* функції $f(x)$ множині X, вся послідовність $\{x_{r_k}^*\}$ збігається до x_*.

Штрафну функцію $P(x,r_k)$ можна будувати різними способами. Наведемо ряд прикладів вибору $P(x,r_k)$ для задачі (4.46):

$$P(x,r_k)=r_k\sum_{i=1}^{m}[\max\{g_i(x),0\}]^q,\ q>0,$$

$$P(x,r_k)=\frac{1}{r_k}\exp\left[r_k\sum_{i=1}^{m}\max\{g_i(x),0\}\right],$$

$$P(x,r_k)=r_k\max_{1\le i\le m}\{\max\{g_i(x),0\}\}.$$

Якщо задача (4.46) містить обмеження-рівності, то штрафну функцію можна вибрати у вигляді:

$$P(x,r_k)=r_k\sum_{i=1}^{m}(g_i(x))^p \text{ , де } p\ge 1 \text{ – фіксоване число.}$$

Алгоритм методу зовнішньої точки

Нехай функції $f(x),\ g_i(x),\ i=\overline{1,m}$ є неперервними.

Початковий етап. Задати початкову точку $x^{(1)}\in E^n$, штрафний параметр $r_1>0$ і число $\beta>0$. Покласти $k=1$ та перейти до основного етапу. Вибрати $\varepsilon>0$ як константу зупинки.

Основний етап

Крок 1. При початковій точці $x^{(k)}$ розв'язати таку задачу безумовної оптимізації:

$$f_k(x)=f(x)+P(x,r_k)\to min.$$

Покласти $x^{(k+1)}$ таким, що дорівнює оптимальному розв'язку цієї задачі і перейти до кроку 2.

Крок 2. Якщо $P(x,r_k)\le\varepsilon$, то зупинитися, у протилежному випадку покласти $r_{k+1}=\beta r_k$, замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Приклад. Розглянемо таку задачу:

$$f(x)=(x_1-2)^4+(x_1-2x_2)^2\to min$$

за умови

$$x_1^{\ 2}-x_2=0,\ x=(x_1,x_2)\in E^2.$$

Згідно методу зовнішньої точки на k-тій ітерації при заданому значенні параметра штрафу r_k буде розв'язуватися така задача:

$$(x_1-2)^4+(x_1-2x_2)^2+r_k\left(x_1^2-x_2\right)^2\to min$$

У табл. 4.4 наведені результати обчислень за методом зовнішньої точки з початкової точки $x^{(1)}=(2.00,1.00)$, число $\beta=10$.

Таблиця 4.5

k	r_k	$x^{(k+1)}$	$f\left(x^{(k+1)}\right)$	$r_k\ P\left(x^{(k+1)},r_k\right)$
1	0,1	(1.4539,0.7608)	0.0935	0.1831
2	1,0	(1.1687,0.7407)	0.5753	0.3908
3	10,0	(0.9906,0.8425)	1.5203	0.1926
4	100,0	(0.9507,0.8875)	1.8917	0.0267
5	1000,0	(0.9461,0.8934)	1.9405	0.0028

2. Метод внутрішньої точки (метод бар'єрів)

Головна незручність попереднього методу полягає в тому, що точка мінімуму x_* функції $f(x)$ множині X апроксимується зовні, тобто різні проміжні значення $x_1^*,\ldots,\ x_k^*$, які отримано при коефіцієнтах штрафу $r_1,\ldots,\ r_k$, не належать допустимій множині X. Це і привело до необхідності розробки інших методів штрафу, в яких оптимум апроксимується зсередини.

Функцію $B(x,r_k)$ назвемо **функцією бар'єру**, якщо

1) $B(x,r_k)$ визначена і невід'ємна в усіх внутрішніх точках допустимої множини;

2) $B(x,r_k)\to+\infty$, якщо x прямує до границі допустимої множини $x^{(1)}\in E^n$;

при неперервних функціях $g_i(x)\le 0$, $i=\overline{1,m}$ функція $B(\,x,r_k\,)$ неперервна у внутрішніх точках множини X.

Прикладами функцій бар'єру є такі функції:

$$B(x,r_k)=\frac{1}{r_k}\sum_{i=1}^{m}\frac{1}{-g_i(x)},$$

$$B(x,r_k)=\frac{1}{r_k}\sum_{i=1}^{m}\frac{1}{g_i^2(x)},$$

$$B(x,r_k)=-\frac{1}{r_k}\sum_{i=1}^{m}ln[-g_i(x)].$$

На відміну від методу зовнішньої точки за початкову точку $x^{(0)}$ вибирається **допустима** точка, тобто $x^{(0)} \in X$. Мінімум функції $f_k(x)$ для методу внутрішньої точки досягається зсередини допустимої області, але не лежить на її границі при усіх $r_k > 0$. Істотною відмінністю від методу зовнішньої точки є вимога існування внутрішніх точок допустимої області.

Алгоритм методу внутрішньої точки співпадає з алгоритмом методу зовнішньої за винятком того, що в алгоритмі методу внутрішньої точки $x^{(0)} \in X$.

Пошук початкового наближення $x^{(0)}$, тобто точки $x^{(0)} \in X$, такої що $g_i\left(x^{(0)}\right) < 0$, можна здійснити за таким алгоритмом.

Алгоритм вибору початкової точки в методі внутрішньої точки

Початковий етап. Задати точку $x^{(1)}$, покласти $k = 1$ та перейти до основного етапу.

Основний етап

Крок 1. Покласти $I = \left\{ i : g_i\left(x^{(k)}\right) < 0 \right\}$, Якщо $I = \{ 1,...,m \}$, то зупинитися: точка $x^{(k)}$ є шуканим початковим наближенням. У протилежному випадку вибрати $j \notin I$ і перейти до кроку 2.

Крок 2. Використати метод бар'єрів для розв'язання такої задачі при початковій точці $x^{(k)}$:

$$g_j(x) \to min, \qquad (4.49)$$

$$g_i(x) < 0, \ i \in I .$$

Покласти $x^{(k+1)}$ таким, що дорівнює оптимальному розв'язку задачі (4.49). Якщо $g_j\left(x^{(k+1)}\right) \geq 0$, то зупинитися, тому що множина $\left\{ x \in E^n : g_i(x) < 0, i = \overline{1,m} \right\}$ є порожньою. У протилежному випадку замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Алгоритм зупиниться не більше, ніж через m ітерацій або знайшовши допустиму початкову точку, яка задовольняє нерівностям $g_i\left(x^{(0)}\right)<0$, або встановивши факт відсутності таких точок.

2. Метод послідовної безумовної оптимізації

(метод Фіако і Мак-Кормика)

Розглянемо задачу:

$$f(x)\to min, \tag{4.50}$$

$$g_i(x)\le 0,\ i=\overline{1,p},$$

$$h_i(x)=0,\ i=\overline{p+1,m}.$$

Тут $h_i(x),\ i=\overline{p+1,m}$ –лінійні функції.

Розширеними штрафними функціями в комбінованих методах можуть бути такі функції:

$$f_k(x)=f(x)+\frac{1}{r_k}\sum_{i=1}^{p}\frac{1}{-g_i(x)}+\sqrt{r_k}\sum_{i=p+1}^{m}h_i^2(x),$$

$$f_k(x)=f(x)-\frac{1}{r_k}\sum_{i=1}^{p}ln\left[-g_i(x)\right]+\sqrt{r_k}\sum_{i=p+1}^{m}h_i^2(x).$$

Процедура мінімізації функції $f_k(x)$ починається з внутрішньої початкової точки $x^{(0)}$, в якій задовольняються всі обмеження у вигляді нерівностей.

Швидкість збіжності методу залежить від початкового вибору параметра r_0 та способу зміни параметра r_k. Існують різні способи вибору початкового параметра r_0. Наприклад,

1. $r_0=1$.
2. $r_0=\dfrac{f'\left(x^{(0)}\right)^T\cdot R'\left(x^{(0)}\right)}{\left\|R'\left(x^{(0)}\right)\right\|^2}$,

   де $R\left(x^{(0)}\right)=\sum_{i=1}^{p}\dfrac{1}{-g_i\left(x^{(0)}\right)}$.

3. $r_0 = \left[\dfrac{f'\left(x^{(0)}\right)^T \cdot \left[R''\left(x^{(0)}\right)\right]^{-1} f'\left(x^{(0)}\right)}{R'\left(x^{(0)}\right)^T \left[R''\left(x^{(0)}\right)\right]^{-1} R'\left(x^{(0)}\right)} \right]^{-1}$.

Що стосується подальшого вибору значень параметра r_k, то як показала практика застосування методів штрафних функцій, методи є досить ефективними, коли послідовність r_1, r_2, …, r_k визначається простим співвідношенням $r_{k+1} = \beta r_k$, де $\beta > 1$. На практиці часто $\beta = 4$.

Умови збіжності методу послідовної безумовної мінімізації

Нехай функції $f(x)$, $g_i(x)$, $i = \overline{1,p}$, $h_i(x)$, $i = \overline{p+1,m}$ є двічі неперервно-диференційовними функціями і виконані такі умови:

1) множина X має внутрішні точки;
2) для будь-якого скінченного C і будь-якого $r_k > 0$ множина точок $X_C(k) = \left\{ x \in E^n : g_i(x) \le 0, i = \overline{1,p},\ f(x) + \sqrt{r_k} \sum\limits_{i=p+1}^{m} h_i^2(x) \le C \right\}$ є обмеженою;
3) функція $f(x)$ є опуклою і сума $\sum\limits_{i=p+1}^{m} h_i^2(x)$ також є опуклою;
4) функції $g_i(x)$, $i = \overline{1,m}$ є опуклими;
5) матриця Гессе розширеної функції $f_k(x)$ не обертається в нуль для жодної точки, яка належить множині $X_1 = \left\{ x \in E^n : g_i(x) \le 0, i = \overline{1,p} \right\}$.

Теорема. Якщо задача нелінійного програмування задовольняє умовам 1–5, то

1) кожна розширена функція $f_k(x)$ має мінімум у деякій точці x_k^* множини X;
2) якщо $\{r_k\}$ – строго зростаюча послідовність, тоді

$$\lim_{k \to \infty} f_k\left(x_k^*\right) = \min_{x \in X} f(x) = f(x_*) = f_*,$$

тобто послідовність безумовних мінімумів функції $f_k(x)$ збігається до умовного мінімуму функції $f(x)$.

Критерієм закінчення ітераційного процесу може бути такий критерій:

$$\frac{1}{r_k}\sum_{i=1}^{p}\frac{1}{g_i\left(x^{(k)}\right)}\le\varepsilon,$$

де ε – задана константа.

Вправи до розділу 4

1. Розв'язати в явному вигляді задачу (4.21), якщо допустима множина X має вигляд:
   - $X=\left\{x\in E^n : \|x-c\|\le R\right\}$ – куля;
   - $X=\left\{x\in E^n : b_j\le x_j\le c_j,\ j=\overline{1,n}\right\}$ – n – вимірний паралелепіпед;
   - $X=\left\{x\in E^n : x_j\ge 0,\ j=\overline{1,n}\right\}$ – невід'ємний октант.
2. Отримати з (4.23) явний вираз для крокового множника α_k, якщо $f(x)=\frac{1}{2}(A\,x,x)-(b,x)$, де A – додатно визначена, симетрична матриця розміру $n\times n$, b – вектор простору E^n.
3. Спростити метод лінеаризації для розв'язання задачі:
   $$f(x)\to min,\ x\ge 0.$$
4. Установити неперервну диференційовність штрафної функції:
   $$P(x,r)=r\sum_{i=1}^{m}\left|\,max\{g_i(x),0\}\right|^{q},\ q>1,$$
   де $g_i(x)$, $i=\overline{1,m}$ неперервно-диференційовні функції. Обчислити градієнт цієї функції.
5. Обчислити декілька ітерацій методу умовного градієнта для функції $f(x)=x_1^2+x_1x_2+x_2^2$ при $x\in X=\left\{x\in E^2 : 0\le x_1\le 1, -1\le x_2\le 0\right\}$, вибираючи за початкове наближення $x^{(0)}=(1,-1)$, $(-1,0)$, $(1,0)$, $(0,0)$.
6. Знайти умовний антиградієнт h функції $f(x)=-\frac{1}{8}x_1^2-\frac{1}{3}x_2^2+2$ в точці $(2,3)$ на множині $X=\left\{x\in E^2 : x_1+x_2\le 6, x_1-x_2\le 1, 2x_1+x_2\ge 6,\right.$

$\left.\frac{1}{2}x_1 - x_2 \geq -4, x_1 \geq 0, x_2 \geq 0\right\}$. Знайти довжину кроку α уздовж h з умови одновимірної мінімізації та по правилу дроблення кроку.

7. Розв’язати задачу методом проекції градієнта Розена:

$$f(x) = -x_1^2 + 2x_1x_2 - x_2^2 + 10x_1 \to min,$$

$$x_1 + x_2 \leq 6, x_1 + 2x_2 \leq 10,\ x_1 \geq 0, x_2 \geq 0,$$

вибираючи за початкове наближення $x^{(0)} = (1, 1)$.

8. Розв’язати задачу методом Зойтендейка при різних початкових наближеннях

$$f(x) = x_1 + x_2 \to min,\ g_1(x) = x_1^2 - 2x_2 \leq 0,\ g_2(x) = x_2 - 1 \leq 0.$$

9. Розв’язати задачу методом Зойтендейка, вибираючи за початкове наближення точку $x^{(0)} = (0, 0, 0, 0)$.

$$f(x) = x_1 + 3x_2 - x_3 + x_4 - x_1^2 + x_1x_3 - \frac{1}{2}x_2^2 - x_3^2 - x_3x_4 - \frac{1}{2}x_4^2 \to min,$$

$$x_1 + 2x_2 + x_3 + x_4 \leq 5,\ 3x_1 + x_2 + 2x_3 - x_4 \leq 4,\ x_j \geq 0,\ j = \overline{1,4},$$

10. Розв’язати методом проекції градієнта та методом можливих напрямків таку задачу:

$$f(x) = exp(x_1) + x_1^2 - x_1x_2 - 3x_2^2 + 4x_1 - 6x_2 \to min,$$

$$2x_1 + x_2 \leq 8, -x_1 + x_2 \leq 2,\ x_1 \geq 1, x_2 \geq 3.$$

11. Розв’язати методом можливих напрямків таку задачу:

$$f(x) = x_1^3 + 2x_2^3 + x_1 - 2x_2 - x_1^2 \to min,$$

$$x_1 + 2x_2 \leq 6, -x_1 + 2x_2 \leq 3,\ x_1 \geq 0, x_2 \geq 0.$$

Чи є точка, яка отримана за алгоритмом методу можливих напрямків, глобальним мінімумом, локальним мінімумом, чи ні тим ні іншим?

12. Розв’язати такі задачі методом Зойтендейка:

Цільова функція: $f(x) = x_1^2 + x_1x_2 + 2x_1^2 - 6x_1 - 2x_2 - 12x_3 \to min,$

- $x_1 + x_2 + x_3 = 2,\ -2x_1 + 2x_2 \leq 3,\ x_j \geq 0,\ j = \overline{1,3},$

- $2x_1^2 + x_2^2 \le 15,\ -x_1 + 2x_2 + x_3 \le 3,\ x_j \ge 0,\ j = \overline{1,3}$

13. Розв'язати модифікованим методом можливих напрямків таку задачу:

$$f(x) = exp(-x_1 - x_2) + (1 - x_1)^2 - 10(x_2 - x_1)^2 + x_1^2 - 2x_1x_2 \to min,$$

$$x_1^2 + x_2^2 \le 16,\ (x_2 - x_1)^2 + x_1 \le 6,\ x_1 + x_2 \ge 2.$$

14. Розв'язати таку задачу методом проекції градієнта:

$$f(x) = exp(-x_1 - x_2) + (1 - x_1)^2 - 10(x_2 - x_1)^2 + x_1^2 - 2x_1x_2 \to min,$$

$$2x_1 + 5x_2 \le 25,\ -x_1 + 2x_2 + x_3 \le 8,\ x_j \ge 0,\ j = \overline{1,2}.$$

15. Узагальнити метод штрафних функцій для розв'язання такої задачі:

$$f(x) = x_1^3 - 6x_1^2 + 11x_1 + x_3 \to min,$$

$$x_1^2 + x_2^2 - x_3^2 \le 0, \quad x_1^2 + x_2^2 + x_3^2 \ge 4,\ x_1 \ge 0, x_2 \ge, 0 \le x_3 \le 5.$$

16. Вибираючи за початкове наближення точку $x^{(0)} = (0,0)^T$, розв'язати методом Зойтендейка і проекції градієнта таку задачу:

$$f(x) = x_1^2 + x_1x_2 + 2x_2^2 - 12x_1 - 18x_2 \to min,$$

$$-3x_1 + 6x_2 \le 9,\ -2x_1 + x_2 \le 1,\ x_j \ge 0,\ j = \overline{1,2}.$$

17. Розв'язати методом штрафних функцій такі задачі:

- $f(x) = x_1 + x_2 \to min,\ x_1^2 - x_2 \le 0,\ -x_1 \le 0$

  (з використанням логарифмічної функції штрафу).

- $f(x) = x^2 - 10x \to min,\ x - 1 \le 0$

  (з використанням квадратичної функції штрафу).

- $f(x) = (x_1 - x_2)^2 + (x_2 - x_3)^4 \to min,\ x_1 + x_1x_2^2 + x_3{}^4 = 3.$

18. Розглянемо таку задачу:

$$f(x) = exp(x_1) + x_1{}^2 + 4x_2^2 + 2x_1x_2 \to min,$$

$$x_1 + 2x_2 - 6 = 0$$

Побудуйте відповідну штрафну функцію при $r = 10$. Виконайте, починаючи з точки $(1,1)$, дві ітерації методу спряжених градієнтів.

19. Розглянемо таку задачу:

$$f(x)=(x_1-6)^2+(x_2-8)^2 \to min,$$

$$x_1^2-x_2 \le 0.$$

Використовуючи допоміжну функцію

$$f(x)=(x_1-6)^2+(x_2-8)^2+r_k \; max\left\{\left(x_1^2-x_2\right),0\right\}$$

і метод по координатного спуску, розв'язати задачу при начальній точці $x^{(0)}=(0,-4)^T$ і $r_1=0.1$.

20.Розглянемо таку задачу:

$$f(x)=x_1{}^2+x_2{}^2 \to min,$$

$$2x_1+x_2-2 \le 0, \; -2x_2+1 \le 0.$$

Побудуйте відповідну штрафну функцію при $r_0=1$. При початковій точці $x^{(0)}=(1,6)^T$ розв'язати отриману задачу одним з методів нульового порядку, першого порядку, вибравши відповідну штрафну функцію.

21.Розв'язати задачу методом послідовної безумовної оптимізації:

$$f(x)=-exp(-x_1-x_2)-x_1{}^2+x_2^2+2x_1x_2 \to min,$$

$$x_1+x_2-4=0, \;\; x_1+x_2 \le 1.$$

22.Розглянемо таку задачу:

$$f(x)=exp(x_1)-x_1x_2+x_2{}^2 \to min,$$

$$2x_1+x_2-4 \le 0, \; x_1 \ge 0, \; x_2 \ge 0.$$

- розв'язати цю задачу методом внутрішньої точки, використовуючи за початкову точку $x^{(0)}=(1,1)^T$ при $r_0=1, \; \beta=0.1$;
- розв'язати цю ж задачу методом зовнішньої точки, використовуючи за початкову точку $x^{(0)}=(2,4)^T$ при $r_0=1, \; \beta=10$.

Побудувати траєкторії руху для обох методів та порівняти їх.

23.Розв'язати задачі методом умовного градієнта:

- $f(x)=-x_1{}^2-x_2{}^2+x_1x_3 \to min,$

$$3x_1+x_2+x_3 \le 4, \;\; x_1+2x_2+2x_3=3, \;\; x_j \ge 0, \, j=\overline{1,3}\,,$$

$$x^{(0)} = (1,0,1)\ .$$

- $f(x) = -x_1x_2 + x_1x_3 - 2x_2x_3 + x_1 - 5 \to max$,

$$x_1 + x_2 + x_3 = 3,\ \ 2x_1 + x_2 \le 5,\ \ x_1 \ge 0,$$

$$x^{(0)} = (2,1,0)\ .$$

- $f(x) = -3x_1^{\ 2} - x_2^2 + x_2 + 7x_3 \to max$,

$$4x_1 + x_2 - 2x_3 \le 5,\ \ 2x_2 + x_3 \le 14,\ \ x_j \ge 0,\ j = \overline{1,3}\ ,$$

$$x^{(0)} = (0,5,0)\ .$$

- $f(x) = x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} \to min$,

$$x_1 + x_2 + x_3 = 3,\ \ 2x_1 - x_2 + x_3 \le 5,$$

$$x^{(0)} = (1,1,1)\ .$$

- $f(x) = 2x_1^{\ 2} + x_2^{\ 2} + 2x_3^{\ 2} - x_1x_3 - 6 \to max$,

$$x_1 - 3x_2 + 2x_3 \le 1,\ \ 2x_1 + x_2 \le 10,\ x_j \ge 0,\ j = \overline{1,3}\ ,$$

$$x^{(0)} = (1,1,1)\ .$$

24. Знайти мінімум функції $f(x)$ (дивись завдання № 6 розділу 3) на заданій множині одним із заданих методів умовної оптимізації. Розглянути:

- випадок лінійних обмежень:

$$x_1 + x_2 \ge 4,\ \ x_1 + 4x_2 - 16 \le 0,\ 11x_1 - 4x_2 - 44 \le 0\ ;$$

- випадок нелінійних обмежень:

$$x_1^{\ 2} - x_2 - 4 \ge 0\,,\ x_1 + 4x_2 - 16 \le 0,\ 11x_1 - 4x_2 - 44 \le 0\ .$$

Розділ 5. Чисельні методи безумовної недиференційовної оптимізації

До задач мінімізації опуклих функцій з розривним градієнтом зводиться велика кількість проблем, які виникають при розв'язуванні складних задач математичного програмування. Володіння методами недиференційовної оптимізації дає можливість гнучко використовувати різноманітні схеми декомпозиції (за змінними, обмеженнями і таке інше), враховуючи специфіку задач великої розмірності, дозволяє ефективно одержувати двоїсті оцінки в задачах дискретного та неперервно-дискретного програмування і для деяких класів багатоекстремальних задач. З'являється можливість вільно використовувати точні негладкі функції штрафу, які дозволяють при скінченних значеннях штрафних параметрів одержувати задачу безумовної оптимізації, повністю еквівалентну початковій задачі опуклого програмування. Штрафні функції у формі функції максимуму успішно використовуються при розв'язуванні задач з великою кількістю обмежень, оскільки при цьому значно зменшуються витрати на обчислювання субградієнта штрафної функції.

Техніко-економічні характеристики об'єктів, які оптимізуються, часто добре апроксимуються Кусково-гладкими функціями від невідомих параметрів, унаслідок чого з'являються задачі, цільова функція та обмеження яких виражаються за допомогою негладких функцій. Відсутність ефективних методів негладкої оптимізації перешкоджала в минулому розв'язуванню зазначених класів задач та примушувала або змінювати постановку задачі, що погіршувало відповідність математичної моделі наявній задачі, або використовувати всілякі способи згладжування. Останні не завжди приводять до успіху, оскільки використання згладжувальних функцій з малим параметром викликає погану обумовленість функції, що мінімізується, і це погіршує обчислювальну стійкість таких ефективних методів гладкої мінімізації, як квазіньютонівські та методи спряжених градієнтів. Слід відзначити, що ефективні методи негладкої оптимізації виявилися на практиці цілком конкурентоспроможними як за надійністю, так і за часом обчислення на ЕОМ та точністю результатів з найбільш ефективними методами розв'язання гладких погано обумовлених задач. Таким чином, область застосування методів негладкої оптимізації дуже широка, тому велика увага приділяється розробці обчислювальних методів негладкої оптимізації.

Матеріал цього розділу цілком базується на роботах Н.З. Шора.

§1. Метод узагальненого градієнтного спуску

Нехай цільова функція $f(x)$ є такою, що можливе практичне обчислення субградієнта $g(x)$ у кожній точці $x \in E^n$. Тоді градієнтний алгоритм безпосередньо узагальнюється на недиференційовний випадок. Для цього достатньо за напрямок спуску взяти на кожному k-ому кроці вектор протилежний субградієнту в точці $x^{(k)}$ (замість вектора антиградієнта у випадку диференційовної функції).

Нехай задана опукла функція $f(x)$ з областю визначення $X \subseteq E^n$ і $\bar{x}$ – внутрішня точка X. Вектор $g(\bar{x})$, який задовольняє співвідношенню

$$\left(g(\bar{x}),\, x-\bar{x}\right) \le f(x) - f(\bar{x}), \quad \forall x \in X,$$

називається субградієнтом (узагальненим градієнтом) функції $f(x)$ у точці $\bar{x}$.

Розглянемо задачу знаходження мінімуму опуклої функції, яка не є скрізь диференційовною

$$f(x) \to min,\ x \in E^n. \qquad (5.1)$$

Нехай $x^{(0)}$ – початкова точка. Відповідно до методу узагальненого градієнтного спуску пошук точки мінімуму здійснюється згідно з таким рекурентним співвідношенням

$$x^{(k+1)} = x^{(k)} - \alpha_k \frac{g\left(x^{(k)}\right)}{\left\| g\left(x^{(k)}\right) \right\|},\ k = 0,1,2,\ldots, \qquad (5.2)$$

Дамо геометричну інтерпретацію методу узагальненого градієнтного спуску (УГС). Оскільки $f(x)$ не має неперервних похідних, то в точці $x^{(k)}$ лінія рівня $\Gamma_k = \left\{ x \in E^n : f(x) = f\left(x^{(k)}\right) \right\}$ може зазнавати злому (рис. 5.1). У цьому випадку замість дотичної площини в точці $x^{(k)}$ існує множина опорних гіперплощин, кожній з яких відповідає узагальнений градієнт $g\left(x^{(k)}\right)$, напрямлений поза множиною $\left\{ x : f(x) \le f\left(x^{(k)}\right) \right\}$. На рис. 5.1 показано, що коли величина α_k не перевищує довжини відрізка $\left[A, x^{(k)}\right]$, то точка $x^{(k+1)}$, яка побудована згідно з (5.2), хоч і не обов'язково, потрапляє в область менших значень функції $f(x)$, все ж буде ближче до точки мінімуму x_*, ніж $x^{(k)}$.

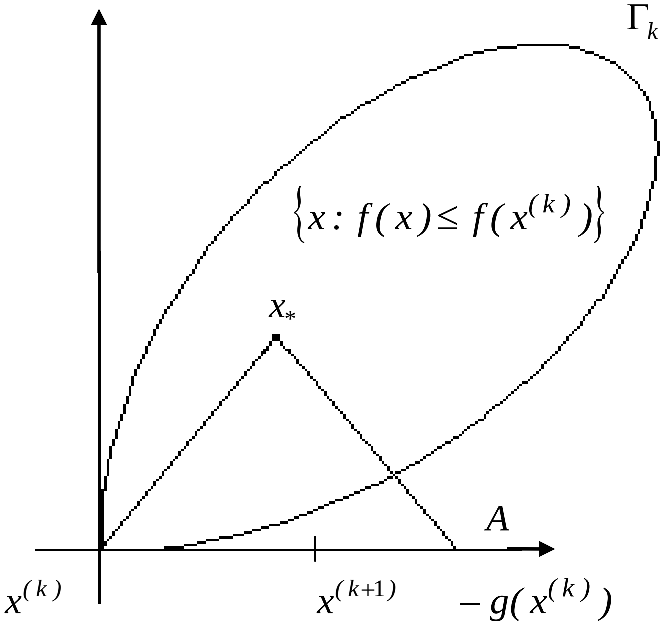


Рис. 5.1

Складність тут полягає в тому, що при русі з точки $x^{(k)}$ у напрямку $\left(-g\left(x^{(k)}\right)\right)$ цільова функція не обов'язково повинна зменшуватися навіть при скільки завгодно малому кроковому множнику α_k, тобто регулювання α_k не може бути поставлено в залежність від поведінки цільової функції $f(x)$, як це було за градієнтним методом, а повинно бути задано якимось іншим способом.

Виявляється, якщо величину кроку α_k на k-тій ітерації вибирати з умови:

$$\alpha_k \to 0,\ k = 0,1,\ldots,\ \sum_{k=0}^{\infty} \alpha_k = \infty,\ \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \tag{5.3}$$

то процедура (5.2) дозволяє знайти мінімум цільової функції $f(x)$. Відзначимо, що розбіжність ряду $\sum_{k=0}^{\infty}\alpha_k$ є природною вимогою, бо у випадку $\sum_{k=0}^{\infty}\alpha_k < \infty$ завжди можна було б вибрати таке далеке від x_* початкове наближення $x^{(0)}$, при якому неможливо було б наблизитись до x_* при скільки завгодно великих k.

Наприклад, за α_k можна узяти $\alpha_k = \frac{1}{k+1}$.

Теорема про збіжність. Нехай функція $f(x)$ є опуклою і визначеною на E^n. Якщо кроковий множник α_k вибирається з умов

$$\alpha_k > 0,\ \alpha_k \to 0,\ k = 0,1,\ldots,\ \sum_{k=0}^{\infty}\alpha_k = \infty,$$

то при будь-якій початковій точці $x^{(0)}$ послідовність $\left\{f\left(x^{(k)}\right)\right\}$ збігається до множини точок мінімуму X_*. Якщо множина точок мінімуму X_* непорожня, то при додатковій умові

$$\sum_{k=0}^{\infty}\alpha_k^2 < \infty$$

послідовність $\left\{x^{(k)}\right\}$ збігається до X_*, якщо $k \to \infty$.

Метод узагальненого градієнтного спуску відзначається простотою, проте його швидкість збіжності не може бути високою.

Алгоритм методу

Початковий етап. Вибрати $\varepsilon > 0$, як константу зупинки. Задати початкову точку $x^{(0)}$. Покласти $k = 0$ та перейти до основного етапу.

Основний етап

Крок 1. Знайти субградієнт $g\left(x^{(k)}\right) \in \partial f\left(x^{(k)}\right)$. Якщо $g\left(x^{(k)}\right) = 0$, то зупинитися: точка $x^{(k)}$ є оптимальним розв’язком задачі. Інакше, перейти до кроку 2.

Крок 2. Знайти α_k. Прийняти

$$x^{(k+1)} = x^{(k)} - \alpha_k \frac{g\left(x^{(k)}\right)}{\left\| g\left(x^{(k)}\right)\right\|},\ k = 0,1,2,\ldots,$$

і перейти до кроку 3.

Крок 3. Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| < \varepsilon$, то зупинитися, у протилежному випадку замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

При певних додаткових припущеннях удається отримати варіант методу узагальненого градієнтного спуску, який збігається зі швидкістю геометричної прогресії.

Нехай x_* –єдина точка мінімуму $f(x)$ і $0 \le \varphi(x) \le \frac{\pi}{2}$,

$$cos\,\varphi(x) = \frac{(g(x), x - x_*)}{\|g(x)\| \, \|x - x_*\|} \text{ при } x \ne x_*$$

Позначимо

$$\varsigma(R) = \sup_{\{x:\|x - x_*\| \le R, x \ne x_*\}} (\sin\varphi(x)).$$

Значення $\varsigma(R)$ можна розглядати як показник, що характеризує ступінь витягнутості ліній (поверхонь) рівня функції $f(x)$, аналог показника «обумовленості» задачі мінімізації $f(x)$. Так, якщо $f(x) = (A\,x, x)$ – квадратична функція з додатно визначеною матрицею A, а ρ – відношення максимального та мінімального власних чисел матриці A, то при будь-якому R маємо $\varsigma(R) = \frac{\rho - 1}{\rho + 1}$. Якщо $\varsigma(R) = 1$ при деякому R, то цільову функцію будемо називати істотно яристою.

Нехай нам, апріорі, відомо, що $\|x - x_*\| < R$ та задано число q: $\varsigma(R) \le q \le 1$. Тоді можна побудувати процес узагальненого градієнтного спуску, який збігається зі швидкістю геометричної прогресії із знаменником q при $q \ge \frac{\sqrt{2}}{2}$ та знаменником $\frac{1}{2\sqrt{1-q^2}}$ при $q \le \frac{\sqrt{2}}{2}$.

Теорема. Нехай функція $f(x)$ є опуклою, точка x_* – єдина точка мінімуму, $\|x - x_*\| < R$ та для усіх $\{x : x \neq x_*,\ \|x - x_*\| < R\}$

$$(g(x), x - x_*) \geq \sqrt{1 - q^2}\, \|g(x)\| \, \|x - x_*\|, \quad q < 1. \tag{5.4}$$

Тоді, якщо прийняти

$$\alpha_0 = \begin{cases} R\sqrt{1 - q^2}, & q \geq \dfrac{\sqrt{2}}{2}, \\ \dfrac{R}{2\sqrt{1 - q^2}}, & q < \dfrac{\sqrt{2}}{2}, \end{cases}$$

$$\alpha_{k+1} = \alpha_k\, r(q), \quad k = 0, 1, 2, \ldots,$$

де

$$r(q) = \begin{cases} q, & q \geq \dfrac{\sqrt{2}}{2}, \\ \dfrac{1}{\sqrt{1 - q^2}}, & q < \dfrac{\sqrt{2}}{2}, \end{cases}$$

то послідовність $\{x^{(k)}\}$, яка обчислюється за формулою узагальненого градієнтного спуску (5.2), або при деякому $k = \bar{k}$ переривається в точці мінімуму, або

$$\|x^{(k)} - x_*\| \leq \begin{cases} R\, q^k, & q \geq \dfrac{\sqrt{2}}{2}, \\ \dfrac{R}{2\sqrt{1 - q^2}}, & q < \dfrac{\sqrt{2}}{2}, \end{cases} \qquad k = 0, 1, 2, \ldots$$

Цей результат має просту геометричну інтерпретацію. Розглянемо сферу $S(x^{(0)}, R)$ з центром у точці $x^{(0)}$ радіуса R – область локалізації точки мінімуму x_*. Обчислимо $g(x^{(0)})$. Гіперплощина $(g(x^{(0)}), x - x^{(0)}) = 0$ відокремлює півпростір $P = \{x : (g(x^{(0)}), x - x^{(0)}) \leq 0\}$ та відповідну півсферу $P \cap S(x^{(0)}, R)$ – область локалізації мінімуму. Умова (5.4) вирізує сферичний конус $K(x^{(0)}, q)$ з вершиною в точці $x^{(0)}$, віссю симетрії, яка збігається з напрямком $\left(-g\left(x^{(0)}\right)\right)$ та твірними, які знаходяться під кутом ς з $\left(-g\left(x^{(0)}\right)\right)$, $sin\,\varsigma = q$.

Таким чином, область локалізації точки мінімуму x_* утворює фігуру $P \cap S\left(x^{(0)},R\right) \cap K\left(x^{(0)},q\right)$, плоский переріз якої зображено на рис. 5.2 (заштрихована область).

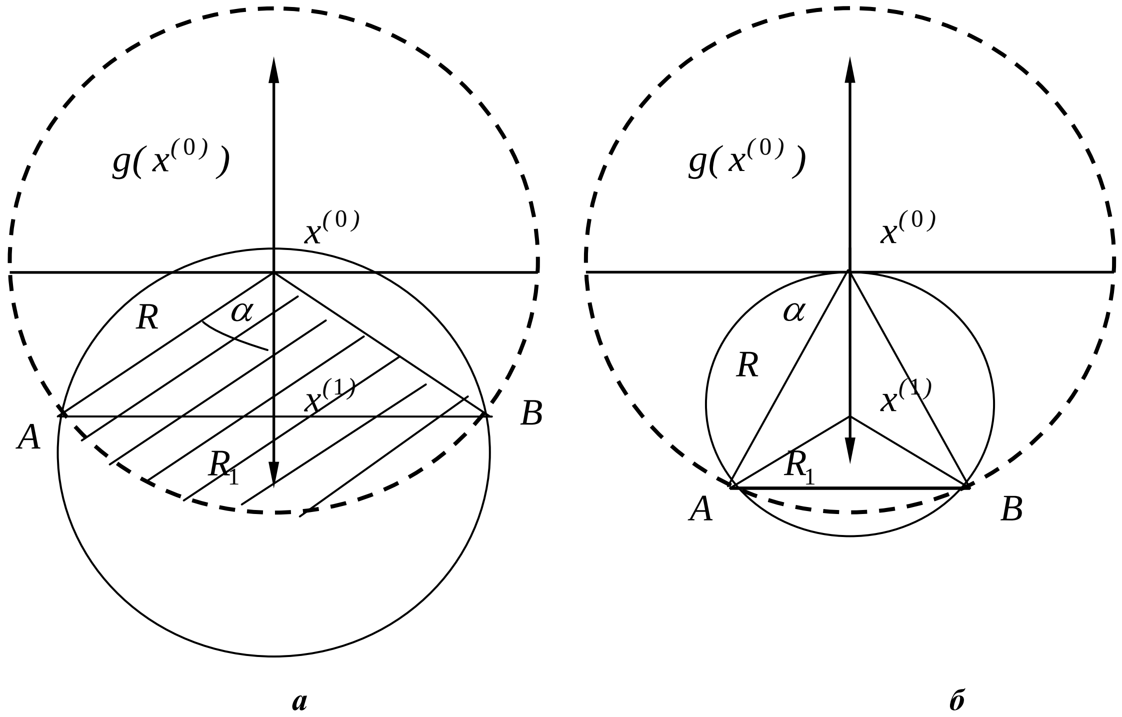


Рис. 5.2

Рис. 5.2,*а* відповідає випадку $\varsigma \geq \dfrac{\pi}{2}$, а рис. 5.2,*б* відповідає випадку $\varsigma < \dfrac{\pi}{2}$. Опишемо навколо заштрихованих фігур сфери мінімального радіуса. У першому випадку центр такої сфери лежить на середині хорди *AB*, а радіус дорівнює $R_1 = R\,sin\varsigma = Rq$. Довжина кроку $\left\|x^{(0)} - x^{(1)}\right\| = R\,\sqrt{1-q^2} = R\,cos\varsigma$. У другому випадку центр сфери збігається з центром описаного навколо трикутника *AOB* кола, а його радіус $R_1 = R/(2\,cos\varsigma) = R/\left(2\sqrt{1-q^2}\right)$ і збігається з α_0.

Приклад. Знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x = (x_1, x_2) \in E^2$$

за допомогою методу узагальненого градієнтного спуску.

Виберемо початкову точку $x^{(0)} = (0.0, 3.0)$ та точність обчислювання $\varepsilon = 10^{-6}$. Після 430 ітерацій дістанемо точки $(2.0365955, 1.0183076)$. Табл. 5.1 наводить результати обчислень за методом узагальненого градієнтного спуску. Оптимальним розв'язком цієї задачі є точка $x_* = (2.00, 1.00)$, $f(x_*) = 0$.

Таблиця 5.1

k	$x^{(k)}$	$f\left(x^{(k)}\right)$	$g\left(x^{(k)}\right)$
40	(2.0978978, 1.0487449)	0.00009202	(0.0045688, –0.0016316)
80	(2.0759439, 1.0382679)	0.00003361	(0.0005682, 0.0023677)
120	(2.0630899, 1.0315036)	0.00001585	(0.0011698, –0.0003307)
160	(2.0561680, 1.0280708)	0.00000995	(0.0007617, –0.0001058)
200	(2.0501518, 1.0250554)	0.00000633	(0.0005865, –0.0001638)
240	(2.0472021, 1.0235684)	0.00000497	(0.0005512, –0.0002610)
280	(2.0431861, 1.0216139)	0.00000348	(0.0002386, 0.0001671)
320	(2.0406370, 1.0203528)	0.00000273	(0.0001310, 0.0002749)
360	(2.0390143, 1.0195435)	0.00000232	(0.0000923, 0.0002906)
400	(2.0375732, 1.0188239)	0.00000200	(0.0000629, 0.0002984)

§2. Оператор розтягу простору

Нехай задано вектор $\xi \in E^n$, такий, що $\|\xi\| = 1$, та число $\alpha \geq 0$. Кожний вектор x однозначно зображується у вигляді

$$x = \gamma_\xi(x)\,\xi + d_\xi(x) \tag{5.5}$$

при умові.

$$\left(\xi, d_\xi(x)\right) = 0\,. \tag{5.6}$$

З (5.5), (5.6) маємо

$$\gamma_\xi(x) = (x,\ \xi),\quad d_\xi(x) = x - (x, \xi)\,\xi\,.$$

Оператором розтягу простору E^n у напрямку ξ з коефіцієнтом α назвемо оператор $R_\alpha(\xi)$, що діє на вектор x, даний у формі (5.5):

$$R_\alpha(\xi)\, x = \alpha\, \gamma_\xi(x)\, \xi + d_\xi(x).$$

З цього означення виходить такі властивості оператора розтягу простору:

1. $R_\alpha(\xi)\, x = \alpha(x, \xi)\, \xi + [x - (x, \xi)\, \xi] = x + (\alpha - 1)(x, \xi)\, \xi$, (5.7)
2. Оператор $R_\alpha(\,\xi\,)$ є лінійним та симетричним:
   $(R_\alpha(\xi)\, x, y) = (\alpha - 1)(x, \xi)(y, \xi) + (x, y) = (x, R_\alpha(\xi)\, y)$.
3. $R_{\alpha\beta}(\xi\,) = R_\alpha\,(\xi) \cdot R_\beta(\xi)$.
4. Оператор $R_0(\xi)$ є оператором проектування на півпростір, ортогональний вектору ξ:
   $$R_0(\xi)\, x = d_\xi(x).$$
5. Оператор $R_\alpha(\,\xi\,)$ у матричній формі зображується у вигляді
   $$R_\alpha(\xi) = I + (\alpha + 1)\, \xi \cdot \xi^T,$$
   де I – одиничний оператор.

§3. Узагальнений градієнтний спуск з розтягом простору в напрямку субградієнта

Основним фактором, який уповільнює збіжність градієнтних методів, є близькість до нуля косинуса кута між напрямком градієнта та напрямком на точку мінімуму. Якщо б ми змогли значно зменшити складову антиградієнта, яка ортогональна напрямку на точку мінімуму, залишаючи майже без зміни складову, яка паралельна цьому напрямку, то отриманий при цьому вектор давав би напрямок спуску кращий, ніж напрямок антиградієнта, і можна було б очікувати прискорення збіжності при відповідному регулюванні кроку. Ця ідея лежить в основі методів узагальненого градієнтного спуску з розтягом простору у напрямку градієнта (субградієнта).

Загальний алгоритм методів

Початковий етап. Задати початкове наближення $x^{(0)} \in E^n$ та початкову неособливу матрицю $B_0 = I$, покласти $k = 0$ та перейти до основного етапу.

Основний етап

Крок 1. Обчислити субградієнт $g\left(x^{(k)}\right)$. Якщо $g\left(x^{(k)}\right)=0$, то обчислення закінчуються: точка $x^{(k)}$ є точкою мінімуму.

Крок 2. Обчислити

$$\tilde{g}_k = B_k^T\ g\left(x^{(k)}\right).$$

Вектор $\tilde{g}_k$ є субградієнтом функції $\varphi_k(y)=f(B_k\ y)$, яка виходить з функції $f(x)$ при застосуванні лінійного перетворення простору $y=A_k\ x$, де $A_k = B_k^{-1}$.

Визначити
$$\xi_k = \frac{\tilde{g}_k}{\left\|\tilde{g}_k\right\|},\ \alpha_k,\ \lambda_k,$$

де α_k – крок переміщення, λ_k – коефіцієнт розтягу простору на k -ому кроці.

Визначити

$$x^{(k+1)} = x^{(k)} - \alpha_k B_k\ \xi_k = x^{(k)} - \alpha_k \frac{B_k\ B_k^T\ g\left(x^{(k)}\right)}{\left\|B_k^T\ g\left(x^{(k)}\right)\right\|},$$

$$B_{k+1} = B_k \cdot R_\mu(\xi_k),\ \mu = \frac{1}{\lambda_k}.$$

Крок 3. Якщо критерій зупинки обчислень виконано, то кінець. Інакше, $k=k+1$ та перейти до кроку 1.

Алгоритм описаний.

Існує стільки ж алгоритмів цього типу, скільки і способів вибору коефіцієнтів α_k, λ_k на кожній ітерації.

Розглянемо **метод розтягу простору в напрямку субградієнта у поточній точці**. У цьому методі на k -ій ітерації вибираємо $\lambda_k = \lambda > 1$ при будь-яких k, а

$$\alpha_k = \rho\ \frac{f\left(x^{(k)}\right) - m}{\left\|B_k^T\ g\left(x^{(k)}\right)\right\|},$$

де ρ – коефіцієнт релаксації, який задовольняє умові $0<\rho\le 2$, m – наближення до оптимального значення цільової функції f_*.

Відносно збіжності методу справедлива така теорема.

Теорема. Нехай опукла функція $f(x)$ має таку властивість: існує число $M>1$ таке, що коли

$$\varphi(\alpha)=f((1-\alpha)\,x+\alpha\, z),\ 0\le\alpha\le 1$$

спадає за α, то виконується нерівність

$$(\,g(x),x-z\,)\le M\,[\,f(x)-f(z)].$$

Окрім того,

$$\lim_{\|x\|\to\infty} f(x)=+\infty .$$

Тоді якщо при використанні алгоритму узагальненого градієнтного спуску з розтягом простору в напрямку субградієнта

$$\lambda_k=\lambda=\frac{M+1}{M-1},\ \rho=\frac{2M}{M+1},$$

причому при $m\ge f_*$ послідовність $\{\alpha_k\}$ є обмеженою і для довільного $\varepsilon>0$ знайдеться $\bar{k}$ таке, що $f\left(x^{\bar{k}}\right)<m+\varepsilon$; якщо $m<f_*$, то послідовність $\{\alpha_k\}$ є необмеженою.

Окрім того, якщо вибрано $m=f_*$, то знайдеться така постійна C, що

$$f(x_l)-m\le C\,\lambda^{l/n}$$

(збіжність $\left\{f(\,x^{(k)}\,)\right\}$ до m є лінійною з множником збіжності $q=\sqrt[n]{\lambda}$).

Для скорочення кількості арифметичних операцій, які пов’язані з перетворенням простору, зручно ввести симетричну, додатно визначено матрицю $H_k=B_k\,B_k^T$.

Алгоритм методу

Початковий етап. Вибрати початкове наближення $x^{(0)}\in E^n$, коефіцієнт розтягу простору λ, коефіцієнт релаксації ρ і наближене значення цільової функції в точці мінімуму m. Вибрати $\varepsilon>0$, як константу зупинки. Покласти $H_0=I$, $k=0$ та перейти до основного етапу.

Основний етап

Крок 1. Знайти субградієнт $g\left(x^{(k)}\right)$. Якщо $g\left(x^{(k)}\right)=0$, то зупинитися: точка $x^{(k)}$ є оптимальним розв'язком задачі. Інакше, перейти до кроку 2.

Крок 2. Обчислити

$$\alpha_k = \rho \frac{f\left(x^{(k)}\right)-m}{g\left(x^{(k)}\right)^T H_k\ g\left(x^{(k)}\right)},$$

$$x^{(k+1)} = x^{(k)} - \alpha_k\ H_k\ g\left(x^{(k)}\right),$$

$$H_{k+1} = H_k - \left(1-\beta^2\right)\frac{\left(H_k\ g\left(x^{(k)}\right)\right)\left(H_k\ g\left(x^{(k)}\right)\right)^T}{g\left(x^{(k)}\right)^T H_k\ g\left(x^{(k)}\right)},$$

де $\beta = \dfrac{1}{\lambda}$.

Крок 3. Якщо $\left\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \right\| < \varepsilon$, то зупинитися, у протилежному випадку замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

§4. *R* – алгоритм Шора

R – алгоритм Шора є методом мінімізації функцій, який використовує операцію розтягу простору в напрямку різниці двох послідовних градієнтів. Розглянемо спочатку необхідність використання методів мінімізації, які використовують розтяг простору.

Використання методу узагальненого градієнтного спуску показало, що за допомогою одного лише регулювання крокового множника, рухаючись на кожному кроці в напрямку узагальненого антиградієнта, важно добитися в досить загальному випадку значного прискорення збіжності. Відзначимо, що повільна збіжність пов'язана з тим, що антиградієнт утворює кут, близький до $\pi/2$ з напрямком на точку мінімуму. У такому випадку відстань до точки мінімуму спадає на величину, набагато меншу, ніж довжина кроку, а отже, і швидкість зменшення кроку не може бути значно великою, якщо ми бажаємо гарантувати збіжність до мінімуму.

З іншого боку, має місце проста можливість зміни кутів між напрямком градієнта і напрямком на точку мінімуму – це використання лінійних неортогональних перетворень простору. Виникає ідея побудови у процесі послідовних наближень лінійних операторів, що змінюють метрику простору, та вибору напрямку спуску, який відповідає напрямку антиградієнта у просторі з новою метрикою. Цей напрямок може вже значно відрізнятися від напрямку антиградієнта.

При аналізі алгоритмів, які збігаються зі швидкістю геометричної прогресії, з’ясувалося, що показник швидкості збіжності визначається величиною $sin\,\tilde{\varphi}$, де $\tilde{\varphi}$ – відома апріорі верхня межа кутів $\varphi(x)$ між напрямком з точки x на точку мінімуму та напрямком узагальненого антиградієнта в даній точці (вважаємо, що $\tilde{\varphi} \geq \pi/4$). Можна спробувати покращити цей показник, застосувавши неортогональне перетворення простору змінних (при ортогональному перетворенні – кути між відповідними променями зберігають свою величину). Так, якщо цільова функція є квадратичною з додатно визначеною матрицею Гессе, то її лінії (поверхні) рівня мають форму концентричних еліпсоїдів і можна підібрати таке лінійне перетворення, при якому лінії (поверхні) рівня стають сферами, тобто у відповідних функцій $\tilde{\varphi} = 0$ (або близька до нуля). Такий підхід фактично використовується в квазіньютоновських методах мінімізації достатньо гладких опуклих функцій.

У загальному випадку добитися ефекту прискорення збіжності за рахунок фіксованого лінійного перетворення не вдається. Це становище пов’язано з тим, що, по-перше, апріорна інформація про функцію, як правило, занадто мізерна для того, щоб вибрати відповідне перетворення, якщо воно існує; по-друге, може статися, що $\tilde{\varphi} = \pi/4$, і ніяке лінійне невироджене перетворення не може змінити зазначеної верхньої межі кутів $\varphi(x)$.

Але, якщо фіксоване лінійне перетворення не дає глобального ефекту, то можна спробувати цілеспрямовано змінювати метрику простору після кожної ітерації. При цьому виникає можливість набувати інформацію про мінімізовну функцію у процесі обчислень. Крім того, послідовність невироджених перетворень простору може наближатися до виродженого перетворення і це дозволяє сподіватися, що і випадок $\tilde{\varphi} = \pi/4$ може бути врахованим.

R – алгоритми при певному регулюванні крокових множників можуть бути зроблені монотонними. Це пов'язано з простим геометричним фактом: якщо ми знаходимось на межі двох «кусків» кусково-гладкої поверхні, причому градієнти до цих кусків, які обчислені в даній точці, утворюють між собою тупий кут, то будь-який розтяг простору в напрямку одного з градієнтів не може перетворити цей кут на гострий, він може лише наблизитися до $\pi/2$, залишаючись тупим. У той же час, розтяг простору в напрямку різниці двох зазначених градієнтів з достатньо великим коефіцієнтом розтягу перетворює тупий кут між градієнтами на гострий.

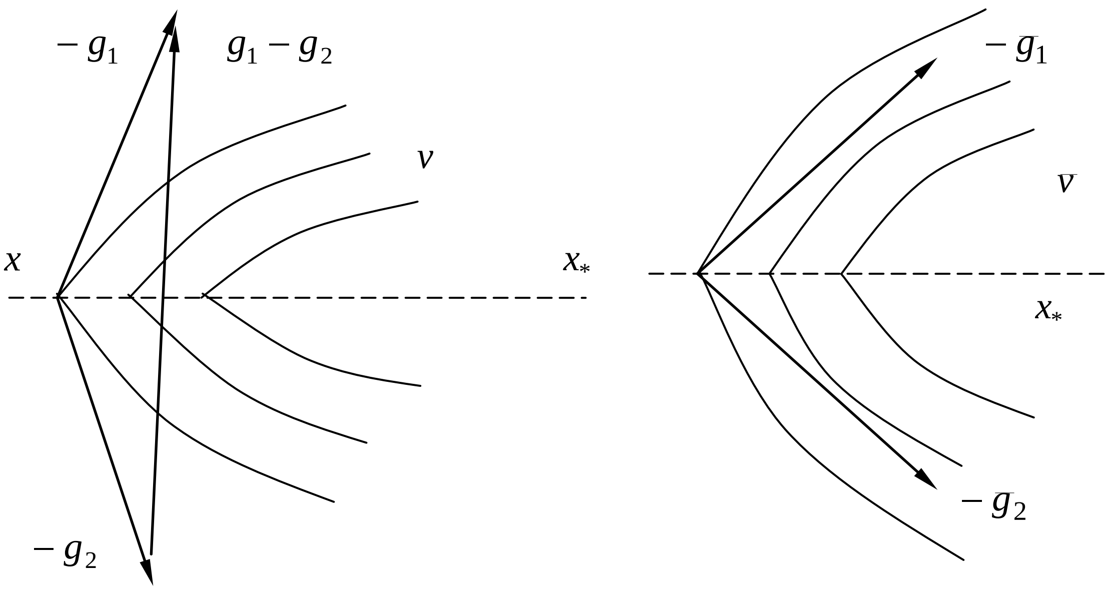


Рис. 5.3

На рис. 5.3 зображено кусково-гладка поверхня рівня деякої опуклої функції $f(x)$, яку позначимо через v. До розтягу простору напрямки антиградієнтів $(-g_1)$ та $(-g_2)$ давали напрямки зростання функції $f(x)$ (поза областю, яка обмежена поверхнею рівня v). Після розтягу простору в напрямку $(g_2 - g_1)$ антиградієнти $(-g_1)$ та $(-g_2)$ дадуть напрямки всередину області, обмеженої відповідної поверхнею v у перетвореному просторі, тобто напрямки $(-\bar{g}_1)$ та $(-\bar{g}_2)$ у перетвореному просторі будуть напрямками спадання цільової функції.

Нехай H_i, $i \ge 0$ – матриця порядку $n \times n$, за допомогою якої буде проводитись перетворення простору.

Алгоритм методу

Початковий етап. Вибрати довільне початкове наближення $x^{(0)} \in E^n$ та коефіцієнт розтягу простору $\lambda > 1$ (часто $\lambda = 2$ або $\lambda = 3$). Покласти $H_0 = I$, $k = 0$. Обчислити $g\left(x^{(0)}\right)$ – градієнт (або субградієнт) цільової функції $f(x)$ у точці $x^{(0)}$ і перейти до основного етапу.

Основний етап

Крок 1. Визначити

$$x^{(k+1)} = x^{(k)} - \alpha_k H_k \, g\left(x^{(k)}\right).$$

Значення α_k визначається із умови мінімуму цільової функції вздовж напрямку $-H_k \, g\left(x^{(k)}\right)$. Якщо $\alpha_k = 0$, то покласти $x^{(k+1)} = x^{(k)}$, $x^{(k)} = x^{(k-1)}$, $g\left(x^{(k+1)}\right) = g\left(x^{(k)}\right)$, $g\left(x^{(k)}\right) = g\left(x^{(k-1)}\right)$.

Крок 2. Обчислити

$$\gamma^{(k)} = g\left(x^{(k+1)}\right) - g\left(x^{(k)}\right),$$

$$H_{k+1} = H_k - (1 - \beta^2) \frac{\left(H_k \gamma^{(k)}\right)\left(H_k \gamma^{(k)}\right)^T}{\gamma^{(k)T} H_k \gamma^{(k)}}, \text{ де } \beta = \frac{1}{\lambda}. \qquad (5.8)$$

Крок 3. Якщо тест зупинки виконано, то кінець алгоритму, інакше замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Зауваження 1. Згідно з алгоритмом, навіть коли він залишається нерухомим в одній і тій самій точці на протязі кількох послідовних ітерацій, матриці H_k продовжують змінюватися (із-за розтягу простору в напрямку різниці двох останніх градієнтів), що дозволяє визначити на кожному кроці ітерації новий напрямок можливого переміщення.

Зауваження 2. Не важко встановити, що R – алгоритм у випадку мінімізації функції однієї змінної вироджується в метод найшвидшого спуску. Тому використання його в цьому випадку не є ефективним.

R – алгоритм має ту перевагу, що ми не фіксуємо перетворення простору, а змінюємо його від кроку до кроку залежно від властивостей цільової функції. У матриці H_k відбувається накопичення інформації про цільову функцію, на основі якої робиться перетворення простору найкращим для функції способом.

Для реалізації знаходження крокового множника α_k з умови мінімуму цільової функції уздовж напрямку ξ з високою точністю необхідна велика кількість обчислень значень цільової функції.

Можна запропонувати такий алгоритм знаходження α_k. Виберемо числа $\mu \geq 1$, $\eta \in [0,1]$, ціле число L та початковий пробний крок h. З точки $x^{(k)}$ робимо крок довжиною h у напрямку ξ: $z_1 = x^{(k)} + h\,\xi$. Обчислюємо $f(z_1)$. Якщо $f(z_1) < f(x^{(k)})$, то з точки z_1 знову робимо крок довжиною h у напрямку ξ: $z_2 = z_1 + h\,\xi$. Обчислюємо $f(z_2)$. Якщо $f(z_2) < f(z_1)$, то з точки z_2 знову робимо крок довжиною h у напрямку ξ, і так до тих пір, поки значення цільової функції на деякому кроці не зросте. Цю останню точку і приймаємо за $x^{(k+1)}$. Нехай l – число кроків у цій процедурі. Тоді якщо $l > L$, то за новий пробний крок приймаємо число $\dfrac{\mu\, l\, h}{L}$. Якщо ж функція зростає на першому кроці, то за новий крок приймаємо число $\eta\, h$, в усіх інших випадках пробний крок не змінюємо. Рекомендується приймати $h = 0.1$, $\mu = 1.25$, $\eta = 0.1$, $L = 10$.

Незважаючи на те, що R – алгоритм зараз є найбільш ефективним і перевіреним практичним способом розв’язування широкого кола задач недиференційовної оптимізації, теоретичне обгрунтування цього класу алгоритмів проведено неповно.

Якщо застосувати до задачі мінімізації опуклої двічі неперервно-диференційовної функції граничний варіант R – алгоритму з відновленням матриці B_k (заміною її одиничною матрицею після кожних n кроків), то справедливою буде така теорема.

Теорема. Нехай опукла функція $f(x)$, яка визначена в E^n, двічі неперервно-диференційовна у певному околі S точки мінімуму x_*, причому в цьому околі матриця інших похідних (гессіан) $H(x)$задовольняє умові Ліпшиця

$\|H(x)-H(z)\| \le L\|x-z\|$, для $\forall\, x, z \in S, \quad L>0$

та $H(x_*)$ – додатно визначена матриця. Тоді знайдеться такий окіл $S' \subset S$ точки x_*, що коли $x^{(0)} \subset S'$, то існує число $C>0$ для якого має місце така оцінка:

$$\| x^{(n)} - x_* \| \le C \| x^{(0)} - x_* \|.^2$$

Тут $x^{(n)}$ – точка, яка одержана після n кроків граничного варіанта R – алгоритму: $\beta_k = 0, \alpha_k$, $k = 0, n-1$, знаходиться з умови мінімуму в напрямку ξ).

Умовою зупинки обчислень може бути виконання однієї з таких нерівностей:

$$\left\| g\left(x^{(n)}\right)\right\| \le \varepsilon_1, \ \left\| x^{(n)} - x^{(n-1)}\right\| \le \varepsilon_2,$$

де ε_1, ε_2 –задані числа.

Приклад. Знайти мінімум функції

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2, \ x = (x_1, x_2) \in E^2$$

за допомогою R – алгоритму.

Результати розв’язання задачі наведені в табл. 5.2, а рис. 5.4 ілюструє проходження до точки мінімуму. Вибрано початкову точку $x^{(0)} = (0.0, 3.0)$ та точність обчислювання $\varepsilon = 10^{-3}$. Після 10 ітерацій була одержана точка $(1.9945165, 0.9972259)$.

Таблиця 5.2

k	$x^{(k)}$	$f\left(x^{(k)}\right)$	$\left\Vert f'\left(x^{(k)}\right)\right\Vert$
0	(0.0000000, 3.0000000)	52.00000	50.1198578
1	(2.7127006, 1.5203232)	0.365554	50.1198578
2	(2.5353696, 1.1773058)	0.114825	1.5324092
3	(2.3400633, 1.2033329)	0.017809	1.2140844
4	(2.2306421, 1.0935800)	0.004728	0.2674976
5	(2.1496673, 1.0849705)	0.000913	0.2208125
6	(2.0665760, 1.0274518)	0.000156	0.0855148
7	(2.0349395, 1.0274518)	0.000016	0.0527394
8	(2.0136800, 1.0060420)	0.000003	0.0168350
9	(1.9931828, 0.9967825)	0.000000	0.0071420
10	(1.9945165, 0.9972259)	0.000000	0.0017103

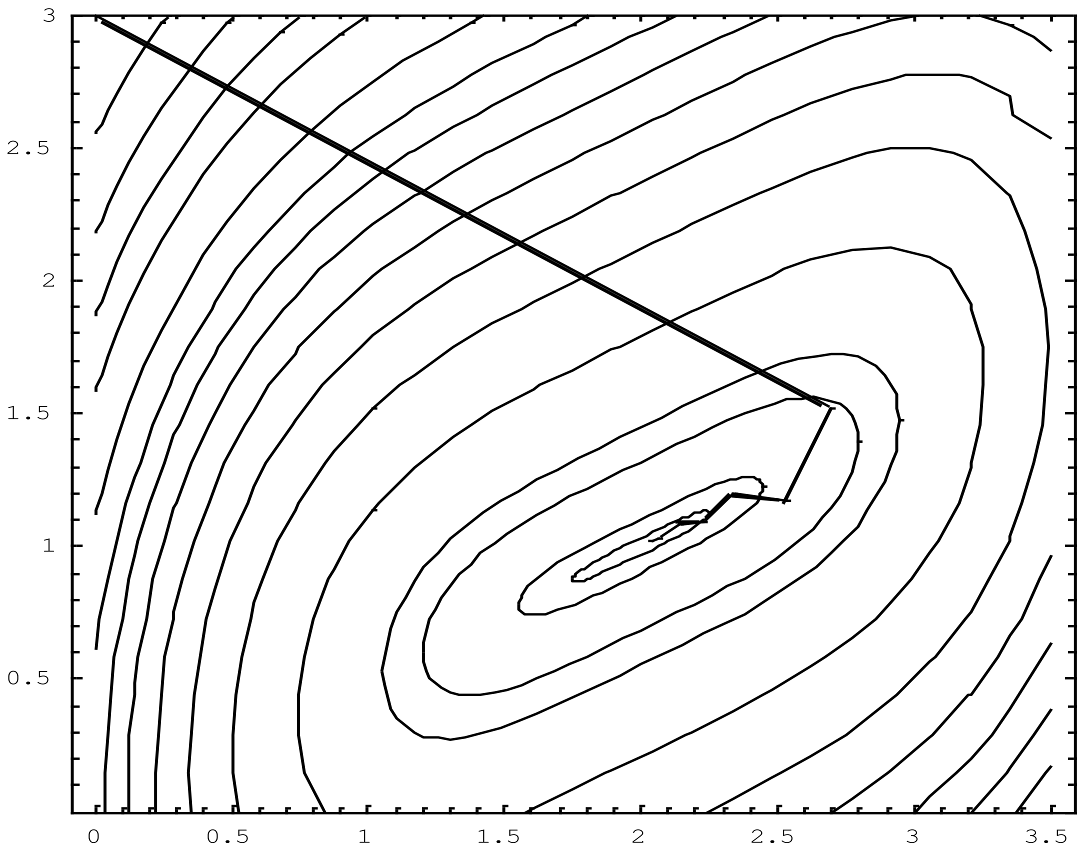


Рис. 5.4

§5. Метод відтинання з розтягом простору (метод еліпсоїдів)

За останні роки велику увагу привернув метод відтинання з розтягом простору (метод еліпсоїдів або метод Шора-Хачияна), який можна розглядати, з одного боку, як окремий випадок методу узагальненого градієнтного спуску з розтягом простору в напрямку субградієнта з постійним коефіцієнтом розтягу і кроковим множником, який змінюється за формулою геометричної прогресії, а з іншого боку, як метод послідовних відтинань, в якому область локалізації розв'язку апроксимується на кожному кроці еліпсоїдом.

Припустимо, що відоме число R та точка $x^{(0)} \in E^n$, яка задовольняє умові $\left\| x^{(0)} - x_* \right\| \le R$, тобто x_* належить кулі $S_0\left(x^{(0)}, R\right)$ радіуса R з центром точці $x^{(0)}$. Точка x_* належить також і півкулі

$$\bar{S}_0\left(x^{(0)}, R\right) = S_0\left(x^{(0)}, R\right) \cap \left\{ x : \left(g\left(x^{(0)}\right), x - x^{(0)}\right) \le 0 \right\}.$$

Побудуємо еліпсоїд мінімального об'єму, який містить $\bar{S}_0\left(x^{(0)}, R\right)$, та за допомогою лінійного перетворення розтягнемо простір таким чином, щоб даний еліпсоїд перетворився на кулю. Зрозуміло, що радіус R_1 цієї нової кулі зменшиться порівняно з $R = R_0$ ($R_1 = R_0 \sqrt[3]{1 - \frac{n}{2}}$). Тому, повторюючи даний процес, можна локалізувати x_* у кулях з радіусом, що послідовно зменшується.

На рис. 5.5 показані два кроки методу у початковому просторі.

Відзначимо, що об'єми еліпсоїдів, які породжуються у процесі процедури, наближаються до нуля зі швидкістю геометричної прогресії із знаменником $$q = \sqrt{\frac{n-1}{n+1}} \left(\frac{n}{\sqrt{n^2 - 1}} \right)^n,$$

який залежить тільки від розміру простору.

Нехай R – мажоранта норми відстані між початковою точкою та оптимальним розв'язком $\left\| x^{(0)} - x_* \right\|$. У цьому методі $\lambda_k = \lambda$ при будь-якому k. А саме,

$$\lambda_k = \lambda = \sqrt{\frac{n-1}{n+1}},$$

$$\alpha_k = \frac{R}{n+1}\left(\frac{n}{\sqrt{n^2-1}}\right)^k \frac{1}{\left\| B_k^T g\left(x^{(k)}\right)\right\|}.$$

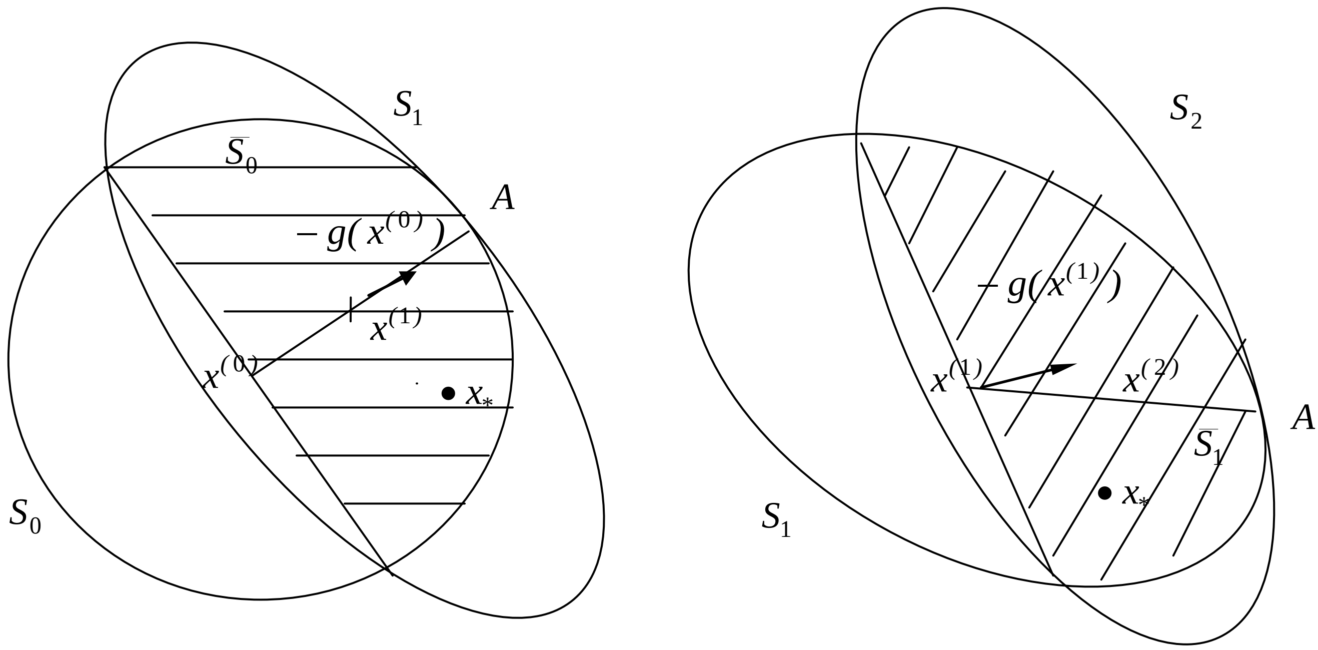


Рис. 5.5.

Алгоритм методу еліпсоїдів

Початковий етап. Вибрати початкове наближення $x^{(0)} \in E^n$, задати $H_0 = R \cdot I$, $k = 0$ і перейти до основного етапу.

Основний етап

Крок 1. Обчислити $g\left(x^{(k)}\right)$ – субградієнт (або градієнт) цільової функції $f(x)$ у точці $x^{(k)}$. Якщо $g\left(x^{(k)}\right) = 0$, то зупинитися: $x^{(k)}$ – шукана точка. Інакше

$$x^{(k+1)} = x^{(k)} - \frac{1}{n+1} \frac{H_k\, g\left(x^{(k)}\right)}{\sqrt{g\left(x^{(k)}\right)^T H_k\, g\left(x^{(k)}\right)}}, \qquad (5.9)$$

$$H_{k+1} = \frac{n^2}{n^2-1}\left(H_k - \frac{2}{n+1} \frac{\left(H_k\ g\left(x^{(k)}\right)\right)\left(H_k\ g\left(x^{(k)}\right)\right)^T}{g\left(x^{(k)}\right)^T H_k\ g\left(x^{(k)}\right)} \right). \qquad (5.10)$$

Крок 2. Якщо тест зупинки виконано, то кінець алгоритму, інакше замінити k на $k+1$та перейти до кроку 1.

Алгоритм описаний.

Теорема. Послідовність $\left\{x^{(k)}\right\}$, яка генерується алгоритмом (5.9)– (5.10), задовольняє нерівності

$$\left\| A_k\left(x^{(k)} - x_*\right) \right\| \le \alpha_k(n+1),\ A_k = B_k^{-1},\ k = 0,1,\ldots$$

Метод еліпсоїдів збігається за функцією у певному розумінні зі швидкістю геометричної прогресії. Але практична його збіжність є повільною.

Позитивною властивістю алгоритму (5.9), (5.10) є те, що гарантована швидкість збіжності залежить лише від розмірності простору і не потребує знання специфічних особливостей цільової функції.

Приклад. Мінімізувати функцію

$$f(x) = (x_1-2)^4 + (x_1-2x_2)^2,\ x = (x_1, x_2) \in E^2$$

за допомогою методу еліпсоїдів.

Задача розв'язувалася при $x^{(0)} = (0.0, 3.0)$, $R = 7$, $m = 0$, $\varepsilon = 10^{-7}$. Після 74 ітерацій була отримана точка $(1.9951310, 0.9975657)$ зі значенням цільової функції $5.62 \cdot 10^{-10}$.

Вправи до розділу 5

1. Для функції двох змінних $f(x_1, x_2) = \begin{cases} 5\sqrt{9x_1^2 + 16x_2^2}, \text{якщо} & x_1 > |x_2| \\ 9x_1 + 16|x_2| \quad , \text{якщо} & x_1 \le |x_2| \end{cases}$

   застосувати метод узагальненого градієнтного спуску з вибором α_k із умови $\min\limits_{\alpha \ge 0} f\left(x^{(k)} - \alpha\ g\left(x^{(k)}\right)\right)$. Пояснити причину невдачі.

2. Знайти розв'язок системи опуклих нерівностей $f_i(x) \le 0,\quad i = 1, 2, \ldots, m$, де $f_i(x)$ – опуклі функції, які визначені на E^n.

1. $m=2$, $f_1(x)=4-x_1-x_2\le 0$, $f_2(x)=x_1^2+x_2^2-16\le 0$.

2. $m=2$, $f_1(x)=2-4\,x_1-3\sqrt{x_2-x_1^2+6}\le 0$, $f_2(x)=2\,x_2-x_1-1\le 0$.

Вказівки. Розв'язування цієї системи зводиться до знаходження мінімуму функції $\varphi(x)$, де $\varphi((x))=max\left\{\max\limits_{1\le i\le m} f_i((x)),0\right\}$. Використати метод узагальненого градієнтного спуску

$$x^{(k+1)}=x^{(k)}-\frac{C}{k+1}\frac{g_\varphi\left(x^{(k)}\right)}{\left\|g_\varphi\left(x^{(k)}\right)\right\|},\ k=0,1,2,\ldots,\ g_\varphi\left(x^{(k)}\right)=g_{i^*}\left(x^{(k)}\right)$$

де i^* – індекс функції, на якій досягається $\max\limits_{1\le i\le m} f_i(x)$, $C>0$.

3. Здійснити перехід від формул перерахування послідовності матриць B_k до формул перерахування послідовності матриць H_k для методу узагальненого градієнтного спуску з розтягом простору в напрямку субградієнта.

4. Розв'язати задачу мінімізації функції вигляду: $f(x)=\max\limits_{1\le i\le m}\ \varphi_i(x)$,

   де $\varphi_i(x)$ гладкі функції, $x\in E^n$, $\varphi_i(x)=b_i\sum\limits_{j=1}^{n}\left(x_j-a_{ij}\right)^2$:

   1) за допомогою *R*-алгоритму;
   2) шляхом зведення цієї задачі до такої задачі опуклого програмування: $y\to min$, $f_i(x)-y\le 0,\quad i=1,2,\ldots,m$.

   Порівняти трудомісткість розв'язування задачі різними методами. Знайти оптимальний розв'язок задачі при $n=5$, $m=10$, $x^{(0)}=(0,\ 0,\ 0,\ 0,\ 1)$, $f(x^{(0)})=80$, $\{b_i\}=\{1,\ 5,\ 10,\ 2,\ 4,\ 3,\ 1.7,\ 2.5,\ 6,\ 3.5\}$, $\lambda=3$

$$A=\{a_{ij}\}=\begin{pmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

Наступні приклади розв’язати за допомогою *R*-алгоритму Шора. Узяти $\varepsilon_1=10^{-6}$, $\varepsilon_2=10^{-7}$, як критерії зупинки $\left\| g(x^{(k)})\right\| \le \varepsilon_1$, $\left\| x^{(k+1)}-x^{(k)}\right\| \le \varepsilon_2$. Розрахунки провести при $\lambda=2$, $\lambda=3$.

5. Мінімізувати функцію $f(x)=100\left(t_1^2-t_2\right)^2+(t_1-1)^2$.

   За початкову обрати точку $x^{(0)}=(-1.2,\ 1)$.

6. Мінімізувати функцію

$$f(x)=\sum_{i=1}^{10}\left(exp(-0.2\,i)+2\,exp(-0.4\,i)-t_1\,exp(-0.2\,t_2\,i)-t_3\,exp(0.4\,t_4\,i)\right)^2.$$

   За початкову обрати точку $x^{(0)}=(0, 0, 0, 0)$.

7. Мінімізувати функцію

$$f(x)=100\left(t_1^2-t_2\right)^2+(t_1-1)^2+90\left(t_3^2-t_4\right)^2+(t_3-1)^2+$$
$$+10.1\left[(t_2-1)^2+(t_4-1)^2\right]+19{,}8(t_2-1)(t_4-1).$$

   За початкову обрати точку $x^{(0)}=(-3,-1,-3,-1)$.

8. Мінімізувати функцію

$$f(x)=(exp(t_1)-t_2)^4+100(t_2-t_3)^6+t_1^4(t_3-t_4)+t_1^8+(t_4-1)^2.$$

   За початкову обрати точку $x^{(0)}=(1, 2, 2, 2)$.

9. Мінімізувати функцію

$$f(x)=(t_1+10t_2)^2+5(t_3-t_4)^3+(t_2-2t_3)^2+10(t_3-t_4)^4.$$

   За початкову обрати точку $x^{(0)}=(10, 10, 10, -10)$.

10. Мінімізувати функцію

$$f(x)=10^{-3}\sum_{i=1}^{10}10^3\left(exp(-0.2\,i)+2\cdot10^3\,exp(-0.4\,i)-\right.$$
$$\left.-t_1\,exp(-0.2\,t_2\,i)-t_3\,exp(0.4\,t_4\,i)\ \right)^2.$$

За початкову обрати точку $x^{(0)}=(500, 0, 2500, 3)$.

1. Мінімізувати функцію $f(x)=x_1^4+(x_1-x_2)^2+(exp(x_2)-1)^2$.

   За початкову обрати точку $x^{(0)}=(1, 1)$.

2. Мінімізувати функцію

$$f(x)=\sum_{\substack{t=0.1\\ \text{з кроком } 0.1}}^{10}\left[exp(-x_1t)-exp(-x_2t)-exp(-t)+exp(-10t)\ \right]^2.$$

За початкову обрати точку $x^{(0)}=(1, 2)$.

Бібліографічні посилання

1. **Акулич И.Л.** Математическое программирование в примерах и задачах. – М.: Высшая школа, 1986.
2. **Аоки М.** Введение в методы оптимизации. – М.: Наука, 1977.
3. **Ашманов С.А.** Линейное программирование. – М.: Наука, 1981.
4. **Базара М.** Нелинейное программирование. Теория и алгоритмы / Базара М., Шетти К. – М.: Мир, 1982.
5. **Бейко И.В.** Методы и алгоритмы решения задач оптимизации / Бейко И.В., Бублик Б.Н., Зинько П.Н. – К.: Вища школа, 1983.
6. **Васильев Ф.П.** Лекции по методам решения экстремальных задач. – М.: МГУ, 1974.
7. **Васильев Ф.П.** Численные методы решения экстремальных задач. – М.: Наука, 1988.
8. **Гольдштейн Е.Г.** Теория двойственности в математическом программировании и ее применение. – М.: Наука, 1971.
9. **Демьянов В.Ф.** Недифференцируемая оптимизация / Демьянов В.Ф., Васильев Л.В. – М.: Наука, 1981.
10. **Ермолаев Ю.М.** Математические методы исследования операций. Учебное пособие для вузов / Ермолаев Ю.М., Ляшко И.И., Михалевич В.С., Тюптя В.Н. – К.: Вища школа, 1979.

11. **Жалдак М.І.** Основи теорії та методов оптимізації: Навчальний посібник / Жалдак М.І., Триус Ю.В. – Черкаси: Брама-Україна, 2005.

12. **Зайченко Ю.П.** Исследование операций. – К.: Вища школа, 1988.

13. **Зайченко Ю.П.** Исследование операций / Зайченко Ю.П., Шумилова С.А. – К.: Вища школа, 1984.

14. **Карманов В.Г.** Математическое программирование. – М.: Наука, 1986.

15. **Кудрявцев Е.М.** ИСО в задачах, алгоритмах, программах. – М.: Радио и связь, 1984.

16. **Ляшенко И.Н.** Линейное и нелинейное программирование / Ляшенко И.Н., Карагодова Е.А., Черникова Н.В., Шор Н.З. – К.: Вища школа, 1975.

17. **Мину М.** Математическое программирование. – М.: Наука, 1992.

18. **Михалевич В.С.** Методы невыпуклой оптимизации / Михалевич В.С., Гупал А.М., Норкин В.И. – М.: Наука, 1987.

19. **Михалевич В.С.** Оптимизационные задачи производственно–транспортного планирования / Михалевич В.С., Трубин В.А., Шор Н.З. – М.: Наука, 1986.

20. **Моисеев Н.Н.** Методы оптимизации / Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М. –М.: Наука, 1978.

21. **Морозов В.В.** Исследование операций в задачах и упражнениях / Морозов В.В., Сухарев А.Г., Федоров В.В. – М.: Высшая школа, 1986.

22. **Пшеничный Б.Н.** Численные методы в экстремальных задача / Пшеничный Б.Н., Данилин Ю.М. – М.: Наука, 1986.

23. **Сухарев А.Г.** Курс методов оптимизации / Сухарев А.Г., Тимохов А.В., Федоров В.В. – М.: Наука, 1986.

24. **Хедли Дж.** Нелинейное и динамическое программирование. – М.: Мир, 1967.

25. **Химмельблау Д.** Прикладное нелинейное программирование. – М.: Мир, 1975.

26. **Шор Н.З.** Методы минимизации недифференцируемых функций и их приложениях. – К.: Наук. думка, 1979.

27. **Шор Н.З.** Квадратичные экстремальные задачи и недифференцируемая оптимизация / Шор Н.З., Стеценко С.И. – К.: Наук. думка, 1989.

28. **Юдин Д.Б.** Линейное программирование. Теория и конечные методы / Юдин Д.Б., Гольштейн Е.Г. – М.: Наука, 1969.

Навчальне видання

Олена Михайлівна Кісельова

Алла Євгенівна Шевельова

Чисельні методи оптимізації

Навчальний посібник

Редактор В.П. Пименов

Технічний редактор В.А. Усенко

Коректор В.П. Пименов

Підписано до друку 28.11.2008. Формат $60\times84\,{}^{1}\!/\!_{16}$. Папір друкарський. Друк плоский. Ум. друк. арк. 12,32. Ум. фарбовідб. 12,32. Обл. вид. арк. 14,0. Тираж 300 пр. Вид. № 1375.Зам. №

Видавництво Дніпропетровського національного університету, пр. Гагаріна, 72, м.Дніпропетровськ,49010.

Друкарня ДНУ, вул. Наукова, 5, м.Дніпропетровськ,49050.

ISBN 978-966-551-269-1