Методи оптимізації. Лекція 22.04.2022

Теореми про збіжність градієнтних методів

Теореми про збіжність сформулюємо для методу найшвидшого спуску. Вони будуть справедливі і для інших градієнтних методів.

Теорема 1 (умови 1-3). Нехай виконані умови:

1) $f(x) \in C^1(E^n)$;

2) $f(x)$ обмежена знизу: $inf\ f(x) = f_* > -\infty$, $x \in E^n$;

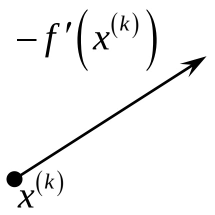
3) градієнт функції $f(x)$ задовольняє умову Ліпшиця:

$$\left\| f'(x') - f'(x'') \right\| \le L \left\| x' - x'' \right\|, \quad \forall x', x'' \in E^n,$$

$L > 0$ – константа Ліпшиця.

Тоді послідовність $\left\{ x^{(k)} \right\}_0^\infty$, яку побудовано за методом (2),(3) є такою, що $\left\| f'\left(x^{(k)}\right) \right\| \to 0$ при $k \to \infty$ для будь-якого початкового наближення $x^{(0)}$.

<u>Доведення:</u> На промені $\left\{ x : x = x^{(k)} - \alpha\, f'\left(x^{(k)}\right), \alpha > 0 \right\}$ візьмемо довільну точку x.



За теоремою про середнє для будь-якої початкової точки $x^{(0)}$:

$$f(x) - f\left(x^{(k)}\right) = \left(f'\left(\overline{x}^{(k)}\right), x - x^{(k)} \right) =$$

{точка $\overline{x}^{(k)} \in \left[x^{(k)}, x \right]$ і може бути представлена як

$$\overline{x}^{(k)} = x^{(k)} + \theta\left(x - x^{(k)}\right), \quad 0 \le \theta \le 1\}$$

$$= \left(f'\left(\overline{x}^{(k)}\right) + f'\left(x^{(k)}\right) - f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) =$$

$$= \left(f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) + \left(f'\left(\overline{x}^{(k)}\right) - f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) \le$$

{застосуємо до другого доданку нерівність Коші-Буняковського: $\left|(a,b)\right| \le \|a\| \cdot \|b\|$}

$$\le \left(f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) + \left\| f'\left(\overline{x}^{(k)}\right) - f'\left(x^{(k)}\right) \right\| \cdot \left\| x - x^{(k)} \right\| \le$$

{за умовою Ліпшиця}

$$\le \left(f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) + L \left\| \overline{x}^{(k)} - x^{(k)} \right\| \cdot \left\| x - x^{(k)} \right\| \le$$

{виберемо таку точку x, щоб $\theta = 1$}

$$\le \left(f'\left(x^{(k)}\right), \quad x - x^{(k)} \right) + L \left\| x - x^{(k)} \right\|^2 =$$

{за ітераційною формулою $x - x^{(k)} = -\alpha\ f'\left(x^{(k)}\right)$}

$$= -\alpha \left\| f'\left(x^{(k)}\right) \right\|^2 + L\alpha^2 \left\| f'\left(x^{(k)}\right) \right\|^2 .$$

Позначимо функцію відносно α як

$$\phi(\alpha) = -\alpha \left\| f'\left(x^{(k)}\right) \right\|^2 + L\alpha^2 \left\| f'\left(x^{(k)}\right) \right\|^2 .$$

Функція $\phi(\alpha)$ є квадратичною і, отже, опуклою. Знайдемо мінімум функції $\phi(\alpha)$ класичним методом: $\phi'(\alpha) = 0$.

$$2L\alpha \left\| f'\left(x^{(k)}\right) \right\|^2 - \left\| f'\left(x^{(k)}\right) \right\|^2 = 0,$$

$$\alpha_{min} = \frac{1}{2L}, \quad \phi(\alpha_{min}) = \frac{\left(\left\| f'\left(x^{(k)}\right) \right\|^2 - 2\left\| f'\left(x^{(k)}\right) \right\| \right)}{4L} = \frac{-\left\| f'\left(x^{(k)}\right) \right\|^2}{4L} .$$

Отже,

$$x^{(k+1)} = x^{(k)} - \alpha_{min} f'\left(x^{(k)}\right),$$

$$f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \le f(x) - f\left(x^{(k)}\right) \le -\phi(\alpha_{min}) = -\frac{\left\| f'\left(x^{(k)}\right) \right\|^2}{4L} .$$

За самою побудовою методу, виконується умова монотонності:

$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right), \ \lim_{k\to\infty}\left[f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \right] = 0,$$

$$\left\| f'\left(x^{(k)}\right) \right\|^2 \le -4L\left[f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \right], \ \left\| f'\left(x^{(k)}\right) \right\|^2 \to 0 \text{ при } k \to \infty . \quad \square$$

Теорема 1 (з умовами 1-3) встановлює збіжність за функцією до нижньої межі.

Теорема 1 (умови 1-4). Нехай виконані умови 1)-3) та

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

тоді

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

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 (з умовами 1-4) 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. Якщо виконуються умови (1)-(4) теореми 1 і, крім того, функція $f(x)$ опукла на E^n, тоді:

1) послідовність $\left\{x^{(k)}\right\}_0^\infty$ є мінімізуючою,

2) послідовність $\left\{x^{(k)}\right\}_0^\infty$ збігається до множини точок мінімуму функції $f\left(x\right)$,

3) має місце оцінка за функцію

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

Тут D – діаметр множини $M\left(x^{(0)}\right)$:

$$D = max\left\|x' - x''\right\|, \quad x', x'' \in M\left(x^{(0)}\right).$$

Теорема 2 стверджує про збіжність за змінною і дає оцінку швидкості збіжності за функцією. Оцінка швидкості збіжності за змінною можлива лише при більш жорстких припущеннях щодо функції f.

Теорема 3. Нехай функція $f\left(x\right) \in C^2\left(E^n\right)$ і її матриця Гессе задовольняє умову:

$$m\left\|z\right\|^2 \le \left(f''\left(x\right)z, z\right) \le M\left\|z\right\|^2,\ 0 < m \le M,\ \forall x, z \in E^n, \tag{8}$$

тоді послідовність $\left\{x^{(k)}\right\}_0^\infty$, побудована за формулами (2), (3) збігається до точки x_* зі швидкістю геометричної прогресії зі знаменником $q = \dfrac{M - m}{M + m}$ (лінійно) незалежно від початкового наближення та справедлива така оцінка

$$\left\|x^{(k+1)} - x_*\right\| \le \frac{M - m}{M + m}\left\|x^{(k)} - x_*\right\|. \tag{9}$$

Функція $f\left(x\right)$, яка задовольняє умову (8), називається ***сильно опуклою функцією***. Числа m та M – це найменше та найбільше власні числа матриці других похідних $f''\left(x\right)$.

Геометричний сенс градієнтних методів

1. Градієнт $f'\left(x^{(k)}\right)$ – **перпендикулярний в точці** $x^{(k)}$ **лінії рівня** $\Gamma_k = \left\{\ x : f(x) = f\left(x^{(k)}\right)\right\}$ (рис. 1). Це властивість є справедлива для всіх градієнтних методів.

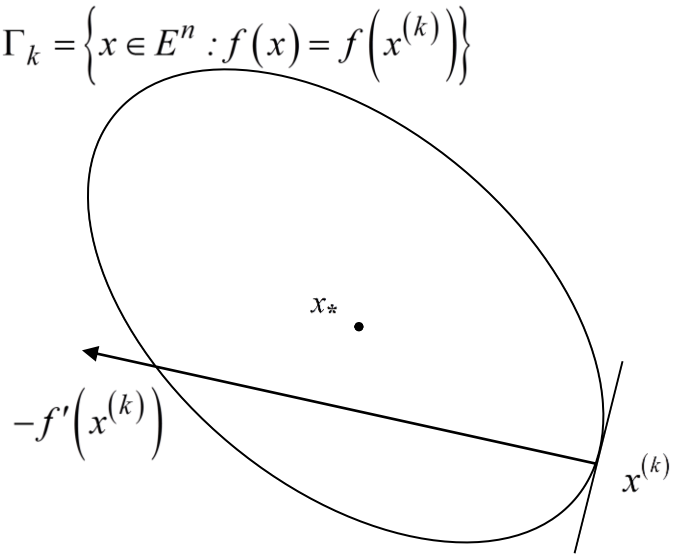


Рис. 1

2. Для методу **найшвидшого спуску** два послідовно знайдених напрямку спуску $\left(-f'\left(x^{(k)}\right)\right)$ та $\left(-f'\left(x^{(k+1)}\right)\right)$ є взаємно ортогональними (рис. 2).

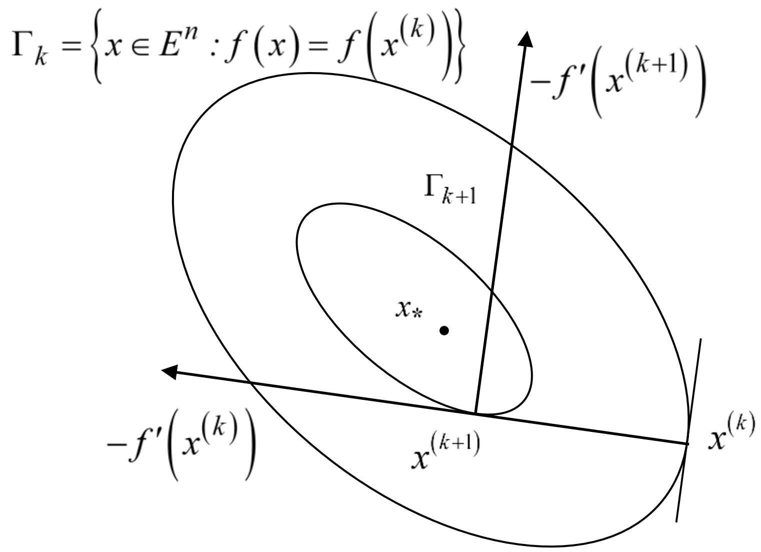


Рис. 2

3. Градієнтний метод (відмінний від методу найшвидшого спуску) генерує послідовність точок $\left\{ x^{(k)} \right\}_0^{\infty}$, які утворюють зигзагоподібну траєкторію $x^{(0)},\ x^{(1)},\ x^{(2)},\ \ldots,\ x^{(k)}$ (рис. 3).

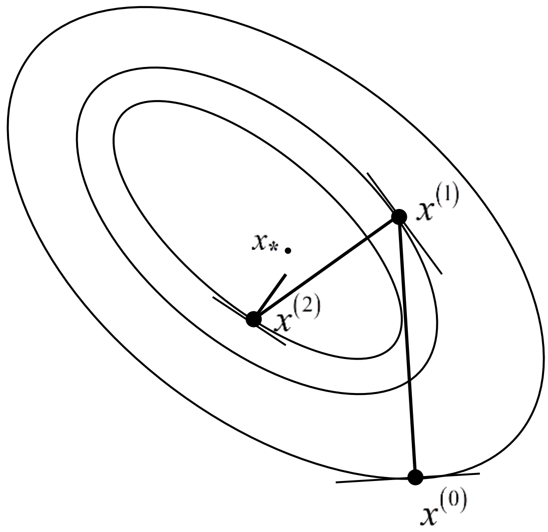


Рис. 3

Переваги градієнтних методів такі:

1) досить 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'язання задачі. Їх застосування доцільне у випадку, коли початкове наближення знаходиться далеко від точки мінімуму цільової функції, а кроки уздовж антиградієнту 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 і сильна опуклість (теорема 3);

2) повільна збіжність для функцій, в яких поверхні (лінії рівня) сильно витягнуті та функція має так званий «яружний» характер (рис. 4).

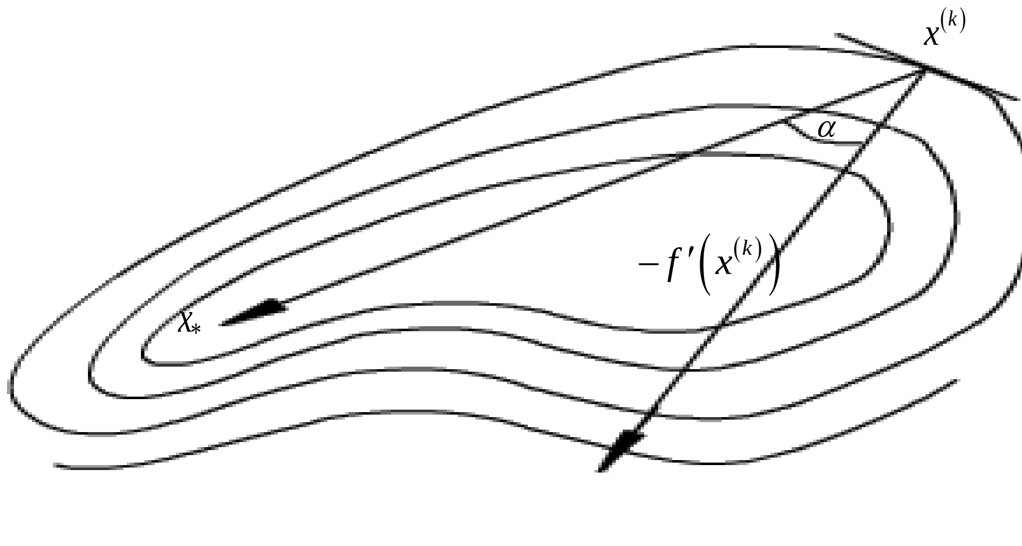


Рис. 4

Вплив кривизни ліній рівня на швидкість збіжності

Нехай

$$f(x) = \left(A(x - x_*),\, x - x_*\right),$$

де A – симетрична додатно визначена матриця.

Функція $f(x) \in C^2(E^n)$ та є сильно опукла, тобто

$$m\|z\|^2 \le (Az, z) \le M\|z\|^2, \ \forall z \in E^n,$$

m та M є найменше і найбільше власні числа матриці A.

Припустимо, що $m = M$, тоді $f(x) = M\|x - x_*\|^2$. Лінії рівня такої цільової функції є концентричними колами.

Чим більше будуть відрізнятися між собою m та M, тим більше будуть витягнутими будуть лінії рівня.

Чим більше витягнуті лінії рівня, тим гірше буде збігатися градієнтний метод. Це також видно з оцінки швидкості збіжності:

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

Матриця $f''(x)$ других похідних буде добре обумовленою, якщо m та M мало відрізняються один від одного і погано обумовленою, якщо $m \ll M$.

Якщо числа m та M мало відрізнятися між собою, тобто матриця $f''(x)$ добре обумовлена, то число q близько до нуля і, отже, збіжність краще. Якщо $m / M \ll 1$, то q близько до 1 і метод збігається повільніше.

Геометричне тлумачення цього факту: зі зменшенням відношення m / M лінії рівня стають більш витягнутими та напрямок спуску (антиградієнт) у більшості точок істотно відхиляється від напрямку на точку мінімуму.

Шляхи прискорення збіжності для функцій з витягнутими лініями рівня:

1) масштабування, тобто перехід до іншої функції заміною змінних.

Наприклад: $f(x) = x_1^2 + 25x_2^2$, зробимо заміну змінних $\xi_1 = x_1$, $\xi_2 = 5x_2$. $g(\xi) = \xi_1^2 + \xi_2^2$ (рис. 5).

x_2 ξ_2

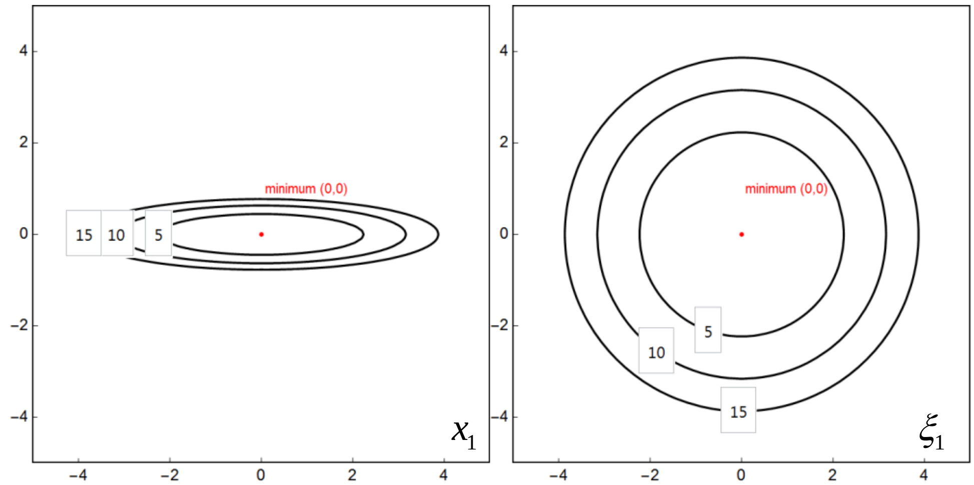


Рис. 5.

2) створення методів, які з геометричних міркувань, враховують кривизну ліній рівня. В основному це евристичні методи (не доведена збіжність в загальному випадку). Наприклад, r-алгоритми Шора.

3) створення методів, які враховує інформацію про кривизну ліній рівня з використанням матриці других похідних (метод Ньютона).