Методи оптимізації. Лекція 06.05.2022

Метод спряжених градієнтів

Постановка задачі.

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

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

Ітераційна формула методу спряжених градієнтів (nonlinear conjugate gradient method):

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)},\ k = 0,1,\ldots \tag{3}$$

де

$$p^{(k)} = \begin{cases} f'\left(x^{(k)}\right), k = 0, \\ f'\left(x^{(k)}\right) - \beta_k p^{(k-1)},\ k = 1, 2, .. \end{cases}$$

$$\beta_k = \begin{cases} \dfrac{\left(f'\left(x^{(k)}\right), f'\left(x^{(k-1)}\right) - f'\left(x^{(k)}\right)\right)}{\left\| f'\left(x^{(k-1)}\right)\right\|^2}, k \in I_1 \\ 0, k \in I_2 \end{cases}$$

Кроковий множник α_k – може вибиратися як в градієнтних методах. Наприклад: $\alpha_k : f\left(x^{(k)} - \alpha_k p^{(k)}\right) = \min\limits_{\alpha \geq 0} f\left(x^{(k)} - \alpha p^{(k)}\right)$.

Нехай I_1, I_2 – множини індексів, $I_1 \cup I_2 = \left\{0,1,2,\ldots\right\}$, $I_1 \cap I_2 = \varnothing$.

Якщо множина задається як $I_1 = \left\{1,2,\ldots\right\}$, а множина $I_2 = \left\{0\right\}$, то в цьому випадку ми отримаємо класичний метод спряжених градієнтів.

Якщо $I_1 = \varnothing$, а множина $I_2 = \left\{0,1,2,\ldots\right\}$ – це метод найшвидшого спуску.

Якщо $I_2 = \left\{0, k, 2k, 3k, \ldots\right\}$ – метод спряжених градієнтів з моментами оновлення методу.

Геометричний сенс методу спряжених градієнтів. У методі спряжених градієнтів за напрямок спуску вибирається вектор, який є лінійною комбінацією антиградієнту і попереднього напрямку спуску. Це дозволяє зменшити кут між напрямком спуску і напрямком на точку мінімуму.

Чутливість методу спряжених градієнтів до кривизни ліній рівня визначається наявністю параметра β_k, який містить інформацію про другу похідну функції через різницю перших похідних : $f'\left(x^{(k-1)}\right) - f'\left(x^{(k)}\right)$.

Теорема про збіжність. Для двічі неперервно-диференційовної, сильно опуклої функції метод спряжених градієнтів збігається із надлінійною швидкістю. Для тричі неперервно-диференційовної функції – з квадратичною швидкістю.

Метод спряжених градієнтів для квадратичної функції

Розглянемо квадратичну функцію

$$f(x) = \frac{1}{2}(Ax, x) - (b, x).$$

Для цієї функції ітераційна формула методу спряжених градієнтів матиме вигляд:

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)},\ k = 0,1,\ldots \tag{4}$$

де

$$p^{(k)} = \begin{cases} f'\left(x^{(k)}\right), k = 0, \\ f'\left(x^{(k)}\right) - \beta_k p^{(k-1)},\ k = 1,2,.. \end{cases}$$

$$\beta_k = \frac{\left(f'\left(x^{(k)}\right), A\, p^{(k-1)}\right)}{\left(A\, p^{(k-1)},\, p^{(k-1)}\right)} = \frac{\left\| f'\left(x^{(k)}\right)\right\|^2}{\left\| f'\left(x^{(k-1)}\right)\right\|^2},$$

$$\alpha_k = -\frac{\left(f'\left(x^{(k)}\right), p^{(k)}\right)}{\left(Ap^{(k)}, p^{(k)}\right)}.$$

Два вектори x і y простору E^n називається ***спряженими*** щодо матриці A або ***A-ортогональними***, якщо

$$(x,\ Ay) = 0.$$

Метод спряжених градієнтів для квадратичної функції генерує послідовність спряжених щодо матриці A напрямків $p^{(0)}, p^{(1)}, \ldots, p^{(n-1)}$, тобто для методу спряжених градієнтів виконується

$$\left(p^{(k)}, Ap^{(j)}\right) = 0, k \neq j\ .$$

Лема про властивості послідовностей $\left\{x^{(k)}\right\}$ і $\left\{p^{(k)}\right\}$.

Нехай $I_1 = \{1,2,\ldots\}$, $I_2 = \{0\}$. Послідовності точок $\left\{x^{(k)}\right\}$ та напрямків $\left\{p^{(k)}\right\}$, які генеруються методом (4) мають такі властивості:

1. $\left(f'\left(x^{(i)}\right), p^{(j)}\right) = 0, \quad i > j$. (5)
2. $\left(f'(x^{(i)}), f'(x^{(j)})\right) = 0, \quad i \neq j$. (6)
3. $\left(p^{(i)}, Ap^{(j)}\right) = 0, \quad i \neq j$. (7)

Властивість (5) означає, що градієнт в точці $x^{(k)}$ є ортогональний всім попереднім напрямкам спуску $p^{(k-1)}, \ldots, p^{(0)}$.

Властивість (6) означає, що напрямки градієнтів в довільних точках послідовності $\left\{x^{(k)}\right\}$ є взаємно ортогональними.

Властивість (7) означає, що напрямки спуску A-ортогональні.

Теорема. Метод спряжених градієнтів для квадратичної функції збігається за ***скінчене число*** кроків, що не перевищує розмірності простору.

Приклад. Знайдемо A–спряжені напрямки. Розглянемо таку задачу:

$$f(x) = 4x_1^2 + 4x_2^2 - 4x_1x_2 - 12x_2 \to min,\ x \in E^2.$$

Матриця Гессе цієї функції має вигляд:

$$A = f''(x) = \begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix}, \quad x = (x_1, x_2).$$

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

$$(1,0)\begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8 \\ -4 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 8a - 4b = 0.$$

Наприклад, можна вибрати $a = 1$, $b = 2$ так що $p_2^T = (1,2)$. На рис. показані її лінії рівня.

Відзначимо, що спряжені напрямки визначаються неоднозначно.

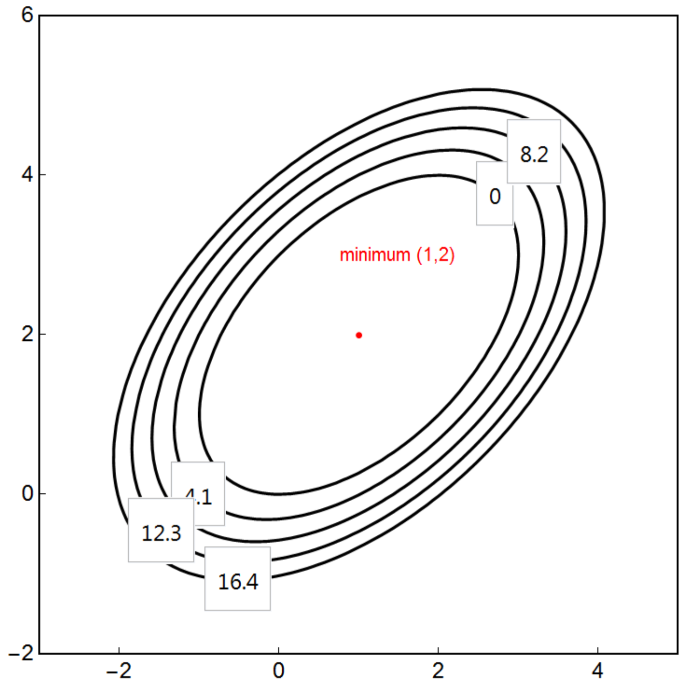


Рис.

Приклад. Розглянемо ту ж задачу:

$$f\left(x_1,x_2\right)=4x_1^2+4x_2^2-4x_1x_2-12x_2\to min,\ x\in E^2.$$

Цільова функція є квадратичною.

Для неї ітераційна формула методу спряжених градієнтів має вигляд (4):

$$x^{(k+1)}=x^{(k)}-\alpha_k p^{(k)},\ k=0,1,\dots$$

де

$$p^{(k)}=\begin{cases} f'\left(x^{(k)}\right), k=0,\\ f'\left(x^{(k)}\right)-\beta_k p^{(k-1)},\ k=1,2,.. \end{cases}$$

$$\beta_k=\frac{\left\|f'\left(x^{(k)}\right)\right\|^2}{\left\|f'\left(x^{(k-1)}\right)\right\|^2},\ \alpha_k=-\frac{\left(f'\left(x^{(k)}\right),p^{(k)}\right)}{\left(Ap^{(k)},p^{(k)}\right)}.$$

Градієнт і матриця других похідних мають вигляд:

$$f'(x)=\begin{pmatrix}8x_1-4x_2\\8x_2-4x_1-12\end{pmatrix},\ A=f''(x)=\begin{bmatrix}8&-4\\-4&8\end{bmatrix}.$$

Виберемо початкове наближення $x^{(0)}=(0,0)^T$. Маємо

$$f\left(x^{(0)}\right)=0,\ f'\left(x^{(0)}\right)=\begin{pmatrix}0\\-12\end{pmatrix},\ \left\|f'\left(x^{(0)}\right)\right\|^2=144.$$

Знайдемо напрямок спуску і кроковий множник:

$$p^{(0)}=f'\left(x^{(0)}\right)=\begin{pmatrix}0\\-12\end{pmatrix},$$

$$\left(Ap^{(0)},p^{(0)}\right)=\left(\begin{bmatrix}8&-4\\-4&8\end{bmatrix}\begin{pmatrix}0\\-12\end{pmatrix},\begin{pmatrix}0\\-12\end{pmatrix}\right)=\left(\begin{pmatrix}48\\-96\end{pmatrix},\begin{pmatrix}0\\-12\end{pmatrix}\right)=$$

$$=48\cdot 0+96\cdot 12=1152,$$

$$\alpha_0=-\frac{\left(f'\left(x^{(0)}\right),p^{(0)}\right)}{\left(Ap^{(0)},p^{(0)}\right)}=\frac{144}{1152}=\frac{1}{8}.$$

Перше наближення

$$x^{(1)}=x^{(0)}-\alpha_0 p^{(0)}=\begin{pmatrix}0\\0\end{pmatrix}-\frac{1}{8}\begin{pmatrix}0\\-12\end{pmatrix}=\begin{pmatrix}0\\3/2\end{pmatrix}.$$

Маємо

$$f\left(x^{(1)}\right)=-9,\ f'\left(x^{(1)}\right)=\begin{pmatrix}-6\\0\end{pmatrix},\ \left\|f'\left(x^{(1)}\right)\right\|^2=36.$$

Знайдемо напрямок спуску:

$$\beta_1 = \frac{\left\| f'\left(x^{(1)}\right)\right\|^2}{\left\| f'\left(x^{(0)}\right)\right\|^2} = \frac{36}{144} = \frac{1}{4},\ p^{(1)} = f'\left(x^{(1)}\right) - \beta_1 p^{(0)} = \begin{pmatrix} -6 \\ 0 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} 0 \\ 12 \end{pmatrix} = \begin{pmatrix} -6 \\ -3 \end{pmatrix}.$$

і кроковий множник

$$\left(Ap^{(1)}, p^{(1)}\right) = \left(\begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix}\begin{pmatrix} -6 \\ -3 \end{pmatrix}, \begin{pmatrix} -6 \\ -3 \end{pmatrix}\right) = \left(\begin{pmatrix} -36 \\ 0 \end{pmatrix}, \begin{pmatrix} -6 \\ -3 \end{pmatrix}\right) = 216,$$

$$\alpha_1 = -\frac{\left(f'\left(x^{(1)}\right), p^{(1)}\right)}{\left(Ap^{(1)}, p^{(1)}\right)} = \frac{36}{216} = \frac{1}{6}.$$

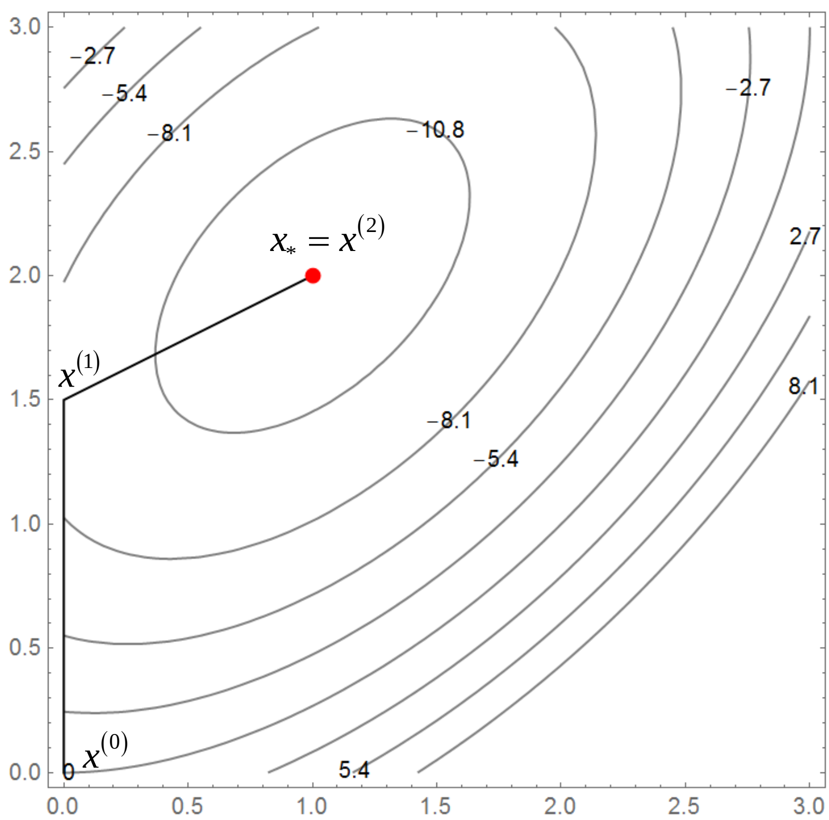
Друге наближення

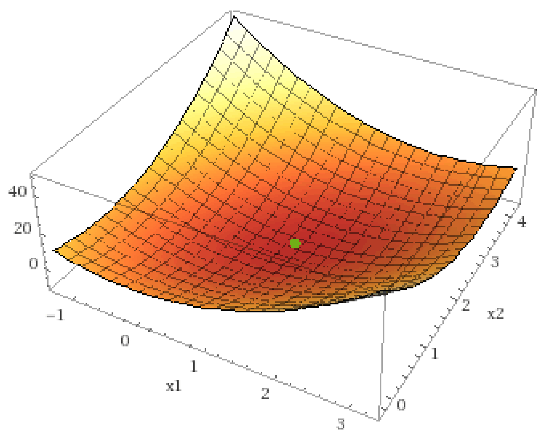
$$x^{(2)} = x^{(1)} - \alpha_1 p^{(1)} = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix} - \frac{1}{6}\begin{pmatrix} -6 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Знайдемо

$$\left\| f'\left(x^{(2)}\right)\right\| = 0.$$

Точка (1,2) – точка мінімуму. Метод спряжених градієнтів для функції двох змінних збігається за два кроки.





Чисельні методи умовної оптимізації

Основна задача опуклого програмування

Постановка основної задачі опуклого програмування:

$$f(x) \to min, \tag{1}$$

$$x \in X = \left\{ x \in E^n : x \in X_0, \quad g_i(x) \le 0, i = \overline{1,m} \right\} \tag{2}$$

Функції $f(x)$ та $g_i(x), i = \overline{1,m}$ вважаються опуклими.

Множина X_0 називається ***множиною простої структури*** (наприклад, куля у *n*–вимірному просторі, *n*–вимірний паралелепіпед, невід'ємний октант). Множина X_0 – опукла, обмежена, замкнута множина.

Множина X також є опуклою.

Функція Лагранжа. Теорема Куна-Таккера

Введемо функцію Лагранжа:

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

яка буде визначена для $x \in X_0$ та $\lambda \in \Lambda_0 = \left\{ \lambda \in E^m : \lambda_i \ge 0 \right\}$.

Сідловою точкою (x_*, λ_*) функції Лагранжа буде точка, яка задовольняє умову:

$$L(x_*, \lambda) \le L(x_*, \lambda_*) \le L(x, \lambda_*), \ \forall x \in X_0, \ \forall \lambda \in \Lambda_0.$$

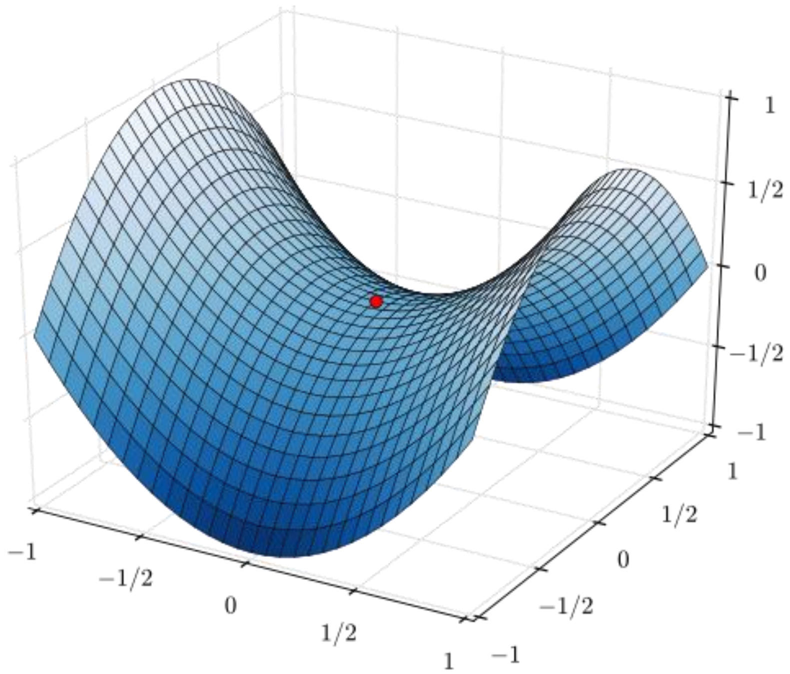


Рис.

Основне місце в теорії опуклого програмування має теорема Куна-Таккера.

Будемо казати, що виконується ***умова Слейтера***, якщо існує точка $\bar{x}$, така, що

$$g_i\left(\bar{x}\right) < 0, \quad i = \overline{1,m}.$$

Це означає, що наша допустима множина містить внутрішні точки.

Теорема Куна-Таккера (в загальній формі). Нехай для задачі (1) виконується умова регулярності Слейтера. Для того, щоб точка x_* була точкою мінімуму функції $f\left(x\right)$ на множині X, необхідно і достатньо, щоб існував такий вектор $\lambda_* \geq 0$, що точка $\left(x_*, \lambda_*\right)$, була сідловою точкою функції Лагранжа $L\left(x, \lambda\right)$ на множині $X_0 \times \Lambda_0$.

Зауваження 1. Згідно з теоремою Куна-Таккера, можна перейти від вихідної задачі опуклого програмування зі складними обмеженнями до задачі опуклого програмування з простими обмеженнями.

Теорема Куна-Таккера (в диференціальної формі). Припустимо, що $X_0 = \left\{ x \in E^n : x_i \geq 0, \quad i = \overline{1,n} \right\}$, $f\left(x\right) \in C^1\left(X_0\right)$, $g_i\left(x\right) \in C^1\left(X_0\right)$, $i = \overline{1,m}$ і опуклі на множині X_0. Для того, щоб точка $\left(x_*, \lambda_*\right)$ була сідловою точкою функції Лагранжа $L\left(x, \lambda\right)$ на множині $X_0 \times \Lambda_0$, необхідно і достатньо, щоб виконувалися умови:

1) $\dfrac{\partial L\left(x_*, \lambda_*\right)}{\partial x} \geq 0, \dfrac{\partial L\left(x_*, \lambda_*\right)}{\partial \lambda} \leq 0,$

2) $\left(x_*, \dfrac{\partial L\left(x_*, \lambda_*\right)}{\partial x} \right) = 0, \left(\lambda_*, \dfrac{\partial L\left(x_*, \lambda_*\right)}{\partial \lambda} \right) = 0,$

3) $x_* \geq 0, \lambda_* \geq 0.$