Optymalizacja nieliniowa

Część III. Optymalizacja z ograniczeniami

- Metoda funkcji kary zewnętrznej

- Metoda funkcji kary wewnętrznej

WPROWADZENIE

Rozwiązywanie rzeczywistych problemów optymalizacyjnych związane jest często z potrzebą uwzględnienia różnorodnych ograniczeń. Ograniczenia te zawężają zbiór rozwiązań i muszą być uwzględnianie w procesie poszukiwania minimum lub maksimum funkcji celu. W modelu matematycznym zagadnienia przybierają one postać równości lub nierówności.

Niech dany będzie problem polegający na wyznaczeniu minimum funkcji celu

$$f : \mathbb{R}^n \to \mathbb{R}$$

przy ograniczeniach:

$$g_j(\mathbf{x}) \leqslant 0 \text{ dla } j = 1, 2, \ldots, k; \tag{1}$$

$$h_j(\mathbf{x}) = 0 \text{ dla } j = 1, 2, \ldots, m, \tag{2}$$

gdzie $g_j, h_j : \mathbb{R}^n \to \mathbb{R}$ są funkcjami ograniczającymi zbiór poszukiwań.

Zbiór tych $\mathbf{x} \in \mathbb{R}^n$, które spełniają układ warunków (1) i (2) nazywamy zbiorem rozwiązań dopuszczalnych lub po prostu zbiorem dopuszczalnym, i oznaczamy przez X_d.

Istota przedstawionych poniżej metod optymalizacji z ograniczeniami polega na pewnej modyfikacji funkcji celu polegającej na włączeniu do niej ograniczeń. W ten sposób zagadnienie optymalizacji z ograniczeniami zostaje przekształcone na zagadnienie optymalizacji bez ograniczeń. Do optymalizacji tak zmodyfikowanej funkcji celu można użyć jednej z metod optymalizacji bez ograniczeń.

Poszczególne metody różnią się sposobem tworzenia zmodyfikowanej funkcji celu. W każdym przypadku ma jednak zastosowanie ten sam ogólny algorytm.

Punktem wyjścia jest ustalenie punktu startowego $x^{(0)}$, tzn. pierwszego przybliżenia szukanego minimum i dokładności obliczeń ε. Następnie wyznaczamy zmodyfikowaną funkcję celu F_i (różną w każdej iteracji) i znajdujemy jej minimum x^*. Znalezione minimum staje się kolejnym przybliżeniem, tzn. $x^{(i)} = x^*$. Jeżeli teraz

$$\|x^{(i)} - x^{(i-1)}\|_2 \leqslant \varepsilon,$$

algorytm kończy działanie - przybliżeniem rzeczywistego minimum jest $x^{(i)}$. W przeciwnym wypadku wyznaczamy F_{i+1} i powtarzamy rozumowanie.

Metoda funkcji kary zewnętrznej

W metodzie funkcji kary zewnętrznej modyfikacji funkcji celu dokonuje się w oparciu o tak zwaną funkcję kary zewnętrznej $S : \mathbb{R}^n \to \mathbb{R}$. Funkcja ta powinna być ciągła i taka, że

$$S(\mathbf{x}) = 0 \text{ dla } \mathbf{x} \in X_d$$

i

$$S(\mathbf{x}) > 0 \text{ dla } \mathbf{x} \notin X_d.$$

Uwaga. Można powiedzieć, że funkcja S przypisuje każdemu punktowi spoza zbioru rozwiązań dopuszczalnych pewną karę liczbową.

Zazwyczaj funkcja kary zewnętrznej jest realizowana przez funkcję kwadratową, tzn. ma pastać:

$$S(\mathbf{x}) = \sum_{j=1}^{k} \Big(\max(0, g_j(\mathbf{x}))\Big)^2 + \sum_{j=1}^{m} \Big(h_j(\mathbf{x})\Big)^2 .$$

Łatwo sprawdzić, że $S(\mathbf{x}) = 0$ tylko wtedy, gdy ograniczenia są spełnione. Natomiast $S(x) > 0$, gdy ograniczenia nie są spełnione.

Uwaga. Funkcja S jest ciągła. Występowanie w powyższym wzorze funkcji max może powodować, że w niektórych przypadkach funkcja S nie będzie różniczkowalna. W konsekwencji nie zawsze będzie można stosować metody gradientowe do optymalizacji.

Posiadając funkcję kary możemy teraz określić ciąg zmodyfikowanych funkcji celu $F_i : \mathbb{R}^n \to \mathbb{R}$ kierując się zasadą, aby wraz ze wzrostem indeksu rosła wartość kary przypisywanej punktom spoza zbioru X_d. Powyższą zasadę spełnia ciąg:

$$F_i(\mathbf{x}) = f(\mathbf{x}) + c_i S(\mathbf{x}),$$

gdzie c_i jest rosnącym ciągiem o wyrazach dodatnich takim, że $\lim_{i \to \infty} c_i = \infty$. Zazwyczaj przyjmuje się

$$\begin{cases} c_1 > 0, \\ c_{i+1} = 2c_i. \end{cases}$$

Metoda funkcji kary wewnętrznej

Niech dany będzie problem polegający na wyznaczeniu minimum funkcji celu

$$f : \mathbb{R}^n \to \mathbb{R}$$

przy ograniczeniach postaci:

$$g_j(\mathbf{x}) \leqslant 0 \text{ dla } j = 1, 2, \ldots, k.$$

Uwaga. Metoda funkcji kary zewnętrznej ma zastosowanie jedynie do zagadnień, w których ograniczenia występują tylko w postaci nierówności.

W metodzie funkcji kary wewnętrznej modyfikacji funkcji celu dokonuje się w oparciu o tak zwaną funkcję barierową

$$S : \mathrm{int}X_d \to \mathbb{R}^+.$$

Funkcja ta powinna być

(1) ciągła we wnętrzu zbioru dopuszczalnego,

(2) o wartościach dodatnich, tzn. $S(\mathbf{x}) > 0$ dla $\mathbf{x} \in \mathrm{int}X_d$,

(3) taka, że dla każdego ciągu punktów $\mathbf{x}^{(n)}$ dążącego do brzegu zbioru X_d, tzn. dla ciągu spełniającego warunek:

$$\lim_{n\to\infty} g_j(\mathbf{x}^{(n)}) = 0$$

dla ustalonego $j = 1, 2, \ldots, k$, zachodzi jednocześnie:

$$\lim_{n\to\infty} S(\mathbf{x}^{(n)}) = \infty.$$

Uwaga. Można powiedzieć, że funkcja barierowa S przypisuje każdemu punktowi wewnątrz zbioru rozwiązań dopuszczalnych pewną karę liczbową i tworzy „barierę'' na brzegu tego obszaru.

Uwaga. Własności funkcji barierowej powinny gwarantować, że ciąg kolejnych przybliżeń rzeczywistego minimum pozostanie w zbiorze dopuszczalnym. Może jednak się zdarzyć, że zastosowana metoda optymalizacji bez ograniczeń spowoduje ''przerzucenie'' tego ciągu ponad barierą i niewłaściwe działanie całej metody.

Uwaga. W celu właściwego działania metody punkt startowy $\mathbf{x}^{(0)}$ musi być tak wybrany, aby $\mathbf{x}^{(0)} \in X_d$. Ponadto w procesie optymalizacji, należy kontrolować czy ciąg kolejnych przybliżeń jest taki, że $\mathbf{x}^{(i)} \in X_d$.

Jako funkcja barierowa można być użyta jedna z dwóch funkcji:

$$S(\mathbf{x}) = -\sum_{j=1}^{k} \frac{1}{g_j(\mathbf{x})} \text{ dla } \mathbf{x} \in X_d$$

lub

$$S(\mathbf{x}) = -\sum_{j=1}^{k} \ln\Big(-g_j(\mathbf{x})\Big) \text{ dla } \mathbf{x} \in X_d.$$

Posiadając funkcję kary wewnętrznej możemy teraz, we wnętrzu zbioru dopuszczalnego, określić ciąg zmodyfikowanych funkcji celu $F_i : \text{int}X_d \to \mathbb{R}$ kierując się zasadą, aby wraz ze wzrostem indeksu malała kara nakładana punktom, tzn. malało znaczenie „bariery''. Powyższą zasadę spełnia ciąg:

$$F_i(\mathbf{x}) = f(\mathbf{x}) + c_i S(\mathbf{x}),$$

gdzie c_i jest malejącym ciągiem o wyrazach dodatnich takim, że $\lim\limits_{i\to\infty} c_i = 0$. Zazwyczaj przyjmuje się

$$\begin{cases} c_1 > 0, \\ c_{i+1} = \alpha \cdot c_i, \end{cases}$$

gdzie $\alpha \in (0, 1)$.