

Wykład 3

3. Optymalizacja z ograniczeniami

Sformułujemy teraz warunki konieczne dla istnienia rozwiązań zagadnienia optymalizacyjnego:

$$\begin{aligned} \bullet \quad & \min f(x) \\ & g_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (\text{ograniczenia LE}) \\ (3.1) \quad & h_j(x) = 0, \quad j = 1, 2, \dots, p, \quad p < n \quad (\text{ograniczenia EQ}). \end{aligned}$$

Założenie. $f, g_i, h_j: U \rightarrow \mathbb{R}^n$ są funkcjami klasy C^1 , zbiór U jest otwarty.

Zbiór $D \subset U$ wektorów, spełniających ograniczenia typu LE i EQ nazywać będziemy **zbiorem dopuszczalnym**. Oznaczmy przez: D_{LE} zbiór wektorów, spełniających wszystkie ograniczenia LE, D_{EQ} zbiór wektorów, spełniających wszystkie ograniczenia EQ. Oczywiście, $D = D_{LE} \cap D_{EQ}$.

Szukamy warunków koniecznych jakie musi spełniać punkt minimum globalnego x^* w zbiorze dopuszczalnym D .

Redukcja 3.1. Wystarczy ograniczyć się do przypadku, gdy wszystkie ograniczenia są typu LE. Każde ograniczenie typu EQ zastąpimy dwoma ograniczeniami: ograniczenie $h_j(x) = 0$ zastąpimy przez $g_{m+2j-1}(x) \leq 0$, $g_{m+2j}(x) \leq 0$ gdzie $g_{m+2j-1}(x) = h_j(x)$ i $g_{m+2j}(x) = -h_j(x)$ dla $j = 1, 2, \dots, p$. Mamy więc $M = m + 2p$ ograniczeń typu LE

Definicja 3.1. Ograniczenie $u(x)$ jest **aktywne w punkcie x^*** gdy zachodzi równość $u(x^*) = 0$.

Oznaczmy przez $AKT = \{i : g_i(x^*) = 0\}$ zbiór ograniczeń aktywnych. Wszystkie ograniczenia EQ są z definicji aktywne więc indeksy o numerach $i = m + 1, \dots, m + 2p$ należą do zbioru AKT. Jednak zbiór AKT może być pusty.

Redukcja 3.2. Wystarczy ograniczyć się do przypadku, gdy wszystkie ograniczenia są aktywne. Niech

$$D_{AKT} = \{x \in U : g_i(x) \leq 0, i \in AKT\}, D_0 = \{x \in U : g_i(x) < 0, i \notin AKT\}$$

Zbiór D_0 jest otwarty (wynika to z ciągłości g_i). Tak więc zagadnienie

$$\begin{aligned} \bullet \quad & \min f(x) \\ & g_i(x) \leq 0, \quad i = 1, 2, \dots, M, x \in U \end{aligned}$$

jest równoważne zagadnieniu

$$\begin{aligned} \bullet \quad & \min f(x) \\ & g_i(x) \leq 0, \quad i \in AKT, x \in U_0 = U \cap D_0 \end{aligned}$$

Warunki można tak przenieumerować aby pierwszych M_0 było aktywnych

Niech $C_i = \{x : x^T \nabla g_i(x^*) \leq (x^*)^T \nabla g_i(x^*)\}$ będzie półprzestrzenią wyznaczoną przez styczną do g_i w punkcie $x^* \in C_i$. Jak łatwo sprawdzić, C_i jest zbiorem wypukłym, zaś $C = \bigcap_{i \in \text{AKT}} C_i$ jako część wspólna wypukłych półpłaszczyzn zbiorem wypukłym, zawierającym x^* .

Niech $C_0 = \{h : h^T \nabla g_i(x^*) \leq 0, i \in \text{AKT}\}$. Wtedy $\forall_{\mu \geq 0} \mu C_0 \subset C_0$, $C = x^* + C_0$ czyli C jest stożkiem o wierzchołku x^* i kierunkach w zbiorze C_0 .

Definicja 3.2. Punkt x^* jest **regularny** gdy dla każdego $h \in C_0$ istnieje otoczenie $U(x^*)$ oraz gładka, klasy C^2 krzywa $\varphi(t)$ taka, że:

1. $\varphi(0) = x^*$, 2. $\varphi'(0) = \mu h$ dla pewnego $\mu > 0$, 3. $\exists_{t_0} \forall_{0 < t < t_0} \varphi(t) \subset U(x^*) \cap D$

Lemat 3.1. Jeżeli punkt x^* jest regularny, $h \in C_0$ to $h^T \nabla f(x^*) \geq 0$

Dowód (niewprost)

$\exists_{h_1} h_1 \in C_0$ takie, że $h_1^T \nabla f(x^*) < 0$. Funkcja $h_1^T \nabla f(x)$ jako funkcja zmiennej x jest ciągła, więc istnieje otoczenie $U(x^*)$ takie, że w tym otoczeniu $h_1^T \nabla f(x) < 0$. Z regularności punktu x^* wynika, że istnieje otoczenie $U_1(x^*) \subset U(x^*)$ i gładka, klasy C^2 krzywa $\varphi(t)$ taka, że:

$\varphi(0) = x^*$, $\varphi'(0) = \mu h_1$ dla pewnego $\mu > 0$, $\exists_{t_0} \forall_{0 \leq t < t_0} \varphi(t) \subset U_1(x^*) \cap D$.

Rozwijając φ w pobliżu 0 w szereg Taylora mamy dla $0 \leq t < t_1 < t_0$

$$\varphi(t) = x^* + \mu t h_1 + \mu t \gamma(t) \rightarrow 0 \text{ gdy } t \rightarrow 0+.$$

Ponieważ $\varphi(t) \subset U_1(x^*) \cap D$ więc dla $0 \leq t < t_1$ $h_1^T \nabla f(\varphi(t)) < 0$

Funkcja $y^T \nabla f(\varphi(t))$ jako funkcja zmiennej y jest ciągła, więc istnieje otoczenie $U_t(h_1)$ takie, że w tym otoczeniu $y^T \nabla f(\varphi(t)) < 0$.

Kładąc $U_0(h_1) = \bigcap_{0 \leq t < t_1} U_t(h_1)$ mamy dla każdego $y \in U_0(h_1)$ i dla każdego $0 \leq t < t_1$,

$$0 > y^T \nabla f(\varphi(t)) = y^T \nabla f(x^* + \mu t (h_1 + \gamma(t))).$$

Łatwo zauważyć, że dla dostatecznie małych $\theta > 0$ dla każdego $y \in U_0(h_1)$ i dla każdego $0 \leq t < t_1$ zachodzi nierówność $y^T \nabla f(x^* + \theta \mu t (h_1 + \gamma(t))) < 0$

Wektor $h_1 + \gamma(t)$ dla dostatecznie małych t należy do $U_0(h_1)$ więc zachodzi dla dostatecznie małych $t > 0$, $(h_1 + \gamma(t))^T \nabla f(x^* + \theta \mu t (h_1 + \gamma(t))) < 0$ a więc dla $h = \mu t (h_1 + \gamma(t))$ mamy $x^* + h \in D$, $h^T \nabla f(x^* + \theta h) < 0$. Wzór ten zachodzi dla dostatecznie małych $t > 0$.

Wtedy, z rozwinięcia Taylora funkcji f mamy $f(x^* + h) = f(x^*) + h^T \nabla f(x^* + \theta h) < f(x^*)$ co stanowi sprzeczność z założeniem, że w x^* jest minimum globalne ■

Twierdzenie 3.1. (Warunek konieczny istnienia minimum, warunki Kuhna-Tuckera, warunki Karusha-Kuhna-Tuckera)

Niech w regularnym punkcie $x^* \in D$ funkcja celu osiąga w tym punkcie minimum w dopuszczalnym obszarze D Wtedy:

$$\sum_{i \in \text{Akt}} u_i \nabla g_i(x^*) + \sum_{j=1}^p v_j \nabla h_j(x^*) = -\nabla f(x^*) \quad (3.2)$$

dla pewnych $u_i \geq 0$ oraz v_j .

Definicja 3.3. Punkty regularne, spełniające warunki (3.2) twierdzenia 3.1 nazywać będziemy **punktami Kuhna-Tuckera (punktami KT)**

Dowód

W dowodzie wykorzystamy lemat Farkasa.

Lemat Farkasa (1902). Niech A będzie macierzą o k wierszach i n kolumnach, b i h n -wymiarowymi wektorami, w wektorze k -wymiarowym. Równoważne są następujące warunki:

$$\begin{aligned} &\text{dla każdego } h \text{ zachodzi implikacja } Ah \geq 0 \Rightarrow b^T h \geq 0, \\ &\exists u \geq 0 \ A^T u = b. \end{aligned}$$

Dla wektora q warunek $q \geq 0$ oznacza, że wszystkie jego współrzędne są nieujemne.

Podstawiając w lemacie Farkasa

$$\begin{aligned} b &= \nabla f(x^*), \\ A &= -(\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_{M_0}(x^*))^T, \end{aligned}$$

uzyskamy z Lematu 3.1, że dla każdego h zachodzi implikacja $Ah \geq 0 \Rightarrow b^T h \geq 0$.

Z Lematu Farkasa mamy, że dla wektora u o nieujemnych składnikach

$$\sum_{i \in \text{Akt}} u_i \nabla g_i(x^*) = -\nabla f(x^*)$$

Dla j -tego ograniczenia typu równości $h_j(x) = 0$,

$$u_{m+2j-1} \nabla h_j(x^*) + u_{m+2j} (-\nabla h_j(x^*)) = (u_{m+2j-1} - u_{m+2j}) \nabla h_j(x^*) = v_j \nabla h_j(x^*),$$

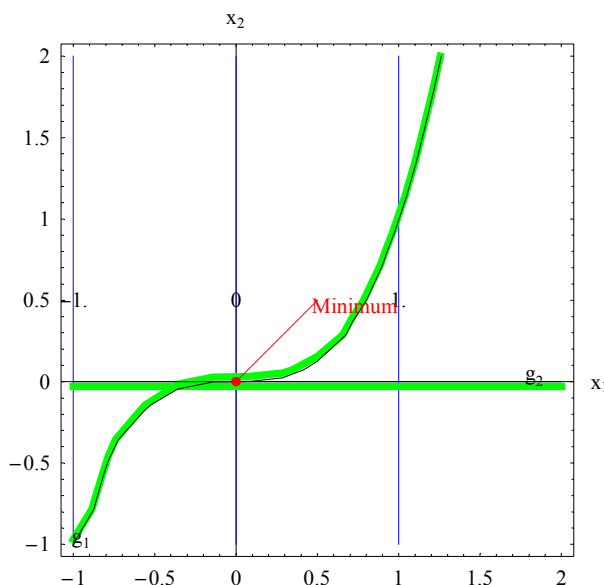
$$\sum_{i \in \text{Akt}} u_i \nabla g_i(x^*) + \sum_{j=1}^p v_j \nabla h_j(x^*) = -\nabla f(x^*)$$

dla pewnych $u_i \geq 0$ oraz v_j .

Warunek regularności jest istotny.

■ Przykład 3.1

Obliczmy minimum globalne funkcji $f(x) = x_1$ z warunkami $x_2 \leq x_1^3$, $x_2 \geq 0$. Rozwiązując to zadanie geometrycznie widzimy, że minimum globalne jest w $x^* = (0, 0)^T$



$$\nabla f \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad g_1 \rightarrow -x_1^3 + x_2 \leq 0 \quad g_2 \rightarrow -x_2 \leq 0,$$

$$\nabla g_1 \rightarrow \begin{pmatrix} -3x_1^2 \\ 1 \end{pmatrix} \quad \nabla g_2 \rightarrow \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_1(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \nabla g_2(x^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Zbiór $C_0 = \{h = (h_1, h_2) : h^T \nabla g_1(x^*) \leq 0, h^T \nabla g_2(x^*) \leq 0\} = \{h = (h_1, 0), h_1 \in R\}$. Nie istnieje krzywa φ leżąca w zbiorze dopuszczalnym dla $h = (-1, 0)$, gdyż wtedy $\varphi_1(t) = -\mu t + t\gamma_1(t) \geq 0$ dla dostatecznie małych $t > 0$. Wtedy $-\mu + \gamma_1(t) \geq 0$ i przechodząc z t do 0 otrzymamy $-\mu \geq 0$ co jest sprzeczne z założeniem, że $\mu > 0$. ■

Sprawdzenie, czy punkt x^* jest regularny jest trudne. Prawdziwe jest

Twierdzenie 3.2. Punkt $x^* \in D$ jest regularny jeżeli wektory:

$$\begin{aligned} \nabla g_i(x^*), \quad i \in Akt, \\ \nabla h_j(x^*), \quad j = 1, 2, \dots, p \end{aligned}$$

są liniowo niezależne.

Z tego wynika, że łączna liczba ograniczeń aktywnych i ograniczeń typu EQ w punkcie regularnym nie może przekraczać n .

Zbadanie, czy układ wektorów jest liniowo niezależny jest ważne przy ustaleniu czy analizowany punkt jest regularny.

Twierdzenie 3.3. (warunek dostateczny na to, by punkt był regularny). Dany jest układ funkcji n zmiennych $x_1, x_2, \dots, x_n$, będących ograniczeniami typu EQ (funkcji $h_1, h_2, \dots, h_p$) i ograniczeniami aktywnymi typu LE (funkcji $g_{i_1}, g_{i_2}, \dots, g_{i_k}$). Punkt x^* jest regularny jeśli macierz $(\nabla g_{i_1}(x^*) \nabla g_{i_2}(x^*) \dots \nabla g_{i_k}(x^*) \nabla h_1(x^*) \nabla h_2(x^*) \dots \nabla h_p(x^*))$ jest rzędu $k+p$.

■ Przykład 3.2 (Minimum w punkcie nieregularnym)

Znajdźmy rozwiązanie zadania optymalizacyjnego:

$$\begin{aligned} f(x, y) = x = \text{MIN}!, \\ x^2 \leq 0. \end{aligned}$$

Gradienty, związane z tym zadaniem mają postać:

$$\begin{aligned} \nabla f(x) = 1, \\ \nabla g(x) = 2x. \end{aligned}$$

Ograniczenie $g(x)$ jest aktywne, gdyż jedyny punkt, spełniający to ograniczenie to punkt 0. Tak więc jest to również rozwiązanie naszego zagadnienia.

Jednocześnie jest to punkt nieregularny, gdyż $\nabla g(0)=0$ a to nie jest wektor liniowo niezależny.

Warunki Kuhna Tuckera wymagają dodatkowych założeń, gdyż w naszym przykładzie, gdy g jest ograniczeniem aktywnym, musiałyby zachodzić:

$$\begin{aligned} u(2x) = -1, \\ x^2 = 0, \\ u \geq 0. \end{aligned}$$

co stanowi sprzeczność. ■

■ Przykład 3.3 (Warunki Kuhna-Tuckera)

Znajdźmy punkty KT dla zagadnienia:

$$\begin{aligned} f(x, y) &= x^2 + y^2 = \text{MIN}!, \\ x + y &\geq 5, \\ x + 2y &= 3. \end{aligned}$$

Sprowadzając do postaci standardowej

$$\begin{aligned} f(x, y) &= x^2 + y^2 = \text{MIN}!, \\ g(x, y) &= 5 - x - y \leq 0, \\ h(x, y) &= x + 2y - 3 = 0. \\ \nabla f(x, y) &= (2x, 2y)^T, \\ \nabla g(x, y) &= (-1, -1)^T, \\ \nabla h(x, y) &= (1, 2)^T \end{aligned}$$

Łatwo sprawdzić, że wszystkie punkty, należące do zbioru dopuszczalnego są regularne. Niech (x^*, y^*) będzie rozwiązaniem naszego zadania. Rozważmy dwa przypadki:

I. $g(x)$ jest aktywne w (x^*, y^*) .

Wtedy muszą zachodzić warunki Kuhna Tuckera:

$$\begin{aligned} u(-1, -1) + v(1, 2) &= -(2x, 2y) \\ 5 - x - y &= 0, \\ x + 2y - 3 &= 0, \\ u &\geq 0. \end{aligned}$$

Układ ten ma rozwiązanie:

$$\begin{aligned} x^* &= 7, \quad y^* = -2, \quad u = 32, \quad v = 18, \\ f(x^*, y^*) &= 53. \end{aligned}$$

II. $g(x)$ nie jest aktywne w (x^*, y^*) .

Warunki Kuhna Tuckera w tym przypadku przybiorą postać:

$$\begin{aligned} v(1, 2) &= -(2x, 2y), \\ 5 - x - y &< 0, \\ x + 2y - 3 &= 0. \end{aligned}$$

Układ ten nie ma rozwiązania, gdyż pierwiastki układu, składającego się z pierwszego i trzeciego równania $x = \frac{3}{5}$ i $y = \frac{6}{5}$ nie spełniają nierówności.

Tak więc rozwiązanie $x^* = 7$ i $y^* = -2$, jako jedyne, jest odpowiedzią w naszym zadaniu. ■

Warunki Kuhna Tuckera można uprościć, wprowadzając dodatkowe zmienne s_i , $i = 1, 2, \dots, m$, które przekształcą wszystkie ograniczenia LE na ograniczenia EQ. Wtedy też wszystkie ograniczenia są aktywne. Nasze zadanie będzie równoważne :

$$\begin{aligned} \bullet \quad &\min f(x) \\ &g_i^*(x, s) = g_i(x) + s_i^2 = 0, \quad i = 1, 2, \dots, m \\ &h_j(x) = 0, \quad j = 1, 2, \dots, p, \quad p < n \end{aligned}$$

Zauważmy, że

$$\begin{aligned}
\nabla f(x, s) &= (\nabla f(x), 0), \\
\nabla g_i^*(x, s) &= (\nabla g_i(x), \\
&\quad 2s_i), \\
\nabla h_j(x, s) &= (\nabla h_j(x), 0)
\end{aligned} \tag{3.2}$$

W powyższych równaniach $0, e_i$ są wektorami m -wymiarowymi: pierwszy składa się z samych 0, drugi ma na i -tym miejscu 1 a na pozostałych miejscach 0.

Zapisując w nowej sytuacji warunki Kuhna Tuckera, otrzymamy:

$$\sum_{i=1}^m u_i \nabla g_i^*(x^*, s^*) + \sum_{j=1}^p v_j \nabla h_j(x^*, s^*) = -\nabla f(x^*, s^*)$$

co po uwzględnieniu (3.2) daje układ (**zmodyfikowane warunki Kuhna-Tuckera (warunki K-T)**)

$$\begin{aligned}
&\sum_{i=1}^m u_i \nabla g_i(x^*) + \sum_{j=1}^p v_j \nabla h_j(x^*) = \\
&\quad -\nabla f(x^*), \\
&\quad u_i s_i^* = 0, \quad u_i \geq 0, \\
&\quad g_i(x^*) + (s^*)_i^2 = 0, \quad i = 1, 2, \dots, m, \\
&\quad h_j(x^*) = 0 \quad j = 1, 2, \dots, p
\end{aligned} \tag{3.3}$$

Gdy $s_i = 0$ to ograniczenie g_i jest aktywne, gdy $s_i \neq 0$ to $u_i = 0$ i ograniczenie g_i nie bierze udziału w wyznaczeniu x^* . Z poprzednich rozważań wynika, że liczba ograniczeń aktywnych, więc takich, dla których $s_i = 0$ wraz z liczbą ograniczeń EQ nie może przekraczać n .

■ Przykład 3.4 (Rozwiązanie przykładu 3.3 zmodyfikowaną metodą K-T)

Zapiszmy od razu warunki (3.3):

$$\begin{aligned}
u(-1, 1) + v(1, 2) &= -(2x, 2y), \\
us &= 0, \quad u \geq 0, \\
-x - y + 5 + s^2 &= 0, \\
x + 2y - 3 &= 0,
\end{aligned}$$

Wyznamy x i y z ostatnich dwóch równań:

$$x = 7 + 2s^2, \quad y = -2 - s^2$$

Podstawiając do pierwszego równania otrzymamy, że $3u = 32 + 10s^2 > 0$, co od razu daje rozwiązanie $s = 0, x = 7, y = -2$. ■