МАТЕМАТИЧНІ МЕТОДИ МОДЕЛЮВАННЯ
ТА ОПТИМІЗАЦІЇ ПРОЦЕСІВ

Лекція № 6. Аналітичні та чисельні методи оптимізації першого порядку (градієнтні методи)

План лекції:

Вступ.
1. Загальні відомості про функції багатьох змінних.
2. Похідні та диференціали функції багатьох змінних.
3. Оптимізація. Визначення оптимальних значень аргументів функції багатьох змінних.
   3.1. Поняття екстремуму, необхідні умови його існування
   3.2. Знаходження екстремуму функцій двох змінних
   3.3. Знаходження умовного екстремуму методом Лагранжа
   3.4. Найбільше і найменше значення функції в замкненій області
4. Метод градієнтного спуску із постійним кроком.
5. Метод найшвидшого градієнтного спуску.

Заключення.

Завдання на СРС.
1. Вивчити матеріал лекції.
2. Виконати вправи (в додатку до лекції): №1, a,c,d, №2, a, №3, b,c, №5, a, №6, a.

Вступ.

До методів оптимізації 1-го порядку відносяться аналітичні та чисельні методи, що використовують значення першої похідної від цільової функції.

Для оптимізації функцій однієї змінної $y = f(x)$, аналітичний метод оптимізації вивчався в курсі Вищої математики. Він полягає в тому, що функція в точці екстремуму має нульову похідну. Тому для пошуку екстремуму треба взяти похідну, прирівняти до нуля, рішити отримане рівняння та найти точу x^*, в якій функція приймає екстремум. Для того, щоб визначити тип екстремуму: максимум чи мінімум, треба взяти другу похідну та перевірити її знак в точці x^*. Якщо вона додатна, то екстремуму є мінімумом, якщо від'ємна – то максимумом.

У випадку умовної оптимізації, тобто пошуку екстремуму на заданому відрізку, треба перевіряти значення функції в локальних екстремумах в середині відрізка та значення функції на кінцях відрізка.

Сьогодні ми розглянемо аналітичні методи оптимізації аналогічні вищезазначеним, але для функцій кількох змінних. Саме такі функції найчастіше використовуються у практичній діяльності інженерів, менеджерів, бізнесменів та інших спеціалістів різних галузей.

До чисельних методів оптимізації першого порядку відносяться алгоритми, в яких в процесі пошуку екстремуму окрім інформації про саму функцію використовується інформація про похідні першого порядку. До групи таких методів відносяться різні градієнтні методи.

Градієнт функції в будь-якій точці показує напрямок найбільшого локального збільшення функції $f(x_1, x_2, \dots, x_n)$. Тому при пошуку мінімуму слід рухатися в напрямку, який є протилежним напрямку градієнта в даній точці, тобто в напрямку найшвидшого спуску.

До методів оптимізації першого порядку відносяться такі методи:

- метод найшвидшого спуску з постійним кроком;
- метод найшвидшого градієнтного спуску;
- метод покоординатного спуску;
- метод Гаусса-Зейделя;
- метод Флетчера-Рівса;
- метод Девідона-Флетчера-Пауелла;
- метод кубічної інтерполяції та інші.

1. Загальні відомості про функції багатьох змінних

При дослідженні процесів часто спостерігають одночасну зміну декількох величин і залежність однієї з них від інших.

Означення 1. Якщо змінна величина w залежить від n незалежних змінних x_1, x_2 **...** x_n то її називають **функцією цих змінних**, а функціональну залежність позначають так:

$$w = f\left(x_1, x_2, \dots, x_n\right), \text{ або } w = f\left(M\right), \text{ де точка } M \in E_n.$$

Незалежні змінні x_1, x_2 **...** x_n називаються ***аргументами***.

Означення 2. Сукупність усіх числових значень, які можуть приймати аргументи x_1, x_2, ... x_n і при яких функція $W = f\left(x_1, x_2, \dots x_n\right)$ приймає певні дійсні значення, називають **областю визначення функції**.

При знаходженні області визначення функції кількох змінних, що задана аналітично, доцільно керуватись такими правилами:

1. Вираз під коренем парного степеня повинен бути невід’ємним;
2. Вираз під знаком логарифма повинен бути додатним;
3. Вираз знаменника дробу не повинен дорівнювати нулю.
4. Модуль виразу, що стоїть під знаком arcsin або arccos, не більше 1 по модулю.

Точка або сукупність точок, в яких функція кількох змінних не визначена, ***називають розривами цієї функції.***

Якщо функція визначена для усіх x_1, x_2 **...** x_n з деякої області D та її межі dD, тоді кажуть, що функція визначена у замкненій області $D = D \cup dD$.

Приклад 1. Знайти область визначення функції

$$u = \sqrt{25 - (x+1)^2 - (y-2)^2 - z^2}\ .$$

Розв'язування. Задана функція в залежності від трьох змінних x, y та z. Вона приймає певні дійсні значення лише при умові

$$25 - (x+1)^2 - (y-2)^2 - z^2 \geq 0 \Rightarrow (x+1)^2 + (y-2)^2 + z^2 \leq 25.$$

Рівність $(x-a)^2 + (y-b)^2 + (z-c)^2 = R^2$ є рівнянням сфери з центром у точці $C(a,b,c)$ і радіусом R.

Отже, одержана нерівність означає, що областю визначення функції u буде куля радіуса 5 з центром в точці $C(-1,2,0)$. Нерівність нестрога, тому функція u визначена на сфері – межі цієї кулі. Отже, задана функція u визначена у замкненій області

$$\bar{D} = \left\{ (x+1)^2 + (y-2)^2 + z^2 \leq 25 \right\}.$$

Згідно з основними поняттями аналітичної геометрії, **функція двох змінних** $z = f(x,y)$ в тривимірному просторі **зображується поверхнею**. Кожна точка цієї поверхні M' має координати (x,y,z). Областю визначення функції $z = f(x,y)$ буде деяка область D площини xOy. Коли точка $M(x,y)$ пробігає область D, тоді точка $M'(x,y,z)$ пробігає поверхню S, рівняння якої $z = f(x,y)$.

Отже, функцію двох змінних x, y можна задати як функцію змінної точки M, що змінюється в області D, тобто $Z = f(M)$.

Аналогічно можна розглядати і функцію ***n*** аргументів, як функцію точки M, що переміщується в області D n-вимірного простору E_n, тобто $W = f(M)$, $M \in D \in E_n$.

Способи задання функції кількох змінних

Функцію однієї змінної можна задавати аналітично, таблично, графічно, мовно і за допомогою комп'ютерної програми. Функцію двох змінних $Z = f(x,y)$, крім цих способів, можна задавати ще й геометрично, за допомогою ліній рівня.

У табличному способі завдання функції $Z = f(x,y)$ використовують <u>таблицю</u> вигляду:

x_k \ y_k	y_1	y_2	...	y_n
x_1				
x_2				
...				

x_n				

У кожній клітинці вказують значення Z для відповідної пари (x,y).

Розглянемо геометричний спосіб задання функції. Нехай графіком функції $Z=f(x,y)$ буде поверхня, зображена на рис. 1. Неважко бачити, що різні точки цієї поверхні знаходяться на різній відстані від площини xOy.

Якщо придати Z постійні значення h_1, h_2,..., то одержимо в площині аргументів ліній $f(x,y)=h_1$ та $f(x,y)=h_2$ ..., які називають ***лініями рівня функції*** $f(x,y)$.

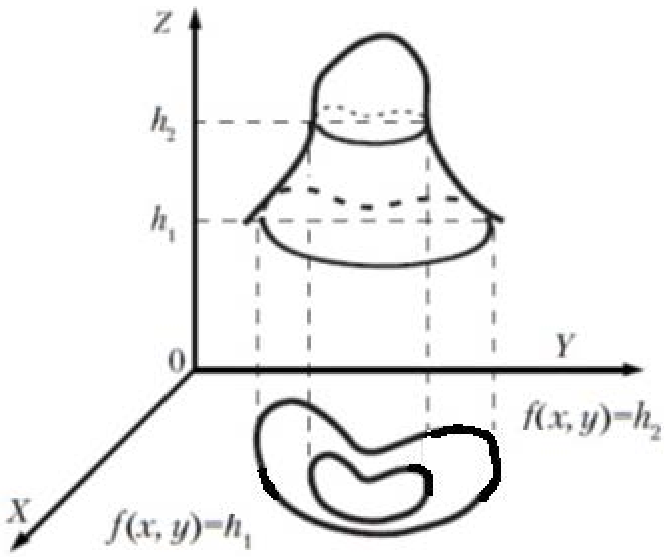

Рис. 1

Означення 3. Криві лінії L, що лежить у площині xOy і мають рівняння $f(x,y)=c$ (c – стала) називають **лініями рівня функції** $Z=f(x,y)$.

Іншими словами: лінія рівня – це множина усіх точок площини xOy, для яких функція $Z=f(x,y)$ приймає одне значення.

Приклад 2. Визначити лінії рівня функції $z=(x-2)^2+(y+3)^2$.

Розв'язування. Згідно з означенням 3 рівняння ліній рівня має вигляд

$$(x-2)^2+(y+3)^2=c.$$

Якщо надати c різні числові значення (наприклад, $c=4$, $c=9$, $c=16$ ...), то одержимо сукупність кіл з центром в точці $C(2,-3)$ з відповідними радіусами (наприклад, 2, 3, 4 ...).

Відмітимо, що лінії рівня широко використовуються в топографії. На топографічних картах нанесені лінії рівня, відстань між якими постійна і дорівнює

h. Величина h вказана на карті (наприклад, $h = 3$м) і дозволяє ефективно використовувати умови місцевості.

У випадку залежності функції від трьох та більше змінних найчастіше використовують аналітичний спосіб задання функції.

Границя та неперервність

Означення 4. Околом радіуса r точки $M_0(x_{10},\ x_{20},\ \ldots,\ x_{n0})$ називають сукупність усіх точок $M(x_1,\ x_2,\ \ldots,\ x_n)$ простору E_n, відстань яких до точки M_0 менше або дорівнює r, тобто виконується співвідношення

$$\sqrt{(x_1 - x_{10})^2 + (x_2 - x_{20})^2 + \ldots + (x_n - x_{n0})^2} \le r .$$

Означення 5. Число A називають **границею** функції $W = f(x_1,\ x_2,\ \ldots,\ x_n)$ (або $W = f(M)$) **в точці** $M_0(x_{10},\ x_{20},\ \ldots,\ x_{n0})$, якщо для будь-якого $\varepsilon > 0$ знайдеться число r таке, що для усіх точок $M(x_1,\ x_2,\ \ldots,\ x_n)$ з околу радіуса r точки M_0, відмінних від точки M_0, виконується нерівність

$$\left| f(x_1,\ x_2,\ \ldots,\ x_n) - A \right| < \varepsilon \text{ або } \left| f(M) - A \right| < \varepsilon .$$

Використовується **позначення**:

$$\lim_{\substack{x_1 \to x_{10} \\ x_2 \to x_{20} \\ \ldots \to \ldots \\ x_n \to x_{n0}}} f(x_1,\ x_2,\ \ldots,\ x_n) = A \text{ або } \lim_{M \to M_0} f(M) = A .$$

Означення 6. Функція $W = f(x_1,\ x_2,\ \ldots,\ x_n)$ $(W = f(M))$ називається **неперервною в точці** $M_0(x_{10},\ x_{20},\ \ldots,\ x_{n0})$, якщо вона визначена в цій точці і $\lim\limits_{M \to M_0} f(M) = f(M_0)$ незалежно від способу прямування точки M до точки M_0.

Функція, неперервна в кожній точці деякої області, називається ***неперервною в цій області***.

Якщо функція неперервна в області D та на її межі dD, тоді кажуть що вона неперервна в замкненій області $\bar{D} = D \cup dD$.

Для знаходження області неперервності функції багатьох змінних доцільно використовувати такі **властивості неперервних функцій:**

1) Області визначення та неперервності функцій співпадають.

2) функція, неперервна в замкненій області D, обмежена, тобто існують такі числа m та M, що виконується співвідношення $m \le f(x_1,\ x_2,\ \ldots,\ x_n) \le M$ для усіх $(x_1,\ x_2,\ \ldots,\ x_n) \in \bar{D}$.

2. Похідні та диференціали функції багатьох змінних

Якщо у функції декількох змінних $W = f\left(x_1, x_2, ..., x_n\right)$ змінна $x_k \left(k = 1, 2, ..., n\right)$ одержить частинний приріст Δx_k, а всі інші незалежні змінні зафіксувати, тоді функція одержить частинний приріст

$$\Delta_{x_k} w = f\left(x_1, x_2, ..., x_{k-1}, x_k + \Delta x_k, ..., x_n\right) - f\left(x_1, x_2, ..., x_{k-1}, x_k, ..., x_n\right)$$

за аргументом x_k.

Означення 7. Якщо існує границя $\lim\limits_{\Delta x_k \to 0} \dfrac{\Delta_{x_k} w}{\Delta x_k}$ незалежна від способу прямування $\Delta x_k \to 0$, тоді її називають **частинною похідною першого порядку функції** $W = f\left(x_1, x_2\ ..., x_n\right)$ **по змінній** $x_k \left(k = 1, 2, ..., n\right)$ і позначають

$$\frac{\partial w}{\partial x_k} \; \textit{або} \; \frac{\partial f\left(x_1, x_2, ..., x_k, ..., x_n\right)}{\partial x_k} \; \textit{або} \; w'_{x_k}.$$

Отже, за означенням частинна похідна буде

$$\frac{\partial w}{\partial x_k} = \lim_{\Delta x_k \to 0} \frac{\Delta_{x_k} w}{\Delta x_k}. \qquad (1)$$

При знаходженні частинної похідної по змінній x_k **усі інші аргументи слід вважати постійними величинами** і тому можна використовувати правила диференціювання та таблицю похідних функцій однієї змінної.

Для функції двох змінних $Z = f\left(x, y\right)$ можна надати геометричну та механічну інтерпретацію частинної похідної першого порядку, а саме:

– похідна $Z'_y \left(Z'_x\right)$ чисельно дорівнює тангенсу кута нахилу дотичної до кривої лінії, яка утворюється перетином поверхні $Z = f\left(x, y\right)$ площиною $x = x_0 \left(y = y_0\right)$;

– механічний зміст $Z'_y \left(Z'_x\right)$ – це швидкість зміни функції Z у напрямку осі $O_y \left(O_x\right)$, коли аргумент $x\left(y\right)$ не змінюється.

Приклад 3. Об’єм продажу нового продукту x залежить від часу t і витрат A підприємства на рекламу. Якщо t вимірювати тижнями, а A в гривнях, тоді ця залежність має вигляд

$$x = 200\left(5 - e^{-0,002A}\right)\left(1 - e^{-t}\right).$$

Знайти $\dfrac{\partial x}{\partial t}$, $\dfrac{\partial x}{\partial A}$ і вказати економічний зміст цих похідних при $t = 1$, $A = 400$.

Розв’язування. Маємо.

$$\frac{\partial x}{\partial t} = 200\left(5 - e^{-0,002A}\right)\left(1 - e^{-t}\right)'_t = 200\left(5 - e^{-0,002A}\right)e^{-t},$$

$$\frac{\partial x}{\partial A} = 200\left(1 - e^{-t}\right)\left(5 - e^{-0,002A}\right)'_A = 0,4\left(1 - e^{-t}\right)e^{-0,002A}.$$

При $t = 1$ та $A = 400$ одержимо

$$\left.\frac{\partial x}{\partial t}\right|_{\substack{t=1\\A=400}} = 200\left(5-e^{-0,8}\right)e^{-1} \approx 335,$$

$$\left.\frac{\partial x}{\partial A}\right|_{\substack{t=1\\A=400}} = 0,4\left(1-e^{-1}\right)e^{-0,8} \approx 0,11.$$

Частинна похідна x'_t характеризує швидкість зміни об'єму продажу нового продукту за тиждень, коли витрати на рекламу не змінюється.

Частинна похідна x'_A характеризує швидкість зміни об'єму продажу продукту при зміні суми витрат на рекламу і постійному t. При витратах на рекламу 400 гривень швидкість зростання об'єму продажу продукту за один тиждень буде 0,11.

Частинну похідну функції $u = f(x,y,z)$ **за напрямом вектора** $\vec{l} = (l_x, l_y, l_z)$ знаходять за формулою

$$\frac{\partial u}{\partial \vec{l}} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma, \tag{2}$$

де $\cos\alpha = \dfrac{l_x}{\left|\vec{l}\right|}$; $\cos\beta = \dfrac{l_y}{\left|\vec{l}\right|}$; $\cos\gamma = \dfrac{l_z}{\left|\vec{l}\right|}$.

Напрям найбільшої швидкості зміни функції $u = f(x,y,z)$ співпадає з напрямом вектора (його називають ***градієнтом функції*** u**)**

$$\operatorname{grad} u = \frac{\partial u}{\partial x}\vec{i} + \frac{\partial u}{\partial y}\vec{j} + \frac{\partial u}{\partial z}\vec{k}, \tag{3}$$

а величина цієї найбільшої швидкості дорівнює довжині цього вектора, тобто

$$\left|\operatorname{grad} u\right| = \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2}. \tag{4}$$

Приклад 4. Знайти величину найбільшої швидкості зміни функції $u = 7x^2y - \frac{7}{2}y^2z + \frac{14}{3}z^3$ в точці $M_0(1,0,9)$.

Розв'язування. Частинні похідні першого порядку в цьому випадку будуть

$$\frac{\partial u}{\partial x} = 14xy; \qquad \frac{\partial u}{\partial y} = 7x^2 - 7yz; \qquad \frac{\partial u}{\partial z} = 14z^2 - \frac{7}{2}y^2.$$

Величину найбільшої зміни заданої функції u в будь-якій точці знайдемо за формулою (4)

$$\left|\operatorname{grad} u\right| = \sqrt{(14xy)^2 + 7^2\left(x^2 - yz\right)^2 + \left(14z^2 - \frac{7}{2}y^2\right)^2} =$$

$$= 7\cdot\sqrt{4x^2y^2 + \left(x^2 - yz\right)^2 + \left(2z^2 - \frac{y^2}{2}\right)^2}.$$

Підставимо замість x, y, z координати точки M_0, тоді

$$\left|\operatorname{grad} u(M_0)\right| = 7\cdot\sqrt{4\cdot1\cdot0 + (1-0)^2 + 4\cdot81} = 7\sqrt{325} =$$

$$= 35\sqrt{13} \approx 35\cdot3,6 = 126 \quad (\text{î äèí èöü âèì ³ðó}).$$

Частинні похідні вищих порядків.

Означення 8. Частинну похідну першого порядку по змінній x_m від частинної похідної першого порядку функції по змінній x_k називають **частинною похідною другого порядку функції по змінним x_k та x_m** і позначають:

$$\frac{\partial^2 w}{\partial x_m \partial x_k} \text{ або } w''_{x_k x_m} \text{ при } k \neq m,$$

$$\frac{\partial^2 w}{(\partial x_k)^2} \text{ або } w''_{x_k x_k} \text{ при } k = m.$$

У випадку функції двох змінних $Z = f(x,y)$ маємо

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right); \quad \frac{\partial^2 z}{(\partial x)^2} = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right); \quad \frac{\partial^2 z}{(\partial y)^2} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right).$$

Якщо мішані частинні похідні другого порядку неперервні, тоді

$$\frac{\partial^2 x}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x},$$

тобто мішана частина похідна другого порядку не залежить від порядку диференціювання функції.

Аналогічно визначають частинні похідні порядку $k > 2$.

3. Оптимізація. Визначення оптимальних значень аргументів функції багатьох змінних.

3.1. Поняття екстремуму, необхідні умови його існування

Функція багатьох змінних $W = f(M)$, $M \in E_n$ **має максимум в точці** M_0, якщо $f(M_0) > f(M)$ для усіх точок M із достатньо малого околу точки M_0.

Функція $W = f(M)$, $M \in E_n$ **має мінімум в точці** M_0, якщо $f(M_0) < f(M)$ для усіх точок M із достатньо малого околу точки M_0.

Максимуми та мінімуми функції кількох змінних називають ***екстремумами функції***, а точку M_0, де функція має екстремум, називають ***точкою екстремуму функції***.

Теорема. ***(Необхідні умови існування екстремуму).*** Якщо функція $W = f(x_1, x_2, ..., x_n)$ має екстремум в точці $M_0(x_{10}, x_{20}, ..., x_{n0})$, то кожна частинна похідна першого порядку функції дорівнює нулю або не існує в цій точці.

Наслідок. *Точки, в яких* $\frac{\partial w}{\partial x_k}(k = 1, 2, ..., n)$ *не існують або дорівнюють нулю називають* ***критичними точками*** *або* ***підозрілими на екстремум.***

Приклад 5. Знайти критичні точки функції

$$z = x^2 - xy + y^2 + 3x - 2y + 1.$$

Розв'язання. Спочатку знайдемо частинні похідні першого порядку заданої функції двох змінних:

$$z'_x = 2x - y + 3;\ \ z'_y = -x + 2y - 2.$$

Ці похідні існують для усіх x та y, тому критичними будуть лише точки, де частинні похідні дорівнюють нулю, тобто

$$\begin{cases} z'_x = 0 \\ z'_y = 0 \end{cases} \Rightarrow \begin{cases} 2x - y = -3 \\ -x + 2y = 2 \end{cases}.$$

Остання система лінійна, неоднорідна, з двома невідомими. Розв'язуючи систему за правилом Крамера, одержимо:

$$x = -4/3;\ \ y = 1/3.$$

Отже, критичною точкою буде $M_0\left(-4/3,\ 1/3\right)$.

3.2. Знаходження екстремуму функцій двох змінних

Необхідні умови існування екстремуму функцій кількох змінних дозволяють знаходити лише критичні точки.

У випадку функції двох змінних за допомогою достатніх умов існування екстремуму можна перевірити кожну критичну точку та виявити, який саме екстремум існує в цій точці.

Теорема. ***(достатні умови існування екстремуму).*** Нехай в околі критичної точки $M_0(x_0,\ y_0)$ функція $Z = f(x,\ y)$ має неперервні частинні похідні до другого порядку включно,

$$a_{11} = \frac{\partial^2 f(x_0, y_0)}{(\partial x)^2} \quad a_{22} = \frac{\partial^2 f(x_0, y_0)}{(\partial y)^2} \quad a_{12} = \frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}$$

Тоді:

1) $f(x,\ y)$ ***має максимум, якщо*** $a_{11} \cdot a_{22} - a_{12}^2 > 0$ ***та*** $a_{11} < 0$***;***

2) $f(x,\ y)$ ***має мінімум, якщо*** $a_{11} \cdot a_{22} - a_{12}^2 > 0$ ***та*** $a_{11} > 0$***;***

3) $f(x,\ y)$ ***не має екстремуму, якщо*** $a_{11} \cdot a_{22} - a_{12}^2 < 0$***;***

4) ***якщо*** $a_{11} \cdot a_{22} - a_{12}^2 = 0$***, тоді екстремум в точці*** M_0 ***може існувати, а може і не існувати , тобто в цьому випадку треба використовувати іншу достатню ознаку.***

Приклад 6. Дослідити на екстремум функцію

$$z = x^2 - xy + y^2 + 3x - 2y + 1.$$

Розв'язування. У прикладі 5 для цієї функції знайдена критична точка $M_0\left(-4/3,\ 1/3\right)$. Застосуємо достатню умову. Маємо:

$$z''_{xx} = \frac{\partial}{\partial x}\left(2x - y + 3\right) = 2;\ z''_{yx} = \frac{\partial}{\partial y}\left(2x - y + 3\right) = -1;\ z''_{yy} = 2.$$

Тому $a_{11} \cdot a_{22} - a_{12}^2 = 2 \cdot 2 - \left(-1\right)^2 = 3 > 0$ та $a_{11} = 2 > 0$.

Згідно з другим твердженням теореми в точці $M_0\left(-4/3,\ 1/3\right)$ задана функція має мінімум:

$$z_{\min} = f\left(-\frac{4}{3},\ \frac{1}{3}\right) = \left(-\frac{4}{3}\right)^2 - \left(-\frac{4}{3}\right)\cdot\frac{1}{3} + \left(\frac{1}{3}\right)^2 + 3\left(-\frac{4}{3}\right) - 2\left(\frac{1}{3}\right) + 1 =$$
$$= \frac{16}{9} + \frac{4}{9} + \frac{1}{9} - 4 - \frac{2}{3} + 1 = -\frac{4}{3}.$$

3.3. Знаходження умовного екстремуму методом Лагранжа

Екстремум функції $Z = f\left(x,\ y\right)$ при виконанні умови $\varphi\left(x,\ y\right) = 0$ називають ***умовним екстремумом функції***.

Умовні екстремуми часто використовуються при дослідженні оптимізації багатьох економічних та соціальних проблем.

Для знаходження умовного екстремуму методом Лагранжа треба:

1) записати функцію Лагранжа вигляду

$$L\left(x,\ y,\ \lambda\right) = f\left(x,\ y\right) + \lambda\varphi\left(x,\ y\right);$$

2) знайти критичні точки $M_k\left(x_k,\ y_k,\ \lambda_k\right)$ функції Лагранжа, використовуючи необхідні умови існування екстремуму:

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \\ \dfrac{\partial L}{\partial y} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} \dfrac{\partial f}{\partial x} + \lambda\dfrac{\partial \varphi}{\partial x} = 0 \\ \dfrac{\partial f}{\partial y} + \lambda\dfrac{\partial \varphi}{\partial y} = 0 \\ \varphi\left(x, y\right) = 0 \end{cases};$$

3) перевірити в кожній критичній точці достатні умови існування екстремуму:

а) якщо в точці $M_k\left(x_k,\ y_k,\ \lambda_k\right)$ визначник третього порядку

$$\Delta\left(M_k\right) = \begin{vmatrix} 0 & \varphi'_x\left(M_k\right) & \varphi'_y\left(M_k\right) \\ \varphi'_x\left(M_k\right) & L''_{xx}\left(M_k\right) & L''_{xy}\left(M_k\right) \\ \varphi'_y\left(M_k\right) & L''_{xy}\left(M_k\right) & L''_{yy}\left(M_k\right) \end{vmatrix}$$

додатний, тоді точка M_k є точкою максимуму і

$$z_{\max} = f\left(M_k\right) = f\left(x_k,\ y_k\right);$$

б) якщо визначник $\Delta(M_k)<0$, тоді точка M_k є точкою мінімуму і

$$z_{\min}=f(M_k)=f(x_{k,}\ y_k).$$

Приклад 7. Знайти екстремум функції $z=xy$ при умові що $x^2+y^2=2$.

Розв'язування. Будемо шукати умовний екстремум з використанням функції Лагранжа

$$L=xy+\lambda\left(x^2+y^2-2\right).$$

Необхідні умови існування тепер мають вигляд

$$\begin{cases} y+2\lambda x=0 \\ x+2\lambda y=0 \\ x^2+y^2-2=0 \end{cases}.$$

Виключаючи з цієї системи λ, одержимо:

$$\begin{cases} \lambda=\dfrac{-y}{2x} \\ x+2y\left(\dfrac{-y}{2x}\right)=0 \\ x^2-y^2-2=0 \end{cases}\Rightarrow\begin{cases} \lambda=\dfrac{-y}{2x} \\ x^2-y^2=0 \\ x^2+y^2-2=0 \end{cases}\Rightarrow\begin{cases} \lambda=\dfrac{-y}{2x} \\ x^2=1 \\ x^2-y^2=0 \end{cases}.$$

Отже, критичними точками будуть:

$M_1(-1,-1),\ M_2(-1,\ 1),\ M_3(1,-1),\ M_4(1,\ 1).$

Для перевірки достатніх умов існування екстремуму запишемо визначник в довільній точці $M(x,y)$, враховуючи

$$\varphi'_x(M)=2x;\ \varphi'_y(M)=2y;\ L''_{xx}=2\lambda=-\frac{y}{x};\ L''_{yy}=-\frac{y}{x};\ L''_{xy}=1,$$

$$\Delta(M)=\begin{vmatrix} 0 & 2x & 2y \\ 2x & -\dfrac{y}{x} & 1 \\ 2y & 1 & -\dfrac{y}{x} \end{vmatrix}=12xy+4\frac{y^3}{x}.$$

Тепер можна знайти значення цього визначника в кожній критичній точці і використати достатні умови:

$$\Delta(M_1)=12(-1)\cdot(-1)+4\frac{(-1)^3}{-1}=12+4=16>0,$$

$$z_{\max}=z(M_1)=(-1)\cdot(-1)=1.$$

$$\Delta(M_2)=12(-1)\cdot(1)+4\frac{(1)^3}{-1}=-12-4=-16<0,$$

$$z_{\min}=z(M_2)=(-1)\cdot(1)=-1.$$

$$\Delta(M_3)=12\cdot1\cdot(-1)+4\frac{(-1)^3}{1}=-12-4=-16<0,$$

$$z_{\min}=z(M_3)=(-1)\cdot(1)=-1.$$

$$\Delta(M_4)=12+4=16>0,$$

$$z_{\max}=z(M_4)=1\cdot1=1.$$

3.4. Найбільше і найменше значення функції в замкненій області

Для знаходження найбільшого та найменшого значень функцій у замкненій області $\bar{D}$, які позначаються $\max\limits_{\bar{D}} f(x,y)$, $\min\limits_{\bar{D}} f(x,y)$, відповідно, треба знайти екстремальні значення функції в точках, що лежать всередині D та на межі області, і обрати найбільше та найменше значення.

Приклад 8. Знайти найбільше та найменше значення функції $z = x^2 y(4 - x - y)$ в трикутнику, обмеженому лініями $x = 0$, $y = 0$; $x + y = 6$.

Розв'язування. Спочатку знайдемо критичні точки всередині області:

$$z'_x = 2xy(4 - x - y) - x^2 y = xy(8 - 3x - 2y),$$

$$z'_y = x^2(4 - x - y) - x^2 y = x^2(4 - x - 2y).$$

Згідно з необхідними умовами існування екстремуму функції двох змінних маємо систему рівнянь

$$\begin{cases} xy(8 - 3x - 2y) = 0 \\ x^2(4 - x - 2y) = 0 \end{cases}.$$

Всередині області $x \neq 0$ та, $y \neq 0$ тому

$$\begin{cases} 3x + 2y = 8 \\ x + 2y = 4 \end{cases} \Rightarrow \begin{cases} x = 2 \\ y = 1 \end{cases}.$$

В критичній точці $M_1(2,\ 1)$ маємо $z(2,\ 1) = 4$.

Тепер проведемо дослідження функції на межі трикутника. На прямій $x + y = 6$ змінна $y = 6 - x$ і функція z приймає вигляд

$$z = x^2(6 - x)\cdot(4 - x + x - 6) = 2x^2(x - 6),\ \ x \in [0,\ 6].$$

Знайдемо найбільше та найменше значення цієї функції однієї змінної x на замкненому відрізку $[0,6]$: $z' = 6x^2 - 24x$.

Із рівності $z' = 0$ знаходимо: $6x(x - 4) = 0$, звідси випливає, що $x_1 = 4$ та $x_2 = 0$. Отже, $z(4) = -64$. При $x = 0$ та $x = 6$: $z(0) = 0$, $z(6) = 0$.

На прямій $y = 0$ маємо $z = 0$.

Отже, задана функція Z має найбільше значення в точці $M_1(2,\ 1)$ всередині області, найменше значення – в точці $M_2(4,\ 2)$ на межі області.

Найбільше значення $\max\limits_{\bar{D}} z = z(2,\ 1) = 4$;

Найменше значення $\min\limits_{\bar{D}} z = z(4,2) = -64$.

4. Метод градієнтного спуску із постійним кроком

Загальні терміни та визначення

Визначення 1. Градієнтом $\nabla f(x^k)$ неперервної та диференційованої функції $f(x)$ в точці x називається вектор-стовбець, елементами якого є частинні похідні першого порядку, що обчислені в даній точці:

$$\nabla f(x)=\begin{pmatrix}\dfrac{\partial f(x)}{\partial x_1}\\ \dfrac{\partial f(x)}{\partial x_2}\\ \vdots\\ \dfrac{\partial f(x)}{\partial x_n}\end{pmatrix}.$$

Визначення 2. Матрицею Гессе $H(x)$ двічі неперервно диференційованої в точці x функції $f(x)$ називається матриця частинних похідних другого порядку, що обчислюються в даній точці:

$$H(x)=\begin{pmatrix}\dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_n}\\ \dfrac{\partial^2 f(x)}{\partial x_2\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2\partial x_n}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial^2 f(x)}{\partial x_n\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2}\end{pmatrix}=\begin{pmatrix}h_{11} & h_{12} & \cdots & h_{1n}\\ h_{21} & h_{22} & \cdots & h_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ h_{n1} & h_{n2} & \cdots & h_{nn}\end{pmatrix},$$

де $h_{ij}=\dfrac{\partial^2 f(x)}{\partial x_i\,\partial x_j}$, $i,j=1,\dots,n$.

Постановка задачі

Нехай задана функція $f(x)$, обмежена знизу на множині R^n та має неперервні часткові похідні у всіх точках.

Потрібно знайти локальний мінімум функції $f(x)$ на множині допустимих рішень $X=R^n$, тобто знайти таку точку $x^*\epsilon\ R^n$, що

$$f(x^*)=\min_{x\epsilon R^n} f(x). \tag{1}$$

Стратегія пошуку

Стратегія рішення задачі полягає в побудові послідовних точок $\{x^k\}$, $k=0,1,\dots,$ таких, що $f(x^{k+1})<f(x^k)$, k=0,1,…, Точки послідовності $\{x^k\}$ обчислюються за правилом

$$x^{k+1}=x^k-t_k\,\nabla f(x^k),\quad k=0,1,\dots, \tag{2}$$

де точка x^0 задається користувачем; $\nabla f(x^k)$ – градієнт функції $f(x)$, обчислений в точці x^k; величина кроку t_k задається користувачем і залишається постійною до тих пір, поки функція спадає в точках послідовності, що контролюється шляхом перевірки виконання умови

$$f(x^{k+1}) - f(x^k) < 0 \text{ або } f(x^{k+1}) - f(x^k) < -\varepsilon\left\|\nabla f(x^k)\right\|^2,\ \ 0 < \varepsilon < 1. \qquad (3)$$

Побудова послідовності $\{x^k\}$ закінчується в точці x^k, для якої $\left\|\nabla f(x^k)\right\| < \varepsilon_1$, де ε_1 – задане мале позитивне число, або $k \geq M$, де M – граничне число ітерацій, або при дворазовому одночасному виконанні двох нерівностей:

$$\left\|x^{k+1} - x^k\right\| < \varepsilon_2, \qquad \left|f(x^{k+1}) - f(x^k)\right| < \varepsilon_2, \qquad (4)$$

де ε_2 – мале додатне число. Питання про те, чи може точка x^k розглядатися як знайдене наближення шуканої точки мінімума, вирішується шляхом проведення додаткового дослідження, яке описано нижче.

Алгоритм

Крок 1. Задати x^0, $0 < \varepsilon < 1$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M – граничне число ітерацій. Знайти градієнт функції в довільній точці $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^T$.

Крок 2. Покласти $k = 0$.

Крок 3. Обчислити $\nabla f(x^k)$.

Крок 4. Перевірити виконання критерію закінчення $\left\|\nabla f(x^k)\right\| < \varepsilon_1$:

а) якщо критерій виконується, то розрахунок закінчено, $x^* = x^k$;

б) якщо критерій не виконується, то перейти до кроку 5.

Крок 5. Перевірити виконання нерівності $k \geq M$:

а) якщо нерівність виконано, то розрахунок закінчено: $x^* = x^k$;

б) якщо ні, то перейти до кроку 6.

Крок 6. Задати величину кроку t_k.

Крок 7. Обчислити $x^{k+1} = x^k - t_k \nabla f(x^k)$.

Крок 8. Перевірити виконання умови

$f(x^{k+1}) - f(x^k) < 0$ (або $f(x^{k+1}) - f(x^k) < -\varepsilon\left\|\nabla f(x^k)\right\|^2$):

а) якщо умова виконана, то перейти до кроку 9;

б) якщо умова не виконана, покласти $t_k = \frac{t_k}{2}$ і перейти до кроку 7.

Крок 9. Перевірити виконання умов

$\left\|x^{k+1} - x^k\right\| < \varepsilon_2$, $\left|f(x^{k+1}) - f(x^k)\right| < \varepsilon_2$:

а) якщо обидві умови виконані при поточному значенні k і $k = k - 1$, то розрахунок закінчено, $x^* = x^{k+1}$;

б) якщо хоча б одна з умов не виконана, покласти $k = k + 1$ і перейти до кроку 3.

Геометрична інтерпретація методу для $n = 2$ наведена на рис. 1.

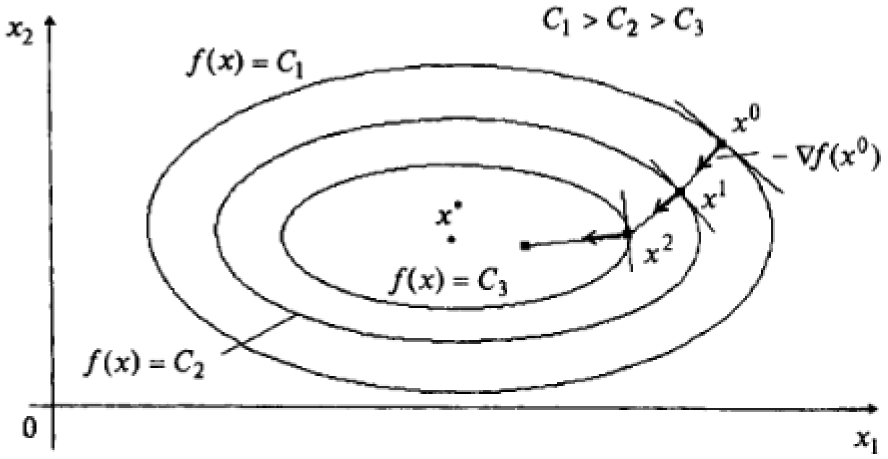

Рис. 1. Метод градієнтного спуску с постійним кроком

Збіжність

Твердження l. Нехай функція $f(x)$ диференційована і обмежена знизу на R^n, а її градієнт задовольняє умові Ліпшиця

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \forall x, y \in R^n, \text{де } L>0.$$

Тоді при довільній початковій точці $x^0 \in R^n$ для методу градієнтного спуску з постійним кроком маємо

$$\lim_{k\to\infty} \|\nabla f(x^k)\| = 0 \tag{5}$$

Зауваження 1.

1. Твердження l. гарантує збіжність послідовності $\{x^k\}$ до стаціонарної точки x^*, де $\nabla f(x^*) = 0$. Отже, знайдена в результаті застосування методу точка x^* потребує додаткового дослідження з метою її класифікації.

2. Метод градієнтного спуску гарантує збіжність послідовності $\{x^k\}$ до точки мінімуму для сильно опуклих функцій.

3. При вирішенні прикладів ітераційний процес підбору вдалої величини t_k, відбивається в індексації кроків 7 і 8. Перший індекс збігається з номером k, а другий з числом поділів поточної величини t_k навпіл.

Швидкість збіжності

Оцінки швидкості збіжності отримані тільки для сильно опуклих функцій, коли послідовність $\{x^k\}$ сходиться до точки мінімуму $f(x)$ зі швидкістю геометричної прогресії:

$$f(x^k) - f(x^*) \leq q^k\big(f(x^0) - f(x^*)\big), \qquad \|x^k - x^*\| \leq C\big(\sqrt{q}\big)^k,$$

де q ϵ (0,1), C > 0 – константи.

Рекомендації щодо аналізу знайденої точки x^k

1. Використовуючи алгоритм градієнтного спуску з постійним кроком, знайти точку x^k, в якій виконано принаймні один з критеріїв закінчення розрахунків.

2. Провести аналіз точки x^k з метою встановити, чи є точка x^k знайденим наближенням рішення задачі. Процедура аналізу визначається наявністю у функції *f(x)* безперервних других похідних. Якщо *f(x)* ϵ C^2. то слід провести перевірку виконання достатніх умов мінімуму: $H(x^*) > 0$, де $H(x^*)$ – матриця Гессе. Якщо

$H(x^k) > 0$. то точка x^k є знайдене наближення шуканої точки x^*. Якщо *f(x)* ϵ C^1, то слід провести перевірку функції *f(x)* на випуклість в Q-околиці точки x^k, використовуючи критерій опуклості для функцій *f(x)* ϵ C^1: функція f *(x)* опукла (строго опукла) в тому і тільки в тому випадку якщо
$f(x+y) \geq f(\text{x}) + (\nabla f(\text{x}), y), \forall x, y \in Q;\ f(x+y) > f(\text{x}) + (\nabla f(\text{x}), y)$. Якщо функція $f(\text{x})$ опукла (строго випукла), то x^k є знайдене наближення точки x^*.

Зауваження 2. Якщо потрібно знайти глобальний мінімум функції $f(x)$, то для строго опуклою $f(x)$ рішення цієї задачі аналогічно пошуку локального мінімуму функції. У випадку, коли $f(x)$ має декілька локальних мінімумів, пошук глобального мінімуму здійснюється в результаті перебору всіх локальних мінімумів.

Приклад 1. Знайти локальний мінімум функції

$$f(x) = 2x_1^2 + x_1x_2 + x_2^2$$

I. Визначення точки x^k, у якій виконаний принаймні один із критеріїв закінчення розрахунків.

1. Задамо $x^0, \varepsilon_1, \varepsilon_2, M$: $x^0 = (0{,}5; 1)^T, \varepsilon_1 = 0{,}1;\ \varepsilon_2 = 0{,}15; M = 10$. Знайдемо градієнт функції в довільній точці $\nabla f(x) = (4x_1 + x_2; x_1 + 2x_2)^T$.

2. Покладемо $k = 0$.

3^0. Обчислимо $\nabla f(x^0)$: $\nabla f(x^0) = (3; 2{,}5)^T$.

4^0. Обчислимо $\|\nabla f(x^0)\|$: $\|\nabla f(x^0)\| = 3{,}9 > 0{,}1$.
Переходимо до кроку 5.

5^0. Перевіримо умову $k \geq M$: $k = 0 < 10 = M$. Переходимо до кроку 6.

6^0. Задамо t_0 =0,5.

7^0. Обчислимо x^1: $x^1 = (0{,}5; 1)^T - 0{,}5(3; 2{,}5)^T = (-1; -0{,}25)^T; f(x^1) = 2{,}31$.

8^0. Порівняємо $f(x^1)$ із $f(x^0)$ = 2. Маємо $f(x^1) > f(x^0)$Висновок: умова $f(x^{k+1}) < f(x^k)$ для *k=0* не виконується. Задамо t^0 = 0,25 , переходимо до повторення кроків 7. 8.

7^{01} . Обчислимо x^1: $x^1 = (0{,}5; 1)^T - 0{,}25(3; 2{,}5)^T = (-0{,}25; 0{,}375)^T; f(x^1) = 0{,}171$.

8^{01} . Порівняємо $f(x^1)$ та $f(x^0)$. Висновок: $f(x^1) < f(x^0)$. Переходимо до кроку 9.

9 °. Обчислимо $\|x^1 - x^0\|$ та $|f(x^1) - f(x^0)|$:
$\|x^1 - x^0\| = 0{,}976 > 0{,}15$; $|f(x^1) - f(x^0)| = 1{,}829 > 0{,}15$. Висновок: вважаємо $k = 1$ і переходимо до кроку 3.

3^1. Обчислимо $\nabla f(x^1)$: $\nabla f(x^1) = (-0{,}625; 0{,}51)^T$.

4^1. Обчислимо $\|\nabla f(x^1)\|$: $\|\nabla f(x^1)\| = 0{,}81$ Переходимо до кроку 5.

5^1. Перевіримо умову $k \geq M$: $k = 1 < 10 = M$. Переходимо до кроку 6.

6^1. Задамо t_1 =0,25.

7^1. Обчислимо x^2: $x^2 = (-0{,}25; 0{,}375)^T - 0{,}25(-0{,}625; 0{,}5)^T = (-0{,}094; 0{,}25)^T; f(x^2) = 0{,}056$.

8^1 . Порівняємо $f(x^2)$ та $f(x^1)$. Висновок: $f(x^2) < f(x^1)$. Переходимо до кроку 9.

9^1. Обчислимо $\|x^2 - x^1\|$ та $|f(x^2) - f(x^1)|$:
$\|x^2 - x^1\| = 0{,}2 > 0{,}15$; $|f(x^2) - f(x^1)| = 0{,}115 < 0{,}15$.

Висновок: вважаємо $k = 2$ і переходимо до кроку 3.

3^2. Обчислимо $\nabla f(x^2)$: $\nabla f(x^2) = (-0{,}126; 0{,}406)^T$.

4^2. Обчислимо $\|\nabla f(x^2)\|$: $\|\nabla f(x^2)\| = 0{,}425 > 0.1$ Переходимо до кроку 5.

5^2. Перевіримо умову $k \geq M$: $k = 2 < 10 = M$. Переходимо до кроку 6.

6^2. Задамо $t_2 = 0{,}25$.

7^2. Обчислимо x^3: $x^3 = (-0{,}094; 0{,}25)^T - 0{,}25(-0{,}126; 0{,}406)^T = (-0{,}063; 0{,}15)^T; f(x^3) = 0{,}021$.

8^2 . Порівняємо $f(x^3)$ та $f(x^2)$. Висновок: $f(x^3) < f(x^2)$. Переходимо до кроку 9.

9^2. Обчислимо $\|x^3 - x^2\|$ та $|f(x^3) - f(x^2)|$:

$\|x^3 - x^2\| = 0{,}105 < 0{,}15$; $\quad |f(x^3) - f(x^2)| = 0{,}035 < 0{,}15$.

Висновок: вважаємо $k = 3$ і переходимо до кроку 3.

3^3. Обчислимо $\nabla f(x^3)$: $\nabla f(x^3) = (-0{,}102; 0{,}237)^T$.

4^3. Обчислимо $\|\nabla f(x^3)\|$: $\|\nabla f(x^3)\| = 0{,}257 > 0.1$ Переходимо до кроку 5.

5^3. Перевіримо умову $k \geq M$: $k = 3 < 10 = M$. Переходимо до кроку 6.

6^3. Задамо $t_3 = 0{,}25$.

7^3. Обчислимо x^4: $x^4 = (-0{,}063; 0{,}15)^T - 0{,}25(-0{,}102; 0{,}237)^T = (-0{,}038; 0{,}091)^T; f(x^4) = 0{,}0076$.

8^3 . Порівняємо $f(x^4)$ та $f(x^4)$: $f(x^4) < f(x^3)$.

9^3. Обчислимо $\|x^4 - x^3\|$, $|f(x^4) - f(x^3)|$:

$\|x^4 - x^3\| = 0{,}064 < 0{,}15$; $\quad |f(x^4) - f(x^3)| = 0{,}015 < 0{,}15$.

Умови $\left\|x^{k+1} - x^k\right\| < \varepsilon_2$, $\left|f(x^{k+1}) - f(x^k)\right| < \varepsilon_2$ виконані при $k = 2{,}3$. Розрахунок закінчено. Знайдена точка $x^4 = (-\,0{,}038;\ 0{,}091)^T$; $f(x^4) = 0{,}0076$.

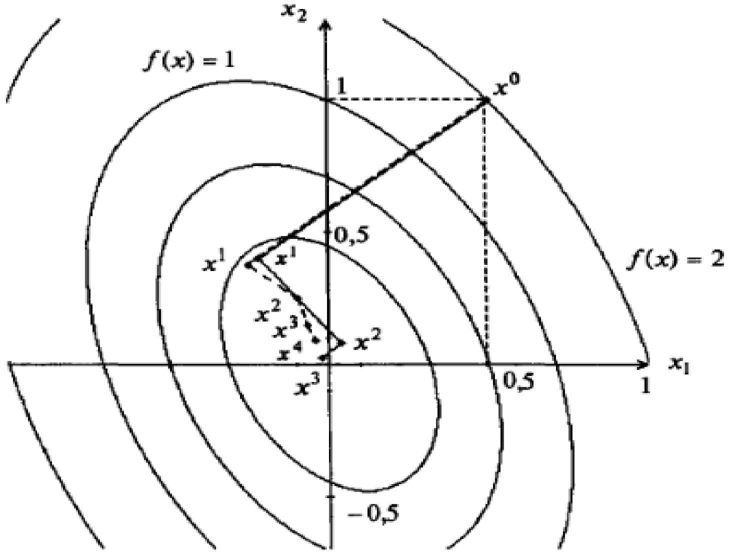

Рис. 2.

II. Аналіз точки x^4.

Функція $f(x) = 2x_1^2 + x_1x_2 + x_2^2$ є двічі диференційованою, тому проведемо перевірку достатніх умов мінімуму в точці x^4. Для цього проаналізуємо матрицю Гессе $H = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$. Матриця постійна і є позитивно визначеною (тобто $H > 0$), оскільки обидва її кутових мінори $\Delta_1 = 4$ і $\Delta_2 = 7$ позитивні. Отже, точка $x^4 = (-0{,}038;\ 0{,}091)^T$ є знайдене наближення точки локального мінімуму $x^* = (0;0)^T$ а значення $f(x^4) = 0{,}0076$ є знайдене наближення значення $f(x^*) = 0$. Зауважимо, що умова $H > 0$, є одночасно умова строгої опуклості функції $f(x) = 2x_1^2 + x_1x_2 + x_2^2$ на R^2 .

Отже, $x^4 = (-0{,}038;\ 0{,}091)^T$, $f(x^4) = 0{,}0076$ є знайдені наближення точки глобального мінімуму $f(x)$ і її найменшого значення на R^2.

5. Метод найшвидшого градієнтного спуску

Постановки задачі

Нехай дана функція $f(x)$, обмежена знизу на множині R^2 і має безперервні часткові похідні у всіх її точках.

Потрібно знайти локальний мінімум функції $f(x)$ на множині допустимих рішень $X = R^2$, тобто знайти таку точку $x^* \epsilon R^n$, що

$$f(x^*) = \min_{x \epsilon R^n} f(x).$$

Стратегія пошуку

Стратегія рішення задачі полягає в побудові послідовних точок $\{x^k\}, k = 0,1, \ldots$, таких, що $f(x^{k+1}) < f(x^k)$, k=0,1,… . Точки послідовності $\{x^k\}$ обчислюються за правилом

$$x^{k+1} = x^k - t_k \nabla f(x^k),$$

де точка x^0 задається користувачем; величина кроку t_k визначається для кожного значення k з умови

$$\varphi(t_k) = f\left(x^k - t_k \nabla f(x^k)\right) \to \min_{t_k}.$$

Рішення завдання може здійснюватися з використанням необхідної умови мінімуму $\frac{\partial^2 \varphi}{\partial {t_k}^2} = 0$ з подальшою перевіркою достатньої умови мінімуму $\frac{\partial^2 \varphi}{\partial t_k^2} > 0$. Такий шлях може бути використаний або при досить простій функції, що мінімізується $\varphi(t_k)$, або при попередній апроксимації досить складної функції $\varphi(t_k) = f\left(x^k - t_k \nabla f(x^k)\right)$ поліномом $P(t_k)$ (як правило, другого або третього ступеня), і тоді умова $\frac{\partial \varphi}{\partial t_k} = 0$ замінюється умовою $\frac{\partial P}{\partial t_k}$=0, а умова $\frac{\partial^2 \varphi}{\partial t_k^2} > 0$ - умовою $\frac{\partial^2 P}{\partial t_k^2} > 0$.

Інший шлях вирішення задачі пов'язаний з використанням чисельних методів, коли відшукується

$$\min_{t_k \in [a,b]} \varphi(t_k) = \min_{t_k \in [a,b]} f\left(x^k - t_k \nabla f(x^k)\right).$$

Границі інтервалу [a,b] задаються користувачем. При цьому ступінь близькості знайденого значення t_k до оптимального значення ${t_k}^*$. задовольняючому умовам $\frac{\partial \varphi}{\partial t_k} = 0, \frac{\partial^2 \varphi}{\partial t_k^2} > 0$ залежить від задання інтервалу [a,b] і точності методів одномірної мінімізації.

Побудова послідовності $\{x^k\}, k = 0,1, \ldots$, закінчується в точці x^k, для якої $\left\|\nabla f(x^k)\right\| < \varepsilon_1$. де ε_1 - задане число, або, якщо , $k \geq M, M$ -граничне число ітерацій, або при дворазовому одночасному виконанні нерівностей $\left\|x^{k+1} - x^k\right\| < \varepsilon_2$, $\left|f(x^{k+1}) - f(x^k)\right| < \varepsilon_2$, де ε_2- мале додатне число. Питання про те, чи може точка x^k розглядатися як знайдене наближення шуканої точки локального мінімуму x^*, вирішується шляхом додаткового дослідження.

Алгоритм

Крок 1. Задати x^0, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M - граничне число ітерацій. Знайти градієнт функції в довільній точці $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right)^T$.

Крок 2. Покласти $k = 0$.

Крок 3. Обчислити $\nabla f(x^k)$.

Крок 4. Перевірити виконання критерію закінчення $\|\nabla f(x^k)\| < \varepsilon_1$:

а) якщо критерій виконаний, розрахунок закінчений, $x^* = x^k$;

б) якщо критерій не виконано, то перейти до кроку 5.

Крок 5. Перевірити виконання нерівності $k \geq M$:

а) якщо нерівність виконано, то розрахунок закінчено: $x^* = x^k$;

б) якщо ні, то перейти до кроку 6.

Крок 6. Обчислити величину кроку t_k^* з умови

$$\varphi(t_k) = f\left(x^k - t_k \nabla f(x^k)\right) \to \min_{t_k}.$$

Крок 7. Обчислити $x^{k+1} = x^k - t_k^* \nabla f(x^k)$.

Крок 8. Перевірити виконання умов

$$\|x^{k+1} - x^k\| < \varepsilon_2, \ |f(x^{k+1}) - f(x^k)| < \varepsilon_2$$

а) якщо обидві умови виконані при поточному значенні k і $k = k - 1$, то розрахунок закінчено, $x^* = x^{k+1}$;

б) якщо хоча б одна з умов не виконана, покласти $k = k + 1$ і перейти до кроку 3.

Геометрична інтерпретація методу для n = 2 наведена на рис. 3.

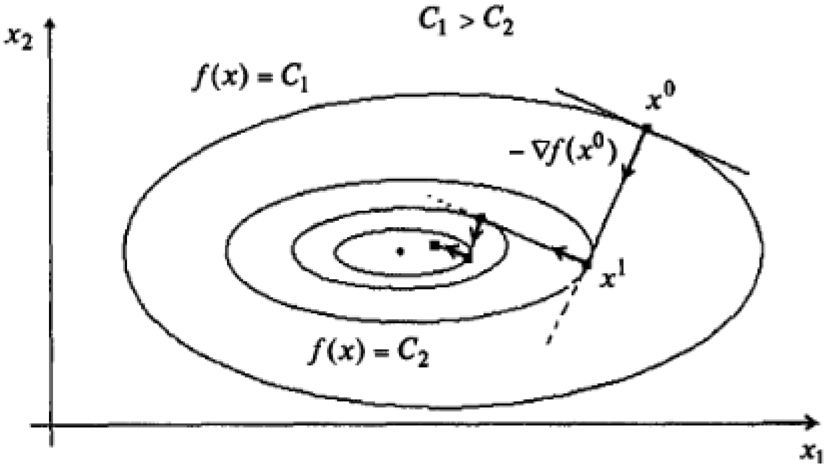

Рис. 3

Збіжність

Твердження 2. *Нехай функція $f(x)$ задовільняє умовам твердження 6.1. Тоді при довільній початковій точці $x^0 \in R^n$ для методу найшвидшого градієнтного спуску маємо $\|\nabla f(x^k)\| \to 0$ при $k \to \infty$.*

Зауваження 3.

1. Твердження гарантує збіжність послідовності $\{x^k\}$ до ста-стаціонарної точки x^*, де $\nabla f(x^*) = 0$. Отже, знайдена в результаті застосування методу точка x^* потребує додаткового дослідження з метою її класифікації.

2. Метод найшвидшого спуску гарантує збіжність послідовності $\{x^k\}$ до точки мінімуму для сильно опуклих функцій.

Швидкість збіжності

Оцінки швидкості збіжності отримані тільки для сильно опуклих функцій, коли послідовність $\{x^k\}$ сходиться до точки мінімуму функції $f(x)$ зі швидкістю геометричної прогресії (лінійна збіжність):

$$\left\|x^{k+1} - x^k\right\| \leq \frac{M-m}{M+m}\left\|x^k - x^*\right\|,$$

де M і m - оцінки найбільшого і найменшого власних значень матриці $H(x)$ функції $f(x)$.

Зауваження 4.

1. Процедура вирішення завдання збігається з описаною в п.1.

2. Щодо процедури пошуку глобального мінімуму функції $f(x)$ залишається справедливим зауваження 2.

Приклад 2. Знайти локальний мінімум функції

$$f(x) = 2x_1^2 + x_1x_2 + x_2^2$$

I. Визначення точки x^k, у якій виконаний принаймні один із критеріїв закінчення розрахунків.

1. Задамо $x^0, \varepsilon_1, \varepsilon_2, M$: $x^0 = (0{,}5; 1)^T, \varepsilon_1 = 0{,}1;\ \varepsilon_2 = 0{,}15; M = 10$. Знайдемо градієнт функції в довільній точці $\nabla f(x) = (4x_1 + x_2; x_1 + 2x_2)^T$.

2. Покладемо $k = 0$.

3^0. Обчислимо $\nabla f(x^0)$: $\nabla f(x^0) = (3; 2{,}5)^T$.

4^0. Обчислимо $\|\nabla f(x^0)\|$: $\|\nabla f(x^0)\| = 3{,}9 > 0{,}1$.

Переходимо до кроку 5.

5^0. Перевіримо умову $k \geq M$: $k = 0 < 10 = M$. Переходимо до кроку 6.. 6^0.

Наступна точка знаходиться за формулою $x^1 = x^0 - t_0\nabla f(x^0) = (0{,}5; 1)^T - t_0(3; 2{,}5)^T = (0{,}5 - 3t_0; 1 - 2{,}5t_0)^T$.

Підставимо отримані вирази $x_1^1 = 0{,}5 - 3t_0, x_2^1 = 1 - 2{,}5 \cdot t_0$ для координат $f(x)$: $\varphi(t_0) = 2 \cdot (0{,}5 - 3t_0)^2 + (0{,}5 - 3t_0) \cdot (1 - 2{,}5t_0)^2$. Знайдемо мінімум функції $\varphi(t_0)$ по t_0 за допомогою необхідних умов безумовного екстремуму:

$\frac{\partial\varphi(t_0)}{\partial t_0} = 4 \cdot (0{,}5 - 3t_0) \cdot (-3) \cdot (1 - 2{,}5t_0) + (-2{,}5) \cdot (0{,}5 - 3t_0) + 2 \cdot (1 - 2{,}5t_0) \cdot (-2{,}5) = -15{,}25 + 63{,}25 \cdot t_0 = 0$.

Звідси $t_0^* = 0{,}24$.

Так як $\frac{\partial^2\varphi(t_0)}{\partial t_0^2} = 63{,}25 > 0$, то знайдене значення кроку забезпечує мінімум функції $\varphi(t_0)$ по t_0.

Зауважимо, що можна отримати формулу для обчислення найкращої величини кроку t_k^* на будь якій ітерації з умови

$$\varphi(t_k) = f\left(x^k - t_k\nabla f(x^k)\right) \to \min_{t_k}.$$

Маємо $\nabla f(x^k) = \left(4x_1^k + x_2^k; x_1^k + x_2^k\right)^T; x^k - t_k\nabla f(x^k) = \left[x_1^k - t_k\left(4x_1^k + x_2^k\right); x_2^k - t_k\left(x_1^k + x_2^k\right)\right]^T$,

$$\varphi(t_k) = 2\left(x_1^k - t_k\left(4x_1^k + x_2^k\right)\right)^2 + \left(x_1^k - t_k\left(4x_1^k + x_2^k\right)\right)\left(x_2^k - t_k\left(x_1^k + x_2^k\right)\right) + \left(x_2^k - t_k\left(x_1^k + x_2^k\right)\right)^2$$

З умови $\frac{\partial\varphi}{\partial t_k} = 0$ отримуємо

$$t_k^* = \frac{\left(4x_1^k + x_2^k\right)^2 + \left(x_1^k + 2x_2^k\right)^2}{4\left(4x_1^k + x_2^k\right)^2 + 2\left(4x_1^k + x_2^k\right)\left(x_1^k + 2x_2^k\right) + 2\left(x_1^k + 2x_2^k\right)^2}$$

Визначимо t_0^*: t_0^*= 0,24.

7^0. Знайдемо $x^1 = x^0 - t_0\nabla f(x^0) = (0{,}5; 1)^T - 0{,}24(3; 2{,}5)^T =$
$= (-0{,}22; 0{,}4)^T$.

8^0. Обчислимо $\|x^1 - x^0\|$: $\|x^1 - x^0\|$ = 0,937> 0,15.

Обчислимо $|f(x^1) - f(x^0)|$: $|f(x^1) - f(x^0)|$ =1,83> 0,15. Висновок: вважаємо $k = 1$ і переходимо до кроку 3.

3^1. Обчислимо $\nabla f(x^1)$: $\nabla f(x^1) = (-0{,}48; 0{,}58)^T$.

4^1. Обчислимо $\|\nabla f(x^1)\|$: $\|\nabla f(x^1)\| = 0{,}752 > 0{,}1$.

5^1. Перевіримо умову $k \geq M$: $k = 1 < 10 = M$

6^1. Обчислимо t_1^*: t_1^*=0,546 див. п.6^0).

7^1. Знайдемо $x^2 = x^1 - t_1^*\nabla f(x^1)$:

$x^2 = (-0{,}22; 0{,}4)^T - 0{,}546(-0{,}48; 0{,}58)^T = (0{,}04; 0{,}08)^T$.

8^1. Обчислимо $\|x^2 - x^1\|$, $|f(x^2) - f(x^1)|$:

$\|x^2 - x^1\|$=0,41>0,15; $|f(x^2) - f(x^1)|$=0,156>0,15

Вважаємо k = 2 і переходимо до кроку 3.

3^2. Обчислимо $\nabla f(x^2)$: $\nabla f(x^2) = (0{,}24; 0{,}2)^T$.

4^2. Обчислимо $\|\nabla f(x^2)\|$: $\|\nabla f(x^2)\| = 0{,}312 > 0{,}1$.

5^2. Перевіримо умову $k \geq M$: $k = 2 < 10 = M$

6^2. Обчислимо t_2^*: t_2^*=0,24 див. п.6^0).

7^2. Знайдемо $x^3 = x^2 - t_2^*\nabla f(x^2)$:

$x^3 = (-0{,}04; 0{,}08)^T - 0{,}24(0{,}24; 0{,}2)^T = (-0{,}0176; 0{,}032)^T$.

8^2. Обчислимо $\|x^3 - x^2\|$, $|f(x^3) - f(x^2)|$:

$\|x^3 - x^2\|$=0,0749<0,15; $|f(x^3) - f(x^2)|$=0,116<0,15

Вважаємо k = 3 і переходимо до кроку 3.

3^3. Обчислимо $\nabla f(x^3)$: $\nabla f(x^3) = (-0{,}012; -0{,}0816)^T$.

4^3. Обчислимо $\|\nabla f(x^3)\|$: $\|\nabla f(x^3)\| = 0{,}082 < 0{,}1$.. Розрахунок закінчено

Знайдена точка x^3 = $(-\,0{,}0176;\ 0{,}032)^T$ $f(x^3)$ =0,00127. На рис. 3 отримані точки виділені і з'єднані суцільною лінією.

II. Аналіз точки x^3. У прикладі 1 було показано, що функція $f(x)$ є строго опуклою і, отже, точка x^3 є знайденим наближенням точки глобального мінімуму x^*.

Заключення.

В питаннях управління підприємствами зв'язку, а також в інших галузях, в бізнесі, менеджменті, аудиті використовуються функцій кількох змінних.

Сьогодні ми вивчили функції багатьох змінних вигляду $f(x, y, z, t)$, способи їх задання, області визначення та значень, границю та неперервність, похідні функції та застосування їх до розв'язування різних задач. А саме, як визначати максимальні та мінімальні значення функцій багатьох змінних вигляду $f(x, y, z, t)$, а також вирішувати задачі оптимізації, коли цільова функція задана як функція багатьох змінних.

Вивчені теоретичні положення складають основу аналітичних методів оптимізації 1-го порядку. Вони можуть бути застосовані тільки у випадку, коли цільова функція задана аналітично, є неперервною в заданій області та має першу частинну похідну по всім аргументам.

У випадку алгоритмічного задання функції, доцільно використовувати чисельні методи, з яких слід відзначити найбільш розповсюджені градієнтні методи.

На лекції були розглянуті метод градієнтного спуску із постійним кроком та метод найшвидшого градієнтного спуску. Це є найрозповсюдженими методами оптимізації першого порядку. Суть цих методів полягає в ітераційному русі в напрямку антиградієнта та поступовому наближені до мінімуму функції.

Завідувач кафедри вищої математики,
математичного моделювання та фізики
кандидат фізико-математичних наук, доцент І.В. Замрій

Додаток до лекції №6 з МММОП

Задачі для самостійного рішення

1. Знайти всі похідні другого порядку від функції:

a) $z = \dfrac{2x}{y+1}$; b) $z = \sqrt{x^2 + y^2}$; c) $z = 2x^2 y - 3y^2$;

d) $z = 5x^3 + 3y^4 + 10$; e) $z = x \ln y - y^2 \ln x$; f) $z = e^{2x+3y}$;

g) $z = (3x + 2y)^5$.

2. Знайти похідні

a) Знайти похідні z'''_{yxx} та $\dfrac{\partial^4 z}{\partial x (\partial y)^3}$ – функції $z = 3x^4 y^5$.

b) Знайти похідні z'''_{xyy} та z'''_{xxy} – функції $z = xe^{2x+y^2}$.

3. Для заданої функції знайти похідну за напрямом $\vec{l}$ та градієнт в точці M_0:

a) $u = \ln\sqrt{2x^2 + 2y^2}$; $\vec{l} = \left(\dfrac{\sqrt{3}}{2}, \dfrac{1}{2}\right)$, $M_0(1, -1)$;

b) $z = 2x^2 - 3y^2$; $\vec{l} = \left(\dfrac{\sqrt{3}}{2}, \dfrac{1}{2}\right)$, $M_0(0, -2)$;

c) $u = x^3 + y^3 - 3xy$; $\vec{l} = \left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right)$, $M_0(2, 1)$;

d) $u = 3x^2 + 2y^3 + z^2$; $\vec{l} = \left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}, \dfrac{\sqrt{2}}{2}\right)$, $M_0(2, -2, 1)$.

4. Знайти повний диференціал функції

a) $z = e^{\frac{x}{y}}$; b) $u = xy + 2yz - z^2x$; c) $u = 3x^2 + 4y^2 + z^2$;

d) $w = x + 2y + 5\left(x^2 + y^2 - 3\right)$; e) $z = \sqrt{1 - \frac{x^2}{9} + \frac{y^2}{4}}$;

f) $w = \sqrt{y^2 - 4x}$; g) $u = \ln\left(4 - x^2 - y^2\right)$.

5. Знайти екстремуми функцій

a) $z = x^2 - xy + y^2 + 9x - 6y + 20$; b) $z = e^{x-y} \cdot \left(x^2 - 2y^2\right)$;

c) $z = e^{\frac{x}{2}} \cdot (x + y^2)$; d) $u = \frac{8}{x} + \frac{x}{y} + y$;

e) $w = x^3 + y^3 - 3xy$; f) $u = x^4 + y^4 - x^2 - 2xy - y^2$;

g) $z = 3x^2 - x^3 + 3y^2 + 4y$; h) $z = x^3 + 3xy^2 - 15x - 12y$.

6. Знайти умовні екстремуми функцій

a) $z = x + y$ при $\frac{1}{x^2} + \frac{1}{y^2} = \frac{1}{2}$;

b) $u = \frac{1}{x} + \frac{1}{y}$ при $3x + y - 2 = 0$;

c) $z = x^2 + y^2 - xy + x + y - 4$ при $x + y + 3 = 0$;

d) $w = xy^2$ при $x + 2y = 1$;

e) $u = 2x + y - 2z$ при $x^2 + y^2 + z^2 = 36$.

7. Знайти найбільше та найменше значення функції в області *D*

a) $z = 1 + x + 12y$; $D = \left\{x \ge 0;\ y \ge 0;\ x + y \le 1\right\}$;

b) $z = x^2y$; $D = \left\{x^2 + y^2 \le 1\right\}$.