Metody optymalizacji. Wykład 5

Kierunki dopuszczalne, c.d.

Niech

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

Cały czas zakładamy, że $\phi, f_1, ..., f_m \in C^1$. Wykazaliśmy ostatnio

Lemat 1 (Lemat 3 z wykładu 4) *Niech X będzie dany przez (1) i niech $x \in X$ będzie lokalnym minimum funkcji ϕ na X. Załóżmy, że $s \in \mathbb{R}^n \setminus \{0\}$ i $\sigma \in \mathbb{R}$ spełniają układ nierówności*

$$(\nabla f_i(x), s) + \sigma \leq 0, \qquad i \in I(x), \tag{2}$$

$$-(\nabla \phi(x), s) + \sigma \leq 0. \tag{3}$$

Wtedy $\sigma \leq 0$.

Za pomocą tego lematu wykażemy następujący rezultat.

Twierdzenie 1 *Jeśli $x \in X$ jest lokalnym minimum funkcji ϕ na X i wektory $\nabla f_i(x)$, $i \in I(x)$, są liniowo niezależne, to istnieją liczby $u_i \geq 0$, $i \in I(x)$, takie, że*

$$\nabla \phi(x) = \sum_{i \in I(x)} u_i \nabla f_i(x). \tag{4}$$

Przed dowodem podamy **przykład:**

Niech $n = 2$, $m = 3$. Dla $x = (x_1, x_2) \in \mathbb{R}^2$, niech $f_1(x) = x_1$, $f_2(x) = x_2$, $f_3(x) = x_1 x_2 - 1$, a więc

$$X = \{x \in \mathbb{R}^2 : f_i(x) \geq 0,\ i = 1, 2, 3\} = \{(x_1, x_2) :\ x_1 \geq 0, x_2 \geq 0, x_1 x_2 \geq 1\}.$$

X jest zatem zbiorem punktów pierwszej ćwiartki $\mathbb{R}^2$ leżących na gałęzi hiperboli o równaniu $x_2 = 1/x_1$ lub nad nią. Nietrudno sprawdzić, że X jest wypukły, choć f_3 nie jest wklęsła.

Naszym zadaniem jest zminimalizowanie na X funkcji $\phi(x) = x_1 + 4x_2$. Minimalizujemy więc funkcję liniową na zbiorze wypukłym (jest to zadanie programowania wypukłego).

Szukamy minimów lokalnych ϕ. Jeśli w punkcie x jest minimum lokalne ϕ i $f_3(x) = x_1 x_2 - 1 = c \geq 0$, to x leży na krzywej

$$F_c = \{x : x_1 \geq 0, x_2 \geq 0, f_3(x) = c\}.$$

Na F_c mamy $x_1 x_2 = 1 + c$, a zatem $x_2 = (1+c)/x_1$. Stąd dla $x \in F_c$, $\phi(x) = x_1 + 4\frac{1+c}{x_1} =: \tilde{\phi}(x_1)$. Zatem warunkiem koniecznym na istnienie minimum lokalnego ϕ w x jest

$$\tilde{\phi}'(x_1) = 1 - \frac{4(1+c)}{x_1^2} = 0,$$

co daje $x_1 = 2\sqrt{1+c}$. Wstawiając to do wzoru na $\tilde{\phi}(x_1)$, dostajemy

$$\tilde{\phi}(x_1) = \phi(x_1, (1+c)/x_1) = 2\sqrt{1+c} + 4\frac{1+c}{2\sqrt{1+c}} = 4\sqrt{1+c}.$$

Ponieważ $X = \bigcup_{c \geq 0} F_c$, nietrudno zauważyć, że minimum lokalne ϕ na X może być osiągnięte tylko na F_0, w punkcie $x^* = (2, 1/2)$. Można sprawdzić, że x^* jest w istocie globalnym minimum ϕ na X.

Sprawdzimy, że w x^* zachodzi warunek (4). Mamy $I(x^*) = \{3\}$ (tylko trzecie ograniczenie jest w $(2, 1/2)$ aktywne) i $\nabla f_3(x_1, x_2) = (x_2, x_1)$, czyli $\nabla f_3(2, 1/2) = (1/2, 2)$. Z drugiej strony, $\nabla\phi(x_1, x_2) = (1, 4)$, a zatem $\nabla\phi(x^*) = 2\nabla f_3(x^*)$ i (4) zachodzi przy $u_3 = 2 \geq 0$.

Dowód twierdzenia 1. Na mocy lematu 1, jeśli przy $s \in \mathbb{R}^n \setminus \{0\}$ zachodzą (2)-(3), to $\sigma \leq 0$, co zapiszemy w postaci

$$(0, s) + 1 \cdot \sigma \leq 0. \tag{5}$$

(Zauważmy, że przy $s = 0$ wynikanie (5) z (2)-(3) jest oczywiste). Zastosujemy do układu równań (2)-(3), (5) twierdzenie Farkasa.

Niech $v \in \mathbb{R}^{n+1}$, $v = (0, ..., 0, 1)$, i niech B będzie macierzą o wymiarach $(|I(x)| + 1) \times (n+1)$ (gdzie przez $|A|$ oznaczamy liczebność zbioru A), którą zdefiniujemy poniżej. Oznaczmy wiersze B przez $b_1, ..., b_{k+1}$, gdzie $k = |I(x)|$, i niech $I(x) = \{i_1, ..., i_k\}$. Połóżmy $b_j = [\nabla f_{i_j}(x), 1]$, $j = 1, ..., k$, $b_{k+1} = [-\nabla\phi(x), 1]$. Nierówności (2)-(3) można zapisać w postaci

$$(s, \sigma) \in Z = \{(\tilde{s}, \tilde{\sigma}) \in \mathbb{R}^{n+1} : B(\tilde{s}, \tilde{\sigma})^T \leq 0\}.$$

Natomiast (5) oznacza, że $(v, (s, \sigma)) \leq 0$. Zatem z twierdzenia Farkasa i implikacji (2)-(3) $\Rightarrow$ (5) mamy

$$v = B^T u, \tag{6}$$

gdzie $B^T = [b_1^T, ..., b_{k+1}^T]$, $u = (u_{i_1}, ...u_{i_k}, u_0)$, $u_{i_j} \geq 0$, $j = 1, ..., k$, $u_0 \geq 0$. Rozpisując (6), dostajemy

$$0 = \sum_{i \in I(x)} u_i \nabla f_i(x) - u_0 \nabla\phi(x), \tag{7}$$

$$1 = \sum_{i \in I(x)} u_i + u_0. \tag{8}$$

Jeśli $u_0 \neq 0$, z (7) mamy tezę (wystarczy podzielić (7) przez u_0). W przeciwnym wypadku (7)-(8) implikują liniową zależność $\nabla f_i(x)$, $i \in I(x)$, co daje sprzeczność z założeniem. $\square$

Zauważmy na koniec, że jeśli w twierdzeniu 1 mamy $x \in int(X)$, to $I(x) = \emptyset$ i pustą sumę w (4) należy rozumieć jako zero. W tym przypadku twierdzenie 1 daje więc znany z analizy warunek konieczny na istnienie ekstremum lokalnego funkcji różniczkowalnej w zbiorze otwartym:

$$\nabla \phi(x) = 0. \tag{9}$$

W teorii optymalizacji (zwłaszcza liniowej) ekstrema znajduje się często na brzegu zbioru dopuszczalnego. Twierdzenie 1 podaje więc uogólnienie warunku (9) do przypadku, w którym x może się znajdować na ∂X.

Własności ekstremalne funkcji wypukłych, c.d. Twierdzenie Kuhna - Tuckera

W przypadku, gdy funkcje f_i, $i = 1, ..., m$, definiujące X, są wklęsłe, warunek liniowej niezależności $\nabla f_i(x)$, $i \in I(x)$, z twierdzenia 1 można zastąpić prostszym warunkiem regularności X. Podamy poniżej dwa takie warunki.

I. *Warunek regularności*: Dla dowolnego $i \in \{1, ..., m\}$, istnieje $x_i \in X$ taki, że $f_i(x_i) > 0$.

II. *Warunek regularności Slatera*: Istnieje $x \in X$ taki, że $f_i(x) > 0$ dla każdego $i \in \{1, ..., m\}$.

Oczywiście II $\Rightarrow$ I. Zauważmy że przy założeniu, że

$$f_i,\ i = 1, ..., m, \ \text{ są wklęsłe,} \tag{10}$$

które będziemy robić w następnych czterech twierdzeniach, warunki I i II są równoważne. Istotnie, załóżmy (10) i I. Niech $\alpha_i > 0$, $i = 1, ..., m$, będą takie, że $\sum_{i=1}^m \alpha_i = 1$ i niech $x = \sum_{i=1}^m \alpha_i x_i$. Na mocy (1) i (10) zbiór X jest wypukły, więc $x \in X$. Z nierówności Jensena, dla każdego $j \in \{1, ..., m\}$ mamy $f_j(x) \geq \sum_{i=1}^m \alpha_i f_j(x_i) > 0$, ponieważ $f_j(x_j) > 0$ na mocy I i $f_j(x_i) \geq 0$, $i = 1, .., m$, gdyż $x_i \in X$.

Twierdzenie 2 *Załóżmy, że funkcje f_i, $i = 1, ..., m$, są wklęsłe, zbiór X dany przez* (1) *jest regularny w sensie Slatera i $x \in X$ jest lokalnym minimum funkcji ϕ na X. Wtedy istnieją $u_i \geq 0$, $i \in I(x)$, takie, że zachodzi* (4).

Dowód. Jak w dowodzie twierdzenia 1 pokazujemy, że istnieją liczby $\tilde{u}_i \geq 0$, $i \in I(x)$, $\tilde{u}_0$ takie, że

$$0 = \sum_{i \in I(x)} \tilde{u}_i \nabla f_i(x) - \tilde{u}_0 \nabla \phi(x), \quad (11)$$

$$1 = \sum_{i \in I(x)} \tilde{u}_i + \tilde{u}_0. \quad (12)$$

Jeśli $\tilde{u}_0 \neq 0$, z (11) mamy tezę (wystarczy podzielić (11) przez u_0). Jeśli $\tilde{u}_0 = 0$, na mocy (12) istnieje $l \in I(x)$ takie, że $\tilde{u}_l > 0$. Z regularności X istnieje $z \in X$ taki, że $f_i(z) > 0$, $i = 1, ..., m$. Jeśli $z = x$, to $I(x) = \emptyset$, $x \in int(X)$ i (4) sprowadza się do $\nabla \phi(x) = 0$, znanego warunku koniecznego na istnienie ekstemum lokalnego. Zakładamy więc $x \neq z$ i niech $s = x - z$. Dla każdego $\alpha \in [0, 1]$, $i = 1, ..., m$, na mocy wklęsłości f_i mamy

$$\begin{aligned} f_i(x - \alpha s) &= f_i(x + \alpha(z - x)) = f_i(\alpha z + (1 - \alpha)x) \\ &\geq \alpha f_i(z) + (1 - \alpha) f_i(x) \geq \alpha \cdot 0 + (1 - \alpha) \cdot 0 = 0, \end{aligned}$$

a zatem $-s$ jest kierunkiem dopuszczalnym w x. Ponieważ f_l jest wklęsła (czyli $-f_l$ jest wypukła), na mocy twierdzenia 2 z poprzedniego wykładu mamy

$$(-\nabla f_l(x), s) = (\nabla f_l(x), z - x) \geq f_l(z) - f_l(x) = f_l(z) > 0, \quad (13)$$

gdzie ostatnia równość wynika z faktu, że $l \in I(x)$. Z (11) (przy $\tilde{u}_0 = 0$),

$$0 = \sum_{i \in I(x)} \tilde{u}_i (\nabla f_i(x), s). \quad (14)$$

Ponieważ kierunek $-s$ jest dopuszczalny w x, na mocy lematu 2 z poprzedniego wykładu, mamy $(\nabla f_i(x), s)) \leq 0$, $i \in I(x)$. Zatem $\tilde{u}_i(\nabla f_i(x), s) \leq 0$, $i \in I(x)$, a na mocy (13) i wyboru l, $\tilde{u}_l(\nabla f_l(x), s) < 0$, co daje sprzeczność z (14). $\square$

Następujące twierdzenie podaje interesującą charakteryzację "geometryczną" lokalnego minimum przy założeniach twierdzenia 2.

Twierdzenie 3 *Jeśli funkcje f_i, $i = 1, ..., m$, są wklęsłe, X jest regularny w sensie Slatera i $x \in X$ jest lokalnym minimum funkcji ϕ na X, to $x = p(x - \nabla \phi(x))$, gdzie dla $v \in \mathbb{R}^n$ $p(v)$ oznacza projekcję (rzut) v na X.*

Zauważmy, że operator rzutowania p jest dobrze określony, bo X jest domknięty i wypukły.

Dowód. Niech $y \in X$. Z wypukłości X, wszystkie punkty na odcinku $[x, y]$ należą do X, więc $-s = y - x$ jest kierunkiem dopuszczalnym w x (o ile $x \neq y$, tzn. $-s$ wyznacza jakiś kierunek w $\mathbb{R}^n$). Na mocy twierdzenia 2,

$$\begin{aligned}((x - \nabla\phi(x)) - x, y - x) &= (\nabla\phi(x), x - y) = (\nabla\phi(x), s) \\ &= \Big(\sum_{i \in I(x)} u_i \nabla f_i(x), s\Big) = \sum_{i \in I(x)} u_i (\nabla f_i(x), s)\end{aligned}$$

z $u_i \geq 0$, $i \in I(x)$. Ponieważ $-s$ jest dopuszczalny w x, z lematu 2 z poprzedniego wykładu mamy $(\nabla f_i(x), s) \leq 0$, $i \in I(x)$, zatem

$$((x - \nabla\phi(x)) - x, y - x) \leq 0, \qquad y \in X.$$

Warunek ten, na mocy geometrycznej charakteryzacji rzutu prostokątnego z drugiego wykładu, jest równoważny temu, że $x = p(x - \nabla\phi(x))$. □

Twierdzenia 1-3 podawały warunki konieczne dla optymalności. Następujące twierdzenie podaje warunki dostateczne.

Twierdzenie 4 *Warunkiem dostatecznym na to, by $x \in X$ był minimum globalnym w podstawowym zadaniu programowania wypukłego jest istnienie liczb $u_i \geq 0$, $i \in I(x)$, takich, że zachodzi* (4).

Dowód. Jak w dowodzie twierdzenia 3, dla dowolnego $y \in X$ wektor $-s = y - x$ jest kierunkiem dopuszczalnym w x i $(\nabla f_i(x), s) \leq 0$, $i \in I(x)$. Ponieważ ϕ jest wypukła, na mocy (4), nieujemności u_i i twierdzenia 2 z poprzedniego wykładu mamy

$$\phi(x) - \phi(y) \leq (\nabla\phi(x), x - y) = (\nabla\phi(x), s) = \sum_{i \in I(x)} u_i (\nabla f_i(x), s) \leq 0,$$

a zatem $\phi(x) \leq \phi(y)$ dla każdego $y \in X$. □

Łącząc twierdzenia 2 i 4, otrzymujemy twierdzenie Kuhna - Tuckera (w przypadku różniczkowalnych ϕ, f_i, ale tylko ten przypadek będziemy omawiać), będące głównym wynikiem tej części naszego kursu.

Twierdzenie 5 (Kuhn - Tucker) *Jeśli funkcje f_i, $i = 1, ..., m$, są wklęsłe, zbiór X dany przez* (1) *jest regularny w sensie Slatera i $\phi : X \to \mathbb{R}$ jest wypukła, to dla optymalności $x \in X$ potrzeba i wystarcza, aby istniały liczby $u_i \geq 0$, $i = 1, ..., m$, takie, że*

$$\nabla\phi(x) = \sum_{i=1}^{m} u_i \nabla f_i(x), \tag{15}$$

$$0 = \sum_{i=1}^{m} u_i f_i(x). \tag{16}$$

Dowód. Jeśli x jest optymalny, to jest też minimum lokalnym ϕ. Z twierdzenia 2, istnieją liczby $u_i \geq 0$, $i \in I(x)$, takie, że zachodzi (4). Połóżmy $u_i = 0$, $i \notin I(x)$. Wtedy z (4) mamy (15). Co więcej,

$$\sum_{i=1}^{m} u_i f_i(x) = \Big(\sum_{i \in I(x)} + \sum_{i \notin I(x)} \Big) u_i f_i(x) = 0 + 0 = 0,$$

gdyż $f_i(x) = 0$ dla $i \in I(x)$, zaś $u_i = 0$ dla $i \notin I(x)$. Zatem (16) również zachodzi.

Odwrotnie, jeśli (15)-(16) zachodzą, to ponieważ $f_i(x) > 0$ dla $i \notin I(x)$ i $f_i(x) \geq 0$, $u_i \geq 0$ dla każdego i, z (16) wynika, że $u_i = 0$ dla $i \notin I(x)$. Zatem (15) implikuje (4). Na mocy twierdzenia 4, x jest więc optymalny. $\square$