Metody optymalizacji. Wykład 6

Metoda Lagrange'a

Na dzisiejszym wykładzie podamy inne sformułowanie twierdzenia Kuhna - Tuckera, używając pojęć funkcji Lagrange'a i punktu siodłowego. Wprowadzone tu idee w naturalny sposób prowadzą do sformułowania zadania dualnego do danego problemu optymalizacyjnego, które wprowadzimy za tydzień. Będzie ono grać ważną rolę w dalszej części naszego kursu.

Rozważmy funkcję $L : \Gamma \times \mathbb{R}^m_+ \to \mathbb{R}$, gdzie $\Gamma \subseteq \mathbb{R}^n$ jest niepustym zbiorem wypukłym. Załóżmy, że dla każdego $y \in \mathbb{R}^m_+$ $L(\cdot, y)$ jest wypukła w Γ i dla każdego $x \in \Gamma$ funkcja $L(x, \cdot)$ jest wklęsła w $\mathbb{R}^m_+$.

Definicja 1 *Parę $(x^*, y^*) \in \Gamma \times \mathbb{R}^m_+$ nazywamy punktem siodłowym funkcji L w $\Gamma \times \mathbb{R}^m_+$, jeśli dla dowolnych $x \in \Gamma$ i $y \in \mathbb{R}^m_+$,*

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*).$$

W praktyce będziemy rozpatrywać $\Gamma = \mathbb{R}^n$ lub $\Gamma = \mathbb{R}^n_+$.

Twierdzenie 1 *Niech $\Gamma = \mathbb{R}^n_+$. Zakładamy, że funkcja $L : \Gamma \times \mathbb{R}^m_+ \to \mathbb{R}$ jest wypukła względem x dla każdego $y \in \mathbb{R}^m_+$, wklęsła względem y dla każdego $x \in \Gamma$, i jest klasy C^1. Wtedy punkt $(x^*, y^*) \in \Gamma \times \mathbb{R}^m_+$ jest punktem siodłowym funkcji L w $\Gamma \times \mathbb{R}^m_+$ wtedy i tylko wtedy, gdy spełnione są następujące warunki:*
(i) $\frac{\partial L}{\partial x}(x^*, y^*) \geq 0$,
(ii) $(x^*, \frac{\partial L}{\partial x}(x^*, y^*)) = 0$,
(iii) $\frac{\partial L}{\partial y}(x^*, y^*) \leq 0$,
(iv) $(y^*, \frac{\partial L}{\partial y}(x^*, y^*)) = 0$.

Uwaga. W (i)-(iv), $\frac{\partial L}{\partial x} = \nabla_x L \in \mathbb{R}^n$, $\frac{\partial L}{\partial y} = \nabla_y L \in \mathbb{R}^m$.

Dowód. Ponieważ $x^* \geq 0$, $y^* \geq 0$, warunki (i)-(iv) są równoważne warunkom (i), (ii'), (iii), (iv'), gdzie
(ii') $x_i^* \frac{\partial L}{\partial x_i}(x^*, y^*) = 0$, $i = 1, ..., n$,
(iv') $y_i^* \frac{\partial L}{\partial y_i}(x^*, y^*) = 0$, $i = 1, ..., m$.

$\Rightarrow$ Jeśli (x^*, y^*) jest punktem siodłowym L w $\Gamma \times \mathbb{R}^m_+$, to dla każdego $i = 1, ..., n$ i $x_i \geq 0$,

$$g_i(x_i) := L(x_1^*, ..., x_{i-1}^*, x_i, x_{i+1}^*, .., x_n^*, y^*) \geq L(x^*, y^*) = g_i(x_i^*),$$

zatem funkcja $g_i : \mathbb{R}_+ \to \mathbb{R}$ ma minimum w x_i^*. Stąd $\frac{d}{dx_i} g_i(x_i^*) = \frac{\partial L}{\partial x_i}(x^*, y^*) \geq 0$ i ostra nierówność może zachodzić tylko wtedy, gdy $x_i^* = 0$. Stąd (i), (ii') są spełnione. Podobnie pokazuje się, że zachodzą warunki (iii), (iv').

$\Leftarrow$ Załóżmy (i)-(iv). Ponieważ L jest wypukła względem x, na mocy twierdzenia 2 z czwartego wykładu, mamy dla każdego $x \in \mathbb{R}_+^n$

$$L(x, y^*) \geq L(x^*, y^*) + \Big(x - x^*, \frac{\partial L}{\partial x}(x^*, y^*)\Big).$$

Stąd, na mocy założeń (i)-(ii) i nierówności $x \geq 0$, otrzymujemy

$$L(x, y^*) \geq L(x^*, y^*) + \Big(x, \frac{\partial L}{\partial x}(x^*, y^*)\Big) - \Big(x^*, \frac{\partial L}{\partial x}(x^*, y^*)\Big) \geq L(x^*, y^*).$$

Podobnie dowodzimy nierówności $L(x^*, y) \leq L(x^*, y^*)$. $\square$

Uwaga. Analogiczne twierdzenie w przypadku $\Gamma = \mathbb{R}^n$ otrzymuje się zastępując warunki (i)-(ii) przez $\frac{\partial L}{\partial x}(x^*, y^*) = 0$.

W dalszej części wykładu rozważać będziemy następujące zadanie optymalizacyjne, które będziemy cytować jako **zadanie 1**.
Niech $\Gamma \subseteq \mathbb{R}^n$ będzie zbiorem domkniętym i wypukłym. Niech funkcje $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., m$, będą wklęsłe. Zdefiniujmy odwzorowanie $f : \mathbb{R}^n \to \mathbb{R}^m$ wzorem $f(x) = (f_1(x), ..., f_m(x))$, $x \in \mathbb{R}^n$, i niech

$$X = \{x \in \Gamma : f(x) \geq 0\}.$$

Niech $\phi : \mathbb{R}^n \to \mathbb{R}$ będzie funkcją wypukłą. Szukamy minimum ϕ na X.

Uwaga. Zadanie 1 jest zadaniem programowania wypukłego. Co więcej, jeśli $\Gamma = \mathbb{R}^n$, to nasze zadanie jest podstawowym zadaniem programowania wypukłego.

Definicja 2 *Funkcję*

$$L(x, y) = \phi(x) - (y, f(x)), \qquad x \in \mathbb{R}^n,\ y \in \mathbb{R}_+^m, \tag{1}$$

nazywamy funkcją Lagrange'a dla zadania 1.

Uwaga. W analizie, do szukania ekstremów warunkowych funkcji ϕ (tzn. ekstremów ϕ na $X = \{x \in \Gamma : f(x) = 0\}$), przydatna jest *metoda mnożników Lagrange'a*, polegająca na szukaniu ekstremów warunkowych ϕ wśród punktów stacjonarnych $L(x, y)$, tzn. (x, y) takich, że $\frac{\partial L}{\partial x}(x, y) = 0$, $\frac{\partial L}{\partial y}(x, y) = 0$. Funkcja Lagrange'a, którą obecnie omawiamy, pełni w zadaniach programowania wypukłego (w szczególności liniowego) podobną rolę, jak mnożniki Lagrange'a w analizie.

Twierdzenie 2 *Niech* $(x^*, y^*) \in \Gamma \times \mathbb{R}^m_+$ *będzie punktem siodłowym funkcji Lagrange'a* (1) *na* $\Gamma \times \mathbb{R}^m_+$. *Wtedy* x^* *jest punktem optymalnym dla zadania 1.*

Dowód. Z definicji punktu siodłowego, dla dowolnych $x \in \Gamma$ i $y \in \mathbb{R}^m_+$ mamy

$$\begin{aligned} \phi(x^*) - (y, f(x^*)) \;=\; L(x^*, y) \;&\leq\; L(x^*, y^*) = \phi(x^*) - (y^*, f(x^*)) \\ &\leq\; L(x, y^*) \;=\; \phi(x) - (y^*, f(x)). \end{aligned} \tag{2}$$

Na mocy pierwszej z tych nierówności,

$$(y, f(x^*)) \geq (y^*, f(x^*)), \qquad y \in \mathbb{R}^m_+, \tag{3}$$

a zatem $f(x^*) \geq 0$. (Istotnie, gdyby $f_i(x^*) < 0$ dla pewnego $i \in \{1, ..., m\}$, kładąc $y_j = \delta_{ij} c$, $j = 1, ..., m$, gdzie $c \to \infty$, mielibyśmy $(y, f(x^*)) = c f_i(x^*) \to -\infty$, co przeczy (3).) W szczególności $x^* \in X$.

Dla $y = 0$, mamy z (3) $0 \geq (y^*, f(x^*))$. Ale $y^* \geq 0$, $f(x^*) \geq 0$, więc $(y^*, f(x^*)) \geq 0$. Zatem

$$(y^*, f(x^*)) = 0. \tag{4}$$

Jeśli $x \in X$, to $f(x) \geq 0$, a więc

$$(y, f(x^*)) \geq 0. \tag{5}$$

Z prawej nierówności w (2), (4) i (5), mamy dla każdego $x \in \Gamma$ (a więc w szczególności dla każdego $x \in X$)

$$\phi(x^*) = \phi(x^*) - (y^*, f(x^*)) \leq \phi(x) - (y^*, f(x)) \leq \phi(x),$$

zatem x^* jest optymalny w X. □

Uwaga. Założenia wypukłości Γ i ϕ czy wklęsłości f_i, $i = 1, ..., m$, nie były w powyższym dowodzie używane, podobnie jak jakiekolwiek założenia regularności (gładkości), a zatem znalezienie punktu siodłowego(x^*, y^*) funkcji Lagrange'a $L(x, y)$ daje optymalność x^* w podstawowym zadaniu programowania matematycznego. W praktyce jednak twierdzenie 2 stosuje zwykle się jako część następującego twierdzenia, w którym wszystkie powyższe założenia są potrzebne.

Twierdzenie 3 (Kuhna - Tuckera, w wersji dla funkcji Lagrange'a) *Niech*

$$X = \{x \in \mathbb{R}^n : f_i(x) \geq 0, i = 1, ...m, x \geq 0\}. \tag{6}$$

Zakładamy, że funkcja ϕ jest wypukła, f_i, $i = 1, ..., m$, wklęsłe, ϕ i f_i, $i = 1, ..., m$, są klasy C^1, zaś zbiór X jest regularny w sensie Slatera. Wtedy warunkiem koniecznym i wystarczającym na to, by $x^* \in X$ był punktem optymalnym w podstawowym zadaniu programowania wypukłego $\min\{\phi(x) : x \in X\}$ *jest istnienie takiego $y^* \in \mathbb{R}^m_+$, że para (x^*, y^*) jest punktem siodłowym funkcji Lagrange'a L danej przez (1) na zbiorze $\mathbb{R}^n_+ \times \mathbb{R}^m_+$.*

Dowód. Dostateczność warunku wynika z twierdzenia 2 z $\Gamma = \mathbb{R}^n_+$.

Dla dowodu konieczności, załóżmy, że x^* jest punktem optymalnym w rozważanym zadaniu. Niech $e_j = (0, ..., 0, 1, 0, ...0)$ (gdzie 1 występuje wyłącznie na j-tej pozycji), $j = 1, ..., n$, będzie bazą kanoniczną w $\mathbb{R}^n$ i niech dla $j = 1, ..., n$, $f_{m+j}(x) = x_j$, $x = (x_1, ..., x_n)$. Zauważmy, że $\nabla f_{m+j} = e_j$, $j = 1, ...n$. Wtedy

$$X = \{x \in \mathbb{R}^n : f_i(x) \geq 0, i = 1, ..., m + n\} \tag{7}$$

(zakładana regularność zbioru dopuszczalnego znaczy, że X opisany przez (7) jest regularny w sensie opisanym w poprzednim wykładzie). Z twierdzenia Kuhna - Tuckera (twierdzenia 5 z ostatniego wykładu), istnieją $y_1^*, ..., y_m^* \geq 0$, $v_1^*, ..., v_n^* \geq 0$ takie, że

$$\nabla\phi(x^*) = \sum_{i=1}^{m} y_i^* \nabla f_i(x^*) + \sum_{j=1}^{n} v_j^* e_j, \tag{8}$$

$$0 = \sum_{i=1}^{m} y_i^* f_i(x^*) + \sum_{j=1}^{n} v_j^* x_j^*. \tag{9}$$

Ale $y_i^* \geq 0$, $f_i(x^*) \geq 0$, $v_j^* \geq 0$ i $x_j^* \geq 0$, zatem (9) implikuje

$$\sum_{i=1}^{m} y_i^* f_i(x^*) = 0, \tag{10}$$

$$\sum_{j=1}^{n} v_j^* x_j^* = 0. \tag{11}$$

Niech $y^* = (y_1^*, ..., y_m^*)$. Oczywiście $y^* \in \mathbb{R}^m_+$. Sprawdzimy, że para (x^*, y^*) spełnia warunki (i)-(iv) twierdzenia 1, a zatem jest punktem siodłowym L w $\mathbb{R}^n_+ \times \mathbb{R}^m_+$. Istotnie, na mocy (1) i (8),

$$
\begin{aligned}
\frac{\partial L}{\partial x}(x^*, y^*) &= \nabla\phi(x^*) - \frac{\partial}{\partial x}\Big(\sum_{i=1}^{m} y_i f_i(x)\Big)_{|(x^*,y^*)} \\
&= \nabla\phi(x^*) - \sum_{i=1}^{m} y_i^* \nabla f_i(x^*) \\
&= \sum_{j=1}^{n} v_j^* e_j \;=:\; v^* \;\geq 0,
\end{aligned}
$$

gdzie $v^* = (v_1^*, ..., v_n^*)$. Zatem warunek (i) twierdzenia 1 zachodzi. Co więcej, na mocy (11),

$$
\Big(x^*, \frac{\partial L}{\partial x}(x^*, y^*)\Big) = (x^*, v^*) = \sum_{j=1}^{n} v_j^* x_j^* = 0,
$$

co daje warunek (ii). Dalej,

$$
\frac{\partial L}{\partial y}(x^*, y^*) = \frac{\partial}{\partial y}\Big(\phi(x) - (y, f(x)\Big)_{|(x^*,y^*)} = -f(x^*) \leq 0,
$$

$$
\Big(y^*, \frac{\partial L}{\partial y}(x^*, y^*)\Big) = -(y^*, f(x^*)) = -\sum_{i=1}^{m} y_i^* f_i(x^*) = 0,
$$

gdzie ostatnia równość wynika z (10), a zatem warunki (iii)-(iv) też są spełnione.

□

Przykład. Niech $n = 2$, $m = 3$. Dla $x = (x_1, x_2) \in \mathbb{R}^2$, niech $\phi(x) = x_1 + 4x_2$, $f_1(x) = 5x_1 + x_2 - 2$, $f_2(x) = 3x_1 + 3x_2 - 3$, $f_3(x) = x_1 + 3x_2 - 2$. Chcemy zminimalizować ϕ na zbiorze X danym przez (6).

Zauważmy, że funkcja celu ϕ jest liniowa, a więc klasy C^1 i wypukła, zaś funkcje f_i są liniowe, a zatem C^1 i wklęsłe. Łatwo sprawdzić, że X spełnia warunek Slatera (można np. wziąć $x = (1, 1)$). Wszystkie założenia twierdzenia 3 są więc spełnione.

Dla $x = (x_1, x_2) \in \mathbb{R}^2$ i $y = (y_1, y_2, y_3) \in \mathbb{R}^3_+$, mamy

$$
\begin{aligned}
L(x, y) &= x_1 + 4x_2 - y_1(5x_1 + x_2 - 2) - y_2(3x_1 + 3x_2 - 3) \\
&= -y_3(x_1 + 3x_2 - 2),
\end{aligned}
$$

a zatem

$$
\begin{aligned}
\nabla_x L(x, y) &= (1 - 5y_1 - 3y_2 - y_3, 4 - y_1 - 3y_2 - 3y_3), \\
\nabla_y L(x, y) &= -f(y) \;=\; (2 - 5x_1 - x_2, 3 - 3x_1 - 3x_2, 2 - x_1 - 3x_2).
\end{aligned}
$$

Warunki na istnienie punktu siodłowego L w punkcie $(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^m_+$: $\nabla_x L(x, y) \geq 0$, $(x, \nabla_x L(x, y)) = 0$, $\nabla_y L(x, y) \leq 0$, $(y, \nabla_y L(x, y)) = 0$ (patrz twierdzenie 1), dają układ równań i nierówności

$$5y_1 + 3y_2 + y_3 \leq 1, \qquad y_1 + 3y_2 + 3y_3 \leq 4,$$
$$x_1(1 - 5y_1 - 3y_2 - y_3) + x_2(4 - y_1 - 3y_2 - 3y_3) = 0,$$
$$2 - 5x_1 - x_2 \leq 0, \quad 3 - 3x_1 - 3x_2 \leq 0, \quad 2 - x_1 - 3x_2 \leq 0,$$
$$y_1(2 - 5x_1 - x_2) + y_2(3 - 3x_1 - 3x_2) + y_3(2 - x_1 - 3x_2) = 0.$$

Ponieważ $x \geq 0$ i $y \geq 0$, równania w powyższym układzie można wzmocnić do

$$x_1(1 - 5y_1 - 3y_2 - y_3) = 0, \qquad x_2(4 - y_1 - 3y_2 - 3y_3) = 0,$$
$$y_1(2 - 5x_1 - x_2) = 0, \quad y_2(3 - 3x_1 - 3x_2) = 0, \quad y_3(2 - x_1 - 3x_2) = 0.$$

Przez żmudne rozpatrywanie wszystkich przypadków można znaleźć rozwiązanie tego układu: $x_1 = 2$, $x_2 = 0$, $y_1 = y_2 = 0$, $y_3 = 1$. Zatem z twierdzenia 3 wynika, że $x^* = (2, 0)$ jest rozwiązaniem naszego zadania optymalizacyjnego.