Методи оптимізації. Лекція 15.04.2022

Чисельні методи безумовної оптимізації

Постановка задачі безумовної оптимізації:

$$f(x) \to min,\ x \in E^n. \tag{1}$$

Загальна ітераційна формула методів розв’язання задачі безумовної оптимізації має вигляд

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

або у покоординатній формі

$$x_1^{(k+1)} = x_1^{(k)} + \alpha_k p_1^{(k)},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_n^{(k+1)} = x_n^{(k)} + \alpha_k p_n^{(k)},$$
$$\alpha_k \in E^1,\ k = 0,1,2,\ldots,\ x = (x_1,\ldots,x_n).$$

Методи монотонного спуску

$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right) \tag{3}$$

називаються ***релаксаційними*** методами. Відповідна послідовність $\left\{x^{(k)}\right\}_0^\infty$ називається також релаксаційною.

Конкретний алгоритм визначається: вибором ***початкової точки*** $x^{(0)}$, правилами вибору ***напрямків спуску*** $p^{(k)}$ та ***крокових множників*** α_k на основі отриманої в результаті обчислень інформації, а також ***умовою зупинки*** ітераційного процесу.

Вектор $p^{(k)} = \left(p_1^{(k)},\ldots,p_n^{(k)}\right)$ 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 $p^{(k)}$, 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.

За характером використання інформації про функцію, що мінімізується, методи поділяються на:

- методи нульового порядку,
- методи першого порядку,
- методи другого порядку.

Методи, які використовують лише інформацію про значення цільової функції, називаються **методами нульового порядку**. Прикладами таких методів є методи покоординатного спуску, пошуку по деформованому багатограннику, конфігурацій, метод Розенброка.

Методи, які використовують інформацію про значення цільової функції та її першої похідної – **методами першого порядку**. Це градієнтні методи, методи спряжених напрямків, методи змінних напрямків, методи змінної метрики.

Методи, що використовують, крім того, інформацію про другі похідні, – **методами другого порядку**. Метод Ньютона.

Збіжність методів оптимізації

Важливою характеристикою нескінченнокрокових методів є збіжність.

Метод (2) **збігається,** якщо

$$x^{(k)} \to x_* \text{ при } k \to \infty,$$

де x_* – розв'язок задачі (1).

У випадку, коли точка мінімуму x_* не єдина, під збіжністю методу розуміється збіжність послідовності $\left\{x^{(k)}\right\}_0^\infty$ до множини X_* точок мінімуму функції $f\left(x\right)$.

Коли

$$f\left(x^{(k)}\right) \to f\left(x_*\right),$$

то метод (2) **збігається за функцією**, при цьому послідовність $\left\{x^{(k)}\right\}_0^\infty$ називається **мінімізуючою**.

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 $q \in \left(0,1\right)$ та k_0, що

$$\left\| x^{(k+1)} - x_* \right\| \le q \left\| x^{(k)} - x_* \right\| \text{ при } k \ge k_0. \quad (4)$$

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

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

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

$$\left\| x^{(k+1)} - x_* \right\| \le C \left\| x^{(k)} - x_* \right\|^2 \text{ при } k \ge k_0. \quad (5)$$

Для характеристики збіжності послідовності наближених значень $f\left(x^{(k)}\right)$ до $f\left(x_*\right)$ 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 сформульовані вимоги до початкового наближення $x^{(0)}$. Нарешті, аналіз швидкості збіжності дає корисну кількісну та якісну характеристику методу оптимізації, який вивчається.

Умови закінчення ітераційного процесу
(Критерії закінчення розрахунків)

На практиці часто використовуються такі умови закінчення розрахунків:

$$\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon_1, \tag{6}$$

$$\left| f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \right| \le \varepsilon_2, \tag{7}$$

$$\left\| f'\left(x^{(k+1)}\right) \right\| \le \varepsilon_3. \tag{8}$$

До початку обчислень вибирається одна з умов (6)–(8) 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 (6)–(8) або всіх трьох.

Градієнтні методи

Постановка задачі.

$$f(x) \to min,\ x \in E^n. \tag{1}$$

Умовою застосування градієнтних методів є ***неперервна диференційовність цільової функції*** $f(x)$ в E^n. Будемо записувати цю умову у вигляді

$$f(x) \in C^1\left(E^n\right). \tag{2}$$

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

$$x^{(k+1)} = x^{(k)} - \alpha_k f'\left(x^{(k)}\right),\ k = 0, 1, 2, \ldots,\ \alpha_k > 0 \tag{3}$$

або в координатній формі

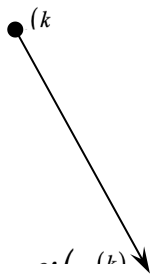
$$x_i^{(k+1)} = x_i^{(k)} - \alpha_k \frac{\partial f\left(x^{(k)}\right)}{\partial x_i},\ i = 1, 2, \ldots, n,\ k = 0, 1, 2, \ldots,\ \alpha_k > 0.$$

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

Різні способи вибору величини α_k у методі (3) визначають різні варіанти градієнтних методів. Наведемо 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

$$g_k(\alpha) = f\left(x^{(k)} - \alpha f'\left(x^{(k)}\right)\right), \quad \alpha \geq 0,$$

і визначимо α_k з умови

$$g_k(\alpha_k) = \min_{\alpha \geq 0} g_k(\alpha), \quad \alpha_k > 0. \tag{4}$$

Метод (3), (4), 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 (4) є важким або неможливим (наприклад, мінімум в (4) при деяких k не досягається), удаються до інших способів вибору α_k.

2. *Градієнтний метод з дробленням кроку*

Процес (3) з дробленням кроку проходить так. Вибираються певні константи $\beta > 0,\ 0 < \lambda < 1$ (часто $\lambda = \frac{1}{2}$). Для коефіцієнта $\alpha = \beta$ перевіряється умова монотонності

$$f\left(x^{(k)} - \alpha f'\left(x^{(k)}\right)\right) < f\left(x^{(k)}\right). \tag{5}$$

Якщо її виконано, то покладають $\alpha_k = \alpha$. Якщо ні, то проводиться дроблення кроку, тобто приймається $\alpha = \lambda \beta$, і знову перевіряється виконання умови (5). Процес дроблення продовжується доти, поки умова (5) не виявиться виконаною. Цей процес не може бути нескінченим, так як $\left(-f'\left(x^{(k)}\right)\right)$ – напрямок спадання. Перше значення α, для якого умова (5) виявиться виконаною, приймається за кроковий множник α_k.

3. *Градієнтний метод з адаптивним вибором кроку* α_k

Цей метод також називають градієнтний метод з автоматичним вибором кроку.

Виберемо α_k з умови:

$$f\left(x^{(k)}-\alpha_k f'\left(x^{(k)}\right)\right)-f\left(x^{(k)}\right)\le -\varepsilon\,\alpha_k \left\| f'\left(x^{(k)}\right)\right\|^2, \qquad (6)$$

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 ускладняється вибір методу α_k з умови (6).

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 (6) при $\alpha_k=\alpha$,
4) якщо нерівність (6) виконано, то значення α вибираємо за α_k і переходимо до наступної ітерації. У протилежному випадку проводимо дроблення числа α (помножуючи його на довільне число $0<\lambda<1$) поки нерівність (6) не буде виконаною. Тоді переходимо до наступної ітерації.

Відзначимо, що з виконання нерівності (6) випливає виконання нерівності (5). Вимога (6) більш сувора, ніж (5), але має той же сенс. Цей градієнтний метод найбільш часто використовується на практиці.

4. *Апріорне завдання величини* α_k

Виберемо α_k з умови:

$$\alpha_k>0,\ k=0,1,...,\ \sum_{k=0}^{\infty}\alpha_k=\infty,\ \sum_{k=0}^{\infty}\alpha_k^2<\infty. \qquad (7)$$

Наприклад, за α_k можна узяти $\alpha_k=\dfrac{1}{k+1}$.

Вибір α_k за формулою (7) 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 (7) застосовується також у субградієнтному методі (методі узагальненого градієнтного спуску), який є аналогом градієнтного методу для мінімізації негладких функцій.