Методи оптимізації. Лекція 20.05.2022

Метод умовного градієнта

Метод умовного градієнта є ***методом лінійної апроксимації (лінеаризації) цільової функції***.

Розглянемо задачу

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

де X – опукла, замкнена, обмежена множина простору E^n, цільова функція $f(x) \in C^1(X)$.

Ітераційна формула методу умовного градієнта має вигляд:

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

де $h^{(k)}$ – вектор, який вказує напрямок спуску цільової функції $f(x)$ в точці $x^{(k)}$, α_k – параметр, який регулює довжину кроку уздовж $h^{(k)}$.

Для вибору $h^{(k)}$ на k-тій ітерації розв'язується задача мінімізації на множині X лінійної апроксимації цільової функції $f(x)$ в точці $x^{(k)}$, тобто такої функції:

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

де $\left(f'\left(x^{(k)}\right), x - x^{(k)}\right)$ – головна лінійна частина приросту цільової функції $f(x)$ в точці $x^{(k)}$. Відкидаючи константу $f\left(x^{(k)}\right)$, цю задачу можна записати у вигляді

$$f_k(x) = \left(f'\left(x^{(k)}\right), x - x^{(k)}\right) \to min,\ x \in X. \tag{3}$$

Задача (3) у випадку, якщо множина X задається лінійними обмеженнями є задачею лінійного програмування.

Нехай $\overline{x}^{(k)}$ – оптимальний розв'язок задачі (3), а $f_k\left(\overline{x}^{(k)}\right) = \left(f'\left(x^{(k)}\right),\ \overline{x}^{(k)} - x^{(k)}\right)$ – оптимальне значення цільової функції задачі (3). За теоремою Вейерштрасса розв'язок $\overline{x}^{(k)}$ задачі (3) завжди існує. Якщо задача (3) має декілька розв'язків, то вибирається один з них.

Враховуючи що $x^{(k)} \in X$, маємо

$$\min_X f_k(x) = f_k\left(\overline{x}^{(k)}\right) \le f_k\left(x^{(k)}\right) = 0.$$

Тому можливі лише два випадки: $f_k\left(\overline{x}^{(k)}\right) = 0$ або $f_k\left(\overline{x}^{(k)}\right) < 0$.

Якщо $f_k\left(\overline{x}^{(k)}\right) = 0$, то $\left(f'\left(x^{(k)}\right), x - x^{(k)}\right) \ge f_k\left(\overline{x}^{(k)}\right) = 0$ для будь-яких точок $x \in X$, тобто точка $x^{(k)}$ – стаціонарна точка задачі (1) за теоремою 3 про екстремальні властивості функцій, які визначені на опуклих множинах.

Робота алгоритму завершується, точку $x^{(k)}$ треба дослідити на екстремальність.

Якщо ж $f\left(x\right)$ – опукла функція на множині X , то відповідно до другої частини теореми 3, точка $x^{(k)}$ – точка мінімуму задачі (1).

Нехай тепер $f_k\left(\overline{x}^{(k)}\right)<0$. В цьому випадку $\overline{x}^{(k)}\neq x^{(k)}$. Тоді у формулі (2) вважаємо

$$h^{(k)}=\overline{x}^{(k)}-x^{(k)}$$

і ітераційна формула (2) запишеться у вигляді

$$x^{(k+1)}=x^{(k)}+\alpha_k\left(\overline{x}^{(k)}-x^{(k)}\right),\ k=0,1,2,\ldots, \tag{4}$$

Вектор $h^{(k)}$ прийнято називати ***умовним антиградієнтом*** цільової функції $f\left(x\right)$ в точці $x^{(k)}$.

Так як множина допустимих розв'язків X є опуклою, то для будь-якого α_k з відрізку $\left[0,1\right]$ точка $x^{(k+1)}\in X$:

$$x^{(k+1)}=x^{(k)}+\alpha_k\left(\overline{x}^{(k)}-x^{(k)}\right)\in X\ .$$

Способи вибору крокового множника α_k

1. Визначимо α_k з умови:

$$g_k\left(\alpha_k\right)=\min_{0\le\alpha\le1} g_k\left(\alpha\right), \tag{5}$$

$$g_k\left(\alpha\right)=f\left(x^{(k)}+\alpha\, h^{(k)}\right).$$

2. Параметр α_k виберемо апріорно:

$$0<\alpha_k\le1,\ k=0,1,\ldots,\ \sum_{k=0}^{\infty}\alpha_k=\infty,\ \sum_{k=0}^{\infty}\alpha_k^2<\infty\,.$$

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

3. Адаптивний (автоматичний) вибір крокового множника. Параметр α_k вибирається за правилом дроблення до тих пір, поки не виконається нерівність

$$f\left(x^{(k)}+\alpha\left(\overline{x}^{(k)}-x^{(k)}\right)\right)-f\left(x^{(k)}\right)\le\delta\,\alpha\ f_k\left(\overline{x}^{(k)}\right),$$

де δ – параметр методу, $0<\delta<1$.

4. Дроблення кроку. Виберемо $\alpha_k=1$ й перевіримо виконання умови монотонності

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

а потім за необхідності дробимо кроковий множник α_k ($\alpha_k=\lambda\,\alpha_k$, $\lambda\in\left(0,1\right)$) поки не буде виконана умова монотонності.

Умови збіжності методу умовного градієнта

Теорема. Нехай X – замкнута, обмежена, опукла множина простору E^n, цільова функція $f(x) \in C^1(X)$, причому її градієнт задовольняє умову Ліпшиця

$$\left\| f'(x) - f'(z) \right\| \le L \left\| x - z \right\| \text{ для } \forall x, z \in X .$$

Тоді будь-яка гранична точка x_* послідовності $\left\{ x^{(k)} \right\}_0^\infty$, яка визначається за методом (4)-(5), є стаціонарною в задачі (1), тобто

$$\left(f'(x_*), x - x_* \right) \ge 0 \text{ для } \forall x \in X .$$

Якщо при цьому цільова функція $f(x)$ опукла на множині X, то x_* – розв'язок задачі (1) і

$$\lim_{k \to \infty} f\left(x^{(k)} \right) = f_*,$$

де $f_* = \min_{x \in X} f(x)$.

Зауваження. Аналогічні теореми можна сформулювати і для інших способів вибору крокової множника α_k.

Критерії завершення ітераційного процесу:

$$\left| f_k\left(\overline{x}^{(k)} \right) \right| \le \varepsilon ,$$

$$\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon .$$

Відзначимо, що якщо допустима область задається нелінійними обмеженнями, то задача (3) є задачею з ***лінійною*** цільовою функцією і ***нелінійними*** обмеженнями. Для розв'язання цієї задачі необхідно застосовувати відповідні методи умовної оптимізації.

Алгоритм методу умовного градієнта

Початковий етап. Вибрати початкову точку $x^{(0)} \in X$. Задати константу δ ($0 < \delta < 1$), константу зупинки $\varepsilon > 0$, покласти $k = 0$ і перейти до основного етапу.

Основний етап

Крок 1. Обчислити $f'\left(x^{(k)} \right)$. Знайти оптимальний розв'язок $\overline{x}^{(k)}$ задачі

$$\left(f'\left(x^{(k)} \right), x - x^{(k)} \right) \to \min,\ x \in X .$$

Крок 2. Якщо $f_k\left(\overline{x}^{(k)} \right) = \left(f'\left(x^{(k)} \right), x - x^{(k)} \right) = 0$, то зупинитися: $x^{(k)}$ – стаціонарна точка, інакше обчислити умовний антиградієнт $h^{(k)} = \overline{x}^{(k)} - x^{(k)}$, покласти $\alpha_k = 1$.

Крок 3. Перевірити виконання умови

$$f\left(x^{(k)} + \alpha \left(\overline{x}^{(k)} - x^{(k)} \right) \right) - f\left(x^{(k)} \right) \le \delta\, \alpha\, f_k\left(\overline{x}^{(k)} \right). \tag{6}$$

Якщо нерівність виконується, то перейти до кроку 4, інакше провести дроблення крокової множника α_k поки не буде виконано нерівність (6).

Крок 4. Обчислити наступне наближення $x^{(k+1)} = x^{(k)} + \alpha_k h^{(k)}$, замінити k на $k+1$. Якщо $\left\| x^{(k+1)} - x^{(k)} \right\| \le \varepsilon$, то зупинитися, в іншому випадку – перейти до кроку 1.

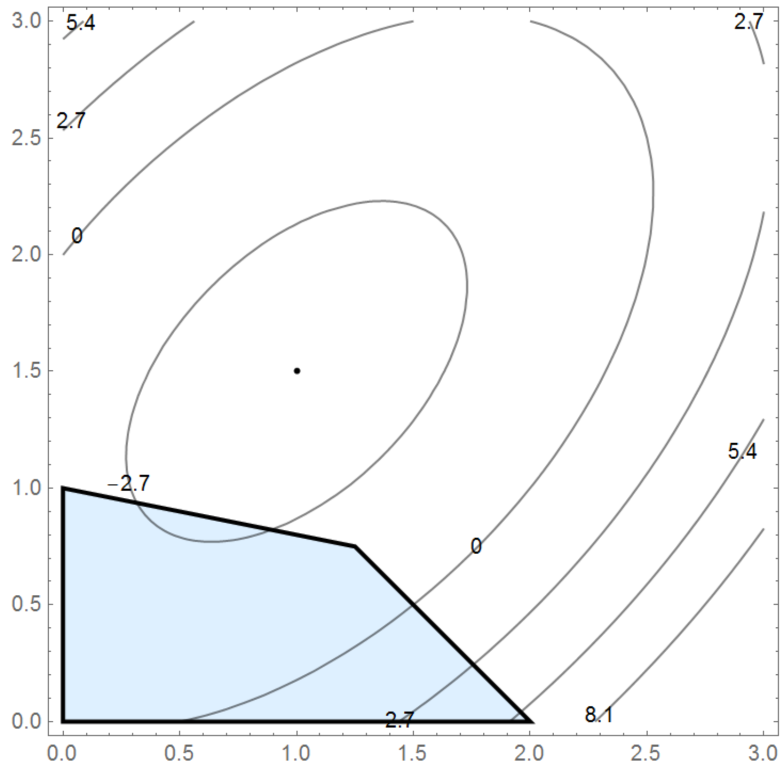
Приклад. Розглянемо задачу

$$f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - x_1 - 4x_2 \to min$$

при обмеженнях

$$x_1 + x_2 \le 2,\ x_1 + 5x_2 \le 5,\ x_1 \ge 0,\ x_2 \ge 0.$$

Лінії рівня цільової функції і допустима область показані на рисунку.



Градієнт цільової функції дорівнює $f'(x) = \begin{pmatrix} 4x_1 - 2x_2 - 1 \\ -2x_1 + 4x_2 - 4 \end{pmatrix}$.

Розв'яжемо цю задачу методом умовного градієнта, взявши за початкову точку $x^{(0)} = (0,0)^T \in X$, $f\left(x^{(0)}\right) = 0$.

Ітерація 1. Пошук напрямку спуску. У початковій точці $x^{(0)} = (0,0)^T$ маємо $f'\left(x^{(0)}\right) = (-1; -4)^T$. Зауважимо, що початкова точка повинна належати множині допустимих розв’язків.

Задача знаходження напрямку $h^{(0)}$ має вигляд:

$$f_0(x) = \left(f'\left(x^{(0)}\right), x - x^{(0)}\right) = \begin{pmatrix} -1 \\ -4 \end{pmatrix} \begin{pmatrix} x_1 - 0 \\ x_2 - 0 \end{pmatrix} = -x_1 - 4x_2 \to min,$$

$$x_1 + x_2 \le 2,$$
$$x_1 + 5x_2 \le 5,$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Сформульована задача є задачею лінійного програмування і може бути розв'язана симплекс-методом.

Оптимальним розв'язком цієї задачі є точка

$$\overline{x}^{(0)} = \left(\frac{5}{4}, \frac{3}{4}\right)^T,\ f_0\left(\overline{x}^{(0)}\right) = -\frac{17}{4} = -4.25.$$

Так як $f_0\left(\overline{x}^{(0)}\right) < 0$, обчислюємо вектор

$$h^{(0)} = \overline{x}^{(0)} - x^{(0)} = \left(\frac{5}{4}, \frac{3}{4}\right)^T.$$

Лінійний пошук. Будь-яка точка $x^{(1)}$, яку знайдено з точки $x^{(0)}$ у напрямку $h^{(0)}$, може бути представлена у вигляді

$$x^{(1)} = x^{(0)} + \alpha h^{(0)},\ \text{де}\ 0 \le \alpha \le 1.$$

Тобто маємо

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 5/4 \\ 3/4 \end{pmatrix} = \left(\frac{5\alpha}{4}, \frac{3\alpha}{4}\right)^T,$$

а відповідне їй значення цільової функції дорівнює $f\left(x^{(0)} + \alpha h^{(0)}\right) = \frac{19}{8}\alpha^2 - \frac{17}{4}\alpha$.

Значення α_0 знаходиться з розв'язання такої задачі одновимірної оптимізації:

$$\frac{19}{8}\alpha^2 - \frac{17}{4}\alpha \to min,\ 0 \le \alpha \le 1.$$

Оптимальний розв'язок дорівнює $\alpha_0 = \frac{17}{19}$.

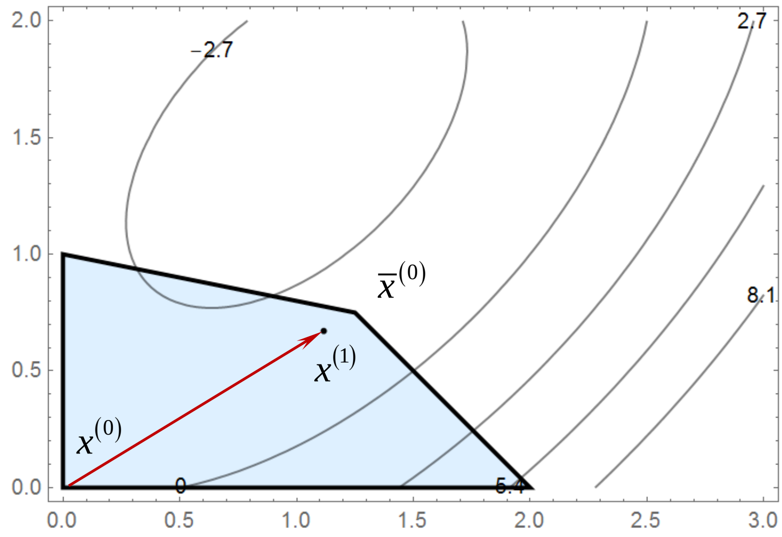
Отже,

$$x^{(1)} = \left(\frac{85}{76}, \frac{51}{76}\right)^T \approx (1.118, 0.671)^T,\ f\left(x^{(1)}\right) = -1.901.$$

Нехай константа зупинки $\varepsilon = 10^{-1}$. Знайдемо норму

$$\left\| x^{(1)} - x^{(0)} \right\| = \sqrt{(1.118 - 0)^2 + (0.671 - 0)^2} = 1.3039 > \varepsilon.$$

Критерій зупинки не виконано.



Ітерація 2. Пошук напрямку. В точці $x^{(1)} = \left(\frac{85}{76}, \frac{51}{76}\right)^T$ маємо $f'\left(x^{(1)}\right) = \left(\frac{81}{38}, -\frac{135}{38}\right)^T$.

Для знаходження напрямку $h^{(1)}$ розв'яжемо таку задачу:

$$f_1(x) = \frac{81}{38}x_1 - \frac{135}{38}x_2 \to min,$$
$$x_1 + x_2 \le 2,$$
$$x_1 + 5x_2 \le 5,$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Оптимальним розв'язком цієї задачі є точка

$$\overline{x}^{(1)} = (0,1)^T,\ f_1\left(\overline{x}^{(1)}\right) = -3.55.$$

Так як $f_1\left(\overline{x}^{(1)}\right) < 0$, обчислюємо вектор

$$h^{(1)} = \overline{x}^{(1)} - x^{(1)} = \left(-\frac{85}{76}, \frac{25}{76}\right)^T.$$

Лінійний пошук. Значення кроку α_1 отримаємо мінімізацією функції

$$f\left(x^{(1)} + \alpha\, h^{(1)}\right) = 3.45\,\alpha^2 - 3.55\,\alpha - 1.9 \text{ за умови } 0 \le \alpha \le 1.$$

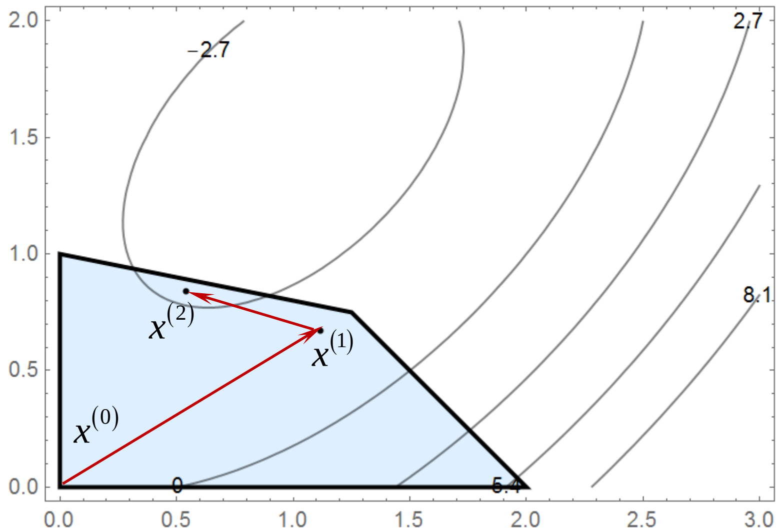
Оптимальний розв'язок $\alpha_1 \approx 0.514$.

Отже, $x^{(2)} = (0.54, 0.84)^T$, $f\left(x^{(2)}\right) = -2.815$.

Знайдемо норму

$$\left\|x^{(2)} - x^{(1)}\right\| = \sqrt{(0.54 - 1.118)^2 + (0.84 - 0.671)^2} = 0.6022 > \varepsilon.$$

Критерій зупинки не виконано.



Ітерація 3. Пошук напрямку. В точці $x^{(2)}=(0.54, 0.84)^T$ маємо $f'\left(x^{(2)}\right)=(-0.5075,-1.7256)^T$.

Задача для знаходження напрямку $h^{(2)}$ має вигляд:

$$f_2(x)=-0.5075x_1-1.7256x_2+1.7235\to min,$$
$$x_1+x_2\le 2,$$
$$x_1+5x_2\le 5,$$
$$x_1\ge 0,\ x_2\ge 0.$$

Оптимальним розв'язком цієї задачі є точка

$$\overline{x}^{(2)}=\left(\frac{5}{4},\frac{3}{4}\right)^T,\ f_2\left(\overline{x}^{(2)}\right)=-0.205.$$

Так як $f_2\left(\overline{x}^{(2)}\right)<0$, обчислюємо вектор

$$h^{(2)}=\overline{x}^{(2)}-x^{(2)}=\begin{pmatrix}1.25\\0.75\end{pmatrix}-\begin{pmatrix}0.54\\0.84\end{pmatrix}=\begin{pmatrix}0.71\\-0.09\end{pmatrix}.$$

Лінійний пошук. Значення кроку α_2 отримаємо мінімізацією функції

$$f\left(x^{(2)}+\alpha\,h^{(2)}\right)=f\left(0.54+0.71\alpha, 0.84-0.09\alpha\right)=$$
$$=1.1522\,\alpha^2-0.205\,\alpha-2.815$$

за умови $0\le\alpha\le 1$.

Оптимальний розв'язок $\alpha_2=0.089$.

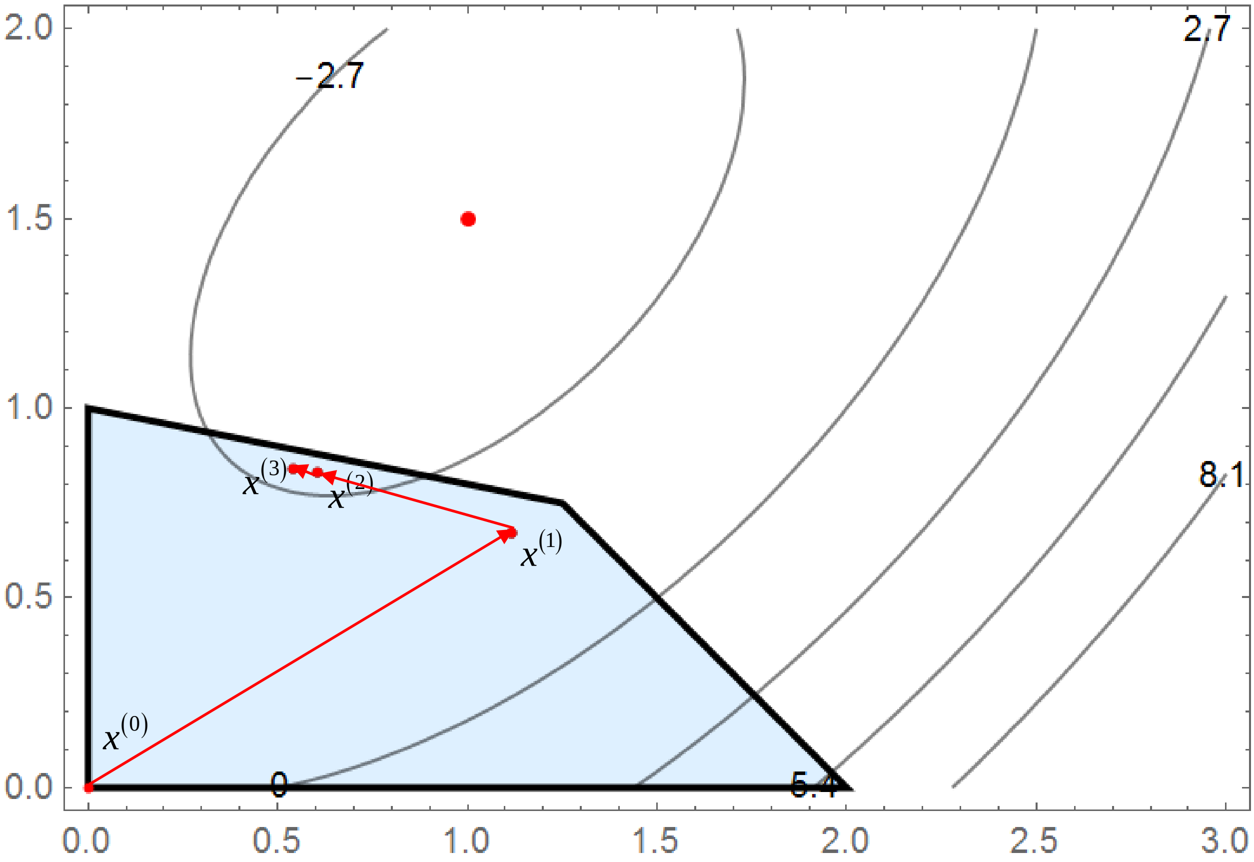
Отже,

$$x^{(3)}=(0.603, 0.832)^T,\ f\left(x^{(3)}\right)=-2.823.$$

Знайдемо норму

$$\left\|x^{(3)}-x^{(2)}\right\|=\sqrt{(0.603-0.54)^2+(0.832-0.84)^2}=0.0635<\varepsilon.$$

Отже, точка $x^{(3)} = (0.603, 0.832)^T$ є оптимальним розв'язком задачі зі заданою точністю.



Метод гіперплощини, що відтинає

Основна ідея методу полягає в тому, що допустима область апроксимується деяким ***багатогранником***, який зменшується від одного ітераційного кроку до іншого, при цьому все краще апроксимуючи допустиму область в околі розв'язку.

Метод гіперплощини, що відтинає (або ***метод Келлі січних площин***) може застосовуватися до загальних задач опуклого програмування з ***нелінійною*** цільовою функцією і ***нелінійними*** обмеженнями:

$$f(x) \to min, \qquad (1)$$

$$g_i(x) \le 0, \quad i = \overline{1,m}.$$

Функції $f(x)$ та $g_i(x)$ є опуклими і неперервно-диференційовними функціями.

Введемо додаткову змінну і перепишемо задачу (1) у вигляді:

$$y \to min,$$

$$f(x) - y \le 0, \qquad (2)$$

$$g_i(x) \le 0, \quad i = \overline{1,m}.$$

Задача (2) є задачею опуклого програмування з ***лінійною*** цільовою функцією і ***нелінійними*** обмеженнями в просторі E^{n+1}.

Оскільки показано, як здійснити перехід від задачі (1) з нелінійною цільовою функцією до задачі (2) з лінійною цільовою функцією, подальше викладення методу проведемо для такої задачі опуклого програмування:

$$(c, x) \to min,$$
$$g_i(x) \le 0, \quad i = \overline{1,m}, \tag{3}$$

де $g_i(x)$ – опуклі і неперервно-диференційовні функції, $x \in E^n$.

Позначимо через $X = \left\{ x \in E^n : g_i(x) \le 0, \quad i = \overline{1,m} \right\}$ множину допустимих розв'язків задачі (3).

Апроксимуємо множину X багатогранною множиною Z. У випадку, коли X – обмежена, замкнута множина, то Z – багатогранник.

Ідея методу полягає у відсікання на кожному кроці частини багатогранної множини, яка не містить точку мінімуму початкової задачі.

Опишемо $k-$й крок методу. Розв'язуємо задачу лінійного програмування:

$$(c, x) \to min,$$
$$x \in Z_k, \tag{4}$$

Нехай $x^{(k)}$ – оптимальний розв'язок задачі (4).

Якщо

$$g_i\left(x^{(k)}\right) \le 0, \quad \forall i = \overline{1,m},$$

то $x^{(k)}$ – оптимальний розв'язок початкової задачі (3).

Інакше, знайдемо

$$g_j\left(x^{(k)}\right) = \max_{1 \le i \le m} g_i\left(x^{(k)}\right).$$

Нова багатогранна множина Z_{k+1} виходить додаванням до множини Z_k додаткового обмеження:

$$g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) \le 0.$$

Отже,

$$Z_{k+1} = Z_k \cap \left\{ x\colon\ g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) \le 0 \right\}. \tag{5}$$

Гіперплощина

$$g_j\left(x^{(k)}\right) + g_j'\left(x^{(k)}\right)\left(x - x^{(k)}\right) = 0 \tag{6}$$

строго відокремлює точку $x^{(k)}$ від множини X і на цій підставі вона називається ***січною гіперплощиною***.

Теорема про збіжність методу січних площин. Нехай $f(x)$, $g_i(x), i = \overline{1,m}$ є опуклими і неперервно-диференційовними функціями. Якщо задача (3) має мінімум на скінченній відстані, то будь-яка точка згущення послідовності $\left\{ x^{(k)} \right\}$, яка генерується методом гіперплощини, що відтинає, є оптимальним розв'язком задачі (3).

Алгоритм методу

Початковий етап. Апроксимувати множину X багатогранною множиною Z_1. Покласти $k=1$, задати $\varepsilon>0$ і перейти до основного етапу.

Основний етап

Крок 1. Розв'язати задачу лінійного програмування (4):

$$(c,x)\to min,$$
$$x\in Z_k.$$

Нехай точка $x^{(k)}$ – оптимальний розв'язок задачі (4).

Якщо $g_i\left(x^{(k)}\right)\le 0$ для всіх $i=\overline{1,m}$, то зупинитися: $x^{(k)}$ – оптимальний розв'язок вихідної задачі. В іншому випадку перейти до кроку 2.

Крок 2. Визначити індекс j, для якого $g_j\left(x^{(k)}\right)=\max\limits_{1\le i\le m} g_i\left(x^{(k)}\right)$. Додати до обмежень, які визначають множину Z_k, додаткове обмеження вигляду

$$g_j\left(x^{(k)}\right)+g_j'\left(x^{(k)}\right)\left(x-x^{(k)}\right)\le 0.$$

Отже, множина Z_{k+1} має вигляд

$$Z_{k+1}=Z_k\cap\left\{\ g_j\left(x^{(k)}\right)+g_j'\left(x^{(k)}\right)\left(x-x^{(k)}\right)\le 0\ \right\}.$$

Замінити k на $k+1$ та перейти до кроку 1.

Крок 3. Якщо $\left\|x^{(k+1)}-x^{(k)}\right\|\le\varepsilon$, то зупинитися: $x^{(k)}$ – наближення до точки мінімуму. В іншому випадку перейти до кроку 1.

Алгоритм описано.