Методи оптимізації. Лекція 08.04.2022

Опуклі функції та їх властивості

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

$$f\left(\lambda x_1 + (1-\lambda) x_2\right) \le \lambda f(x_1) + (1-\lambda) f(x_2) \tag{1}$$

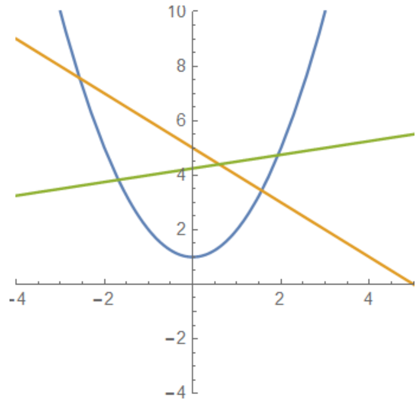
для усіх $x_1, x_2 \in X$ та $\lambda \in [0,1]$.

Геометричний сенс нерівності (1): функція $f(x)$ опукла на опуклій множині X, якщо для будь-яких точок $x_1, x_2 \in X$ графік функції лежить не вище хорди, що з'єднує точки $(x_1, f(x_1))$ та $(x_2, f(x_2))$.

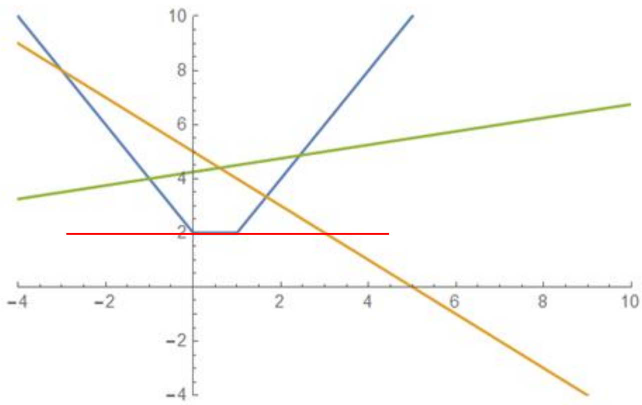
Функція $f(x)$, яка визначена на опуклій множині $X \subset E^n$ називається ***строго опуклою***, якщо нерівність (1) виконується як строга для всіх значень $x_1, x_2 \in X$ та усіх $\lambda x_1 + (1-\lambda) x_2$.

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.



строго опукла функція



опукла функція

Такі функції є опуклими:

- $f_1(x) = 3x + 4,$
- $f_2(x) = x^2 - 2x,$
- $f_3(x) = |x|,$
- $f_4(x) = -\sqrt{x},$
- $f_5(x) = 2(x_1^2 + x_2^2 - 2x_1x_2),$
- $f_6(x) = x_1^4 + 2x_2^2 + x_3^2 - 4x_1 - 4x_2x_3.$

Деякі властивості опуклих функцій:

1. Нехай $f_i(x)$ – опуклі функції, визначені на непорожній опуклій множині X. Тоді функція вигляду

$$f(x) = \sum_{i=1}^{l} \theta_i f_i(x),\ \theta_i \ge 0,\ i = \overline{1,l}$$

є опуклою. Отже, ***лінійна комбінація опуклих функцій з невід'ємними коефіцієнтами є функцією опуклою***.

Доведення. Множина X є опуклою. Візьмемо точки $x_1, x_2 \in X$. За означенням опуклої множини $x = \lambda x_1 + (1-\lambda) x_2 \in X$ для всіх λ, $0 \le \lambda \le 1$, також $f_i(x)$ – опуклі функції, тому за означенням опуклої функції

$$f_i\left(\lambda x_1 + (1-\lambda) x_2\right) \le \lambda f_i(x_1) + (1-\lambda) f_i(x_2).$$

Помножимо обидві частини нерівності на θ_i і підсумуємо за $i = \overline{1,l}$:

$$\underbrace{\sum_{i=1}^{l} \theta_i f_i\left(\lambda x_1 + (1-\lambda) x_2\right)}_{f(x)} \le \sum_{i=1}^{l} \theta_i \left[\lambda f_i(x_1) + (1-\lambda) f_i(x_2)\right] =$$

$$= \lambda \underbrace{\sum_{i=1}^{l} \theta_i f_i(x_1)}_{f(x_1)} + (1-\lambda) \underbrace{\sum_{i=1}^{l} \theta_i f_i(x_2)}_{f(x_2)}.$$

Отже,

$$f\left(\lambda x_1 + (1-\lambda) x_2\right) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

2. Нерівність Йенсена (узагальнення поняття опуклої функції).

Нехай функція $f(x)$ – опукла функції, яка визначена на опуклій множині X. Тоді для неї має місце нерівність:

$$f\left(\sum_{i=1}^{l} \lambda_i x_i\right) \le \sum_{i=1}^{l} \lambda_i f(x_i),\ \lambda_i \ge 0,\ i = \overline{1,l},\ \sum_{i=1}^{l} \lambda_i = 1,\ x_i \in X. \qquad (2)$$

Доведення. Проводиться методом математичної індукції по l.

Нехай $l = 2$.

$$f_i\left(\lambda_1 x_1 + \lambda_2 x_2\right) \le \lambda_1 f(x_1) + \lambda_2 f(x_2),\ \lambda_1 + \lambda_2 = 1,\ \lambda_1 \ge 0,\ \lambda_2 \ge 0.$$

Позначимо $\lambda_1 = \lambda$, тоді $\lambda_2 = 1 - \lambda$, маємо визначення опуклої функції.

Припустимо, що нерівність (2) має місце при $l = k$.

Покажемо, що (2) має місце при $l = k+1$.

$$f\left(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k + \lambda_{k+1} x_{k+1}\right) \le \lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_k f(x_k) + \lambda_{k+1} f(x_{k+1}),$$

$$\lambda_1 + \lambda_2 + \ldots + \lambda_k + \lambda_{k+1} = 1,\ \lambda_i \ge 0,\ i = \overline{1,k+1}.$$

Без обмеження загальності міркувань припустимо, що $0 \le \lambda_i < 1$, $i = \overline{1,k+1}$. Якщо $\lambda_{k+1} = 1$, то $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 0$. Тоді

$$\lambda_1 + \lambda_2 + \ldots + \lambda_k = 1 - \lambda_{k+1},\ \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{1 - \lambda_{k+1}} = 1 \text{ або } \frac{\sum_{i=1}^{k} \lambda_i}{1 - \lambda_{k+1}} = 1.$$

Можна записати

$$f\left(\lambda_1 x_1+\lambda_2 x_2+\ldots+\lambda_k x_k+\lambda_{k+1}x_{k+1}\right)=f\left(\left(1-\lambda_{k+1}\right)\frac{\sum_{i=1}^{k}\lambda_i x_i}{1-\lambda_{k+1}}+\lambda_{k+1}x_{k+1}\right).$$

Позначимо x_{k+1} за $\overline{x}_2$, $\overline{x}_2 \in X$, а за $\overline{x}_1=\dfrac{\sum_{i=1}^{k}\lambda_i x_i}{1-\lambda_{k+1}}$. Точка $\overline{x}_1$ – опукла комбінація точок множини X. Тоді маємо

$$f\left(\left(1-\lambda_{k+1}\right)\overline{x}_1+\lambda_{k+1}\overline{x}_2\right)\le \lambda_{k+1}f\left(\overline{x}_2\right)+\left(1-\lambda_{k+1}\right)f\left(\overline{x}_1\right)=$$

$$=\lambda_{k+1}f\left(x_{k+1}\right)+\left(1-\lambda_{k+1}\right)f\left(\frac{\sum_{i=1}^{k}\lambda_i x_i}{1-\lambda_{k+1\,1}}\right)\le$$

$$=\lambda_{k+1}f\left(x_{k+1}\right)+\sum_{i=1}^{k}\lambda_i f\left(x_i\right).$$

3. Теорема про зв'язок опуклості та неперервності.

Опукла функція $f\left(x\right)$ неперервна в кожній ***внутрішній*** точці своєї множини визначення.

Нагадаємо деякі означення.

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

$$f\left(x+\alpha g\right)=f\left(x\right)+\alpha\left(f'\left(x\right),g\right)+o\left(x,g,\alpha\right), \tag{3}$$

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

Умова (3) одночасно визначає ***градієнт*** $f'\left(x\right)$ функції $f\left(x\right)$, при цьому $f'\left(x\right)=\left(\dfrac{\partial f\left(x\right)}{\partial x_1},\ldots,\dfrac{\partial f\left(x\right)}{\partial x_n}\right)$, де $\dfrac{\partial f\left(x\right)}{\partial x_i}=\lim\limits_{\alpha\to 0}\dfrac{f\left(x+\alpha e^i\right)-f\left(x\right)}{\alpha}$ – частинна похідна функції $f\left(x\right)$ у точці x по аргументу x_i, $e^i=\left(0,\ldots,0,1,0,\ldots,0\right)$ є i-тим координатним ортом.

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

Нехай функція $f\left(x\right)$ 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*** $g(x) \in E^n$, $\|g\| = 1$ 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 $f'(x,g) = (f'(x), g)$.

4. Теорема про диференційовність опуклої функції за напрямом.

Опукла функція $f(x)$ в кожній внутрішній точці опуклої множини X має похідну за будь-яким напрямком $g(x) \in E^n$, $\|g\| = 1$.

Деякі екстремальні властивості функцій, які визначені на опуклих множинах

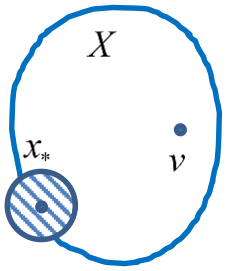
Теорема 1. Нехай $f(x)$ – опукла функція, яку визначено на опуклій множині $X \subset E^n$. Тоді

1) будь-яка точка її локального мінімуму є точкою глобального мінімуму,
2) множина точок мінімуму $X_* = \{ x \in X : f(x) = f(x_*) \}$ є опуклою, якщо вона не є порожньою.
3) Строго опукла функція має не більш ніж одну точку мінімуму.

Доведення.

1). Нехай x_* – точка локального мінімуму функції $f(x)$ на опуклій множині X. За означенням точки локального мінімуму існує такий окіл точки x_*

$O(x_*, \varepsilon) = \{ x \in E^n : \|x - x_*\| \le \varepsilon \}$, що $f(x_*) \le f(x)$ для $\forall x \in X \cap O(x_*, \varepsilon)$.



Візьмемо довільну точку $x \in X$. Так як X – опукла множина, то $x = x_* + \lambda(v - x_*)$, $0 < \lambda \le 1$.

При достатньо малих λ точка $x \in X \cap O(x_*, \varepsilon)$.

Доведення проведемо методом від супротивного.

Будем вважати, що існує точка $v \ne x_*$ така, що $f(v) < f(x_*)$.

Тобто, v – точка глобального мінімуму , яка не співпадає з точкою локального мінімуму x_*. Розглянемо

$$f(x_*) \le f(x) = f(x_* + \lambda(v - x_*)) = f(\lambda v + (1-\lambda)x_*) \overset{\{\text{за означенням опуклої функції}\}}{\le}$$
$$\le \lambda f(v) + (1-\lambda) f(x_*) < \lambda f(x_*) + (1-\lambda) f(x_*) = f(x_*).$$

Отримали протиріччя: $f(x_*) < f(x_*)$.

2). Нехай існує дві точки мінімуму x_* та y_* функції $f(x)$ на опуклій множині X . Розглянемо їх лінійну комбінацію

$$x = \lambda x_* + (1-\lambda) y_*,\ 0 \le \lambda \le 1,$$

і покажемо, що $x \in X_*$.

$$f(x_*) = f(y_*) \le f(x),\ \forall x \in X .$$

Маємо

$$f(x_*) = f(y_*) \le f(x) = f(\lambda x_* + (1-\lambda) y_*) \le \lambda f(x_*) + (1-\lambda) f(y_*) = f(x_*).$$

Отримали $f(x) \le f(x_*)$. Отже, $f(x) = f(x_*)$ і, тому, $x \in X_*$. За визначенням опуклої множини маємо, що множина X_* – опукла.

3). Доведення проведемо методом від супротивного.

Нехай функція $f(x)$ є строго опуклою та має дві точки мінімуму x_* та y_*.

Маємо

$$f(x_*) = f(y_*) \le f(x) = f(\lambda x_* + (1-\lambda) y_*) \overset{\{\text{за означенням строго опуклої функції}\}}{<}$$

$$< \lambda f(x_*) + (1-\lambda) f(y_*) = f(x_*).$$

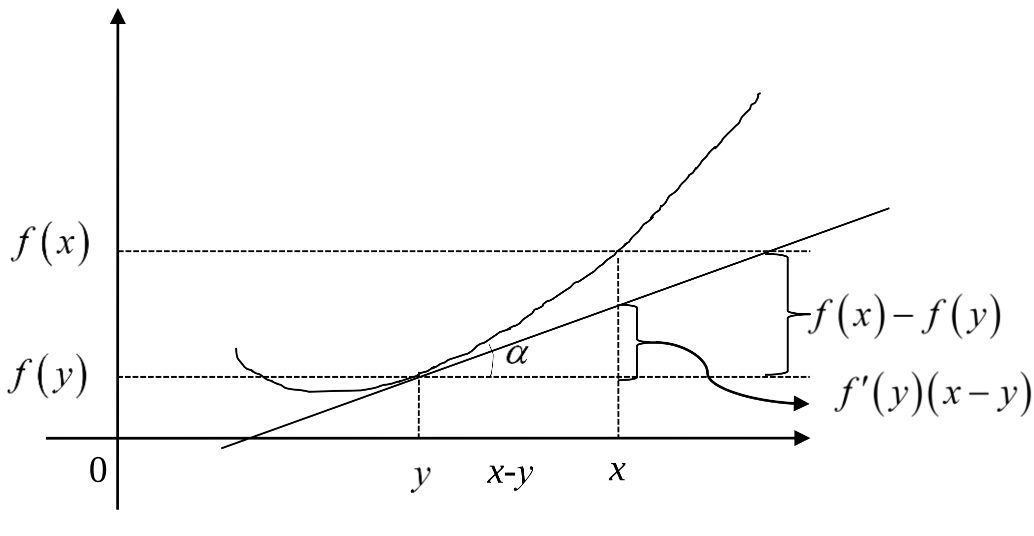
Отримали $f(x_*) < f(x_*)$. Отже, маємо протиріччя.

Теорема 2. Нехай $f(x)$ – опукла функція, яку визначено на опуклій множині $X \subset E^n$ та $f(x) \in C^1(X)$. Тоді має місце нерівність

$$(f'(y), x-y) \le f(x) - f(y) \le (f'(x), x-y) \text{ для всіх } x, y \in X . \qquad (4)$$

Без доведення.

Наведемо геометричний сенс нерівності (4), а саме лівої частини цієї нерівності.



$$f'(y) = tg\alpha,\ f'(y)(x-y) \le f(x) - f(y).$$

Якщо функція є неперервно диференційовною, то графік функція лежить ***не нижче*** за дотичну, яку проведено до кривої $f(x)$ в точці y.

Теорема 3. Нехай функція $f(x)$ визначена на опуклій множині $X \subset E^n$ та $f(x) \in C^1(X)$. Тоді

1. Для того, щоб точка x_* була точкою мінімуму функції $f(x)$ на множині X, ***необхідно*** виконання нерівності
$$\left(f'(x_*), x - x_* \right) \geq 0 \text{ для } \forall x \in X. \tag{5}$$
2. Якщо точка x_* є внутрішньою точкою множини X, тобто $x_* \in int\, X$, то нерівність (5) еквівалентна умові $f'(x_*) = 0$.
3. ***Критерій оптимальності опуклої функції першого порядку***.

Якщо функція $f(x)$ – опукла на множині X, то умова (5) є також ***достатньою*** умовою того, що x_* – точка мінімуму.

Без доведення.

Функція називається ***двічі диференційовною в точці*** x, якщо разом з градієнтом існує симетрична матриця $f''(x)$ порядку $n \times n$ така що
$$f(x+h) = f(x) + \left(f'(x), h \right) + \frac{1}{2}\left(f''(x)h, h \right) + \beta(h,x), \quad \text{де} \quad \frac{\beta(h,x)}{\| h \|^2} \to 0 \quad \text{при}$$
$\| h \| \to 0$.

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
$$f''(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j=1,n}$$
– ***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.

Критерій опуклості в термінах других похідних.

Двічі неперервно-диференційовна на опуклій, непорожній множині X функція $f(x)$ опукла тоді і тільки тоді, коли
$$\left(f''(x)h, h \right) \geq 0, \ \forall x \in X, \ \forall h \in E^n. \tag{6}$$

Умова (6) означає, що матриця Гессе (гессіан) $f''(x)$ невід’ємно визначена для $\forall x \in X$.

Критерій оптимальності другого порядку.

Необхідна умова. Нехай $f(x) \in C^2(X)$ і x_* є точка мінімуму $f(x)$ на множині X. Тоді

$$\left(f''(x_*)h, h\right) \geq 0, \ \forall h \in E^n.$$

Достатня умова. Нехай $f(x) \in C^2(X)$ і квадратична форма є додатно визначеною

$$\left(f''(x_*)h, h\right) > 0, \ \forall h \in E^n, \ h \neq 0.$$

Тоді x_* є точкою мінімуму $f(x)$ на множині X.