Metody optymalizacji. Wykład 3

Punkty ekstremalne zbiorów wypukłych

Rozpoczniemy od omówienia punktów ekstremalnych (wierzchołkowych) zbiorów wypukłych. Pojęcie to okaże się kluczowe w optymalizacji liniowej.

Definicja 1 *Punkt $x_0 \in A$ nazywamy punktem ekstremalnym (wierzchołkowym) zbioru $A \subseteq \mathbb{R}^n$, jeśli nie istnieją punkty $x, y \in A$, $x \neq y$, takie, że dla pewnego $\alpha \in (0,1)$, mamy $x_0 = \alpha x + (1-\alpha)y$.*

Innymi słowy, punkt wierzchołkowy A nie jest punktem wewnętrznym żadnego odcinka o końcach w A. Oczywiście punkt wewnętrzny A nie może być punktem wierzchołkowym tego zbioru. Zatem punktów ekstremalnych A trzeba szukać na ∂A.

Przykłady.

1. Koło na płaszczyźnie. Łatwo sprawdzić, że każdy punkt na okręgu - brzegu koła - jest punktem ekstremalnym tego koła.

2. Trójkąt (z wnętrzem lub bez). Nietrudno sprawdzić, że jedynymi jego punktami ekstremalnymi są wierzchołki tego trójkąta.

Twierdzenie 1 *Niech $A \subseteq \mathbb{R}^n$ będzie zbiorem domkniętym, wypukłym i ograniczonym. Wtedy każdy punkt $x_0 \in A$ jest kombinacją wypukłą skończonej liczby punktów ekstremalnych A.*

Dowód. Indukcja względem n-wymiaru przestrzeni.

$n = 1$: w tym przypadku A jest odcinkiem domkniętym (dlaczego?), więc każdy punkt A jest kombinacją wypukłą jego końców - punktów ekstremalnych A.

Załóżmy, że teza twierdzenia jest prawdziwa dla $n = k-1$. Wykażemy ją dla $n = k$. Dowód przeprowadzimy w dwóch przypadkach.

I. $x_0 \in \partial A$.

Niech $\Pi = \{x \in \mathbb{R}^n : (c, x) = \lambda\}$, $c \neq 0$, będzie hiperpłaszczyzną podpierającą A w punkcie x_0, tzn. taką, że $x_0 \in \Pi$ i dla każdego $x \in A$ mamy $(c, x) \leq \lambda$. Wtedy $X = A \cap \Pi$ jest domkniętym, wypukłym i ograniczonym podzbiorem Π (hiperpłaszczyzny $k-1$-wymiarowej). Na mocy założenia indukcyjnego istnieją punkty ekstremalne $x_1, ..., x_N$ zbioru X takie, że $x_0 = \sum_{i=1}^N \alpha_i x_i$, $\alpha_i \geq 0$, $i = 1, ..., N$, $\sum_{i=1}^N \alpha_i = 1$. Pozostaje wykazać, że x_i są punktami ekstremalnymi A.

Niech $i \in \{1, ..., N\}$. Załóżmy, że istnieją $x', x'' \in A$, $x' \neq x''$, i $\alpha \in (0,1)$ takie, że $x_i = \alpha x' + (1-\alpha)x''$. Ale $(x_i, c) = \lambda$, bo $x_i \in \Pi$, a jednocześnie

$(c, x') \le \lambda$ i $(c, x'') \le \lambda$. Jeśli choć jedna z ostatnich nierówności jest ostra,

$$\lambda = (x_i, c) = (\alpha x' + (1-\alpha)x'', c) = \alpha(x', c) + (1-\alpha)(x'', c) < \alpha\lambda + (1-\alpha)\lambda = \lambda,$$

co daje sprzeczność. Zatem $x', x'' \in X$. To z kolei daje sprzeczność z faktem, że x_i jest punktem ekstremalnym zbioru X.

II. $x_0 \in int(A)$.

Niech l będzie dowolną linią prostą przechodzącą przez x_0. Wtedy $l \cap A$ jest odcinkiem domkniętym o końcach $\bar{x}, \bar{\bar{x}} \in \partial A$ (porównaj przypadek $n = 1$). W szczególności, dla pewnego $\alpha \in (0, 1)$, mamy $x_0 = \alpha\bar{x} + (1-\alpha)\bar{\bar{x}}$. Na mocy przypadku I, $\bar{x}, \bar{\bar{x}}$ są kombinacjami wypukłymi punktów ekstremalnych A, tzn.

$$\bar{x} = \sum_{i=1}^{N_1} \beta_i y_i, \qquad \beta_i \ge 0, \quad i = 1, ..., N_1, \quad \sum_{i=1}^{N_1} \beta_i = 1,$$
$$\bar{\bar{x}} = \sum_{i=1}^{N_2} \gamma_i z_i, \qquad \gamma_i \ge 0, \quad i = 1, ..., N_2, \quad \sum_{i=1}^{N_2} \gamma_i = 1,$$

gdzie y_i, z_i są punktami ekstremalnymi A. Zatem

$$x_0 = \sum_{i=1}^{N_1} \alpha\beta_i y_i + \sum_{i=1}^{N_2} (1-\alpha)\gamma_i z_i,$$

gdzie $\alpha\beta_i \ge 0$, $(1-\alpha)\gamma_i \ge 0$ i

$$\sum_{i=1}^{N_1} \alpha\beta_i + \sum_{i=1}^{N_2} (1-\alpha)\gamma_i = \alpha + (1-\alpha) = 1.$$

Wykazaliśmy w ten sposób, że x_0 jest kombinacją wypukłą punktów ekstremalnych y_i, z_i zbioru A. □

Twierdzenie Farkasa

Definicja 2 *Zbiór $K \subseteq \mathbb{R}^n$ nazywamy stożkiem, jeśli dla dowolnych $x \in K$ i $\lambda > 0$, mamy $\lambda x \in K$.*

Przykłady:

1. $\mathbb{R}^n$.

2. $\mathbb{R}^n_+ = [0, \infty)^n$.

3. Jeśli A jest macierzą o wymiarach $m \times n$, to zbiory

$$\{x \in \mathbb{R}^n : Ax \leq 0\}, \qquad Y = \{y \in \mathbb{R}^m : y = Ax, x \in \mathbb{R}^n_+\}$$

są stożkami. Można również pokazać, że Y jest domknięty (ćwiczenia).

Następujące twierdzenie, zwane również lematem Farkasa, będzie potrzebne w naszych dalszych rozważaniach. Jest ono również przydatne w innych dziedzinach (np. matematyce finansowej).

Twierdzenie 2 (Farkasa) *Niech B będzie macierzą o wymiarach $m \times n$ i niech $v \in \mathbb{R}^n$. Niech $Z = \{x \in \mathbb{R}^n : Bx \leq 0\}$. Następujące warunki są równoważne:*

(i) $(v, x) \leq 0$ dla każdego $x \in Z$,

(ii) istnieje $u \in \mathbb{R}^m$, $u \geq 0$, taki, że $v = B^T u$.

Dowód. (ii) $\Rightarrow$ (i): Jeśli $v = B^T u$ dla pewnego $u \in \mathbb{R}^m$, $u \geq 0$, to dla każdego $x \in \mathbb{R}^n$ takiego, że $Bx \leq 0$, mamy

$$(v, x) = (B^T u, x) = (u, Bx) \leq 0.$$

(i) $\Rightarrow$ (ii): Niech $Y = \{y \in \mathbb{R}^n : y = B^T u, \ u \in \mathbb{R}^m_+\}$. Jak wiemy, Y jest domkniętym stożkiem. Jeśli $v \in Y$, mamy tezę. Łatwo sprawdzić, że Y jest wypukły (ogólniej, liniowy obraz zbioru wypukłego jest wypukły - ćwiczenie). Jeśli więc $v \notin Y$, to z twierdzenia o oddzielaniu punktu od zbioru wypukłego i domkniętego hiperpłaszczyzną, istnieje $c \in \mathbb{R}^n \setminus \{0\}$ takie, że

$$(c, y) < (c, v), \qquad y \in Y. \tag{1}$$

Ponieważ dla $\lambda > 0$ i $y \in Y$ mamy $\lambda y \in Y$, $\lambda(c, y) = (c, \lambda y) < (c, v)$ dla każdego $\lambda > 0$, a zatem $(c, y) \leq 0$. Stąd

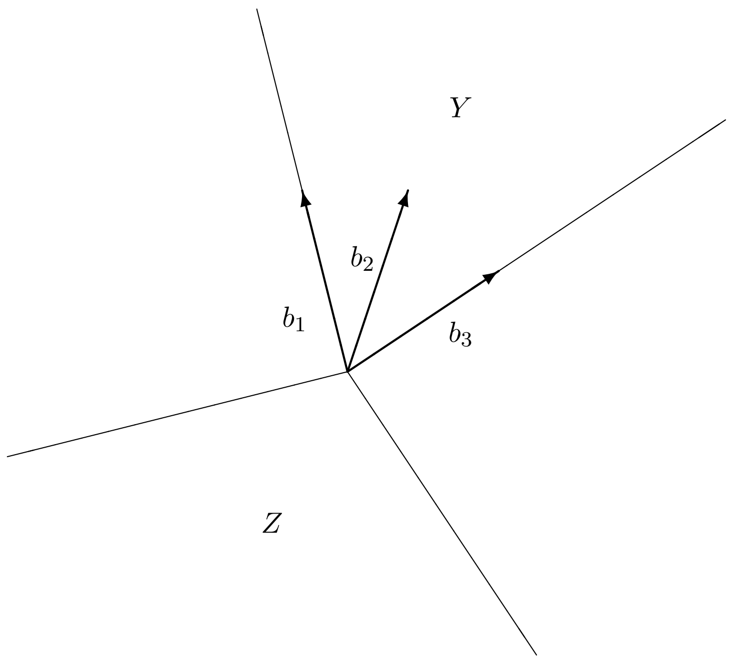
$$(c, y) = (c, B^T u) = (Bc, u) \leq 0, \qquad u \in \mathbb{R}^m_+.$$

Stąd $Bc \leq 0$ (istotnie, gdyby $(Bc)_i > 0$ dla pewnego i, kładąc $u = (u_j)$, $u_j = \delta_{ij}$, $j = 1, ..., n$, gdzie $\delta_{ij} = 1$ dla $i = j$ i 0 w przeciwnym razie, mielibyśmy $u \geq 0$, $(Bc, u) = (Bc)_i > 0$). W szczególności, $c \in Z$.

Ponieważ $0 = B^T 0 \in Y$, z (1) mamy $(c, v) > 0$. Ale na mocy (i), $(c, v) = (v, c) \leq 0$. Otrzymana sprzeczność kończy dowód.

Interpretacja geometryczna. Niech

$$B = \begin{bmatrix} b_1^T \\ ... \\ b_m^T \end{bmatrix},$$



gdzie b_1^T,..., b_m^T są wierszami B (tzn. wektory

$$b_i = \begin{bmatrix} b_{i1} \\ ... \\ b_{in} \end{bmatrix} \in \mathbb{R}^n, \qquad i = 1, ..., m,$$

wyobrażamy sobie jako kolumny). Wtedy

$$Z = \{x \in \mathbb{R}^n : Bx \leq 0\} = \{x \in \mathbb{R}^n : b_i^T x \leq 0,\ i = 1, ..., m\}$$

jest stożkiem tych $x \in \mathbb{R}^n$, które tworzą z każdym wektorem b_i, $i = 1, ..., m$, kąt nieostry. Natomiast

$$\begin{aligned} Y &= \{y \in \mathbb{R}^n : y = B^T u, u \in \mathbb{R}^m_+\} \\ &= \left\{ y \in \mathbb{R}^n : y = [b_1, ..., b_m] \begin{bmatrix} u_1 \\ ... \\ u_m \end{bmatrix},\ u_1 \geq 0, ...u_m \geq 0 \right\} \\ &= \{y \in \mathbb{R}^n : y = u_1 b_1 + ... + u_m b_m,\ u_1 \geq 0, ...u_m \geq 0\} \end{aligned}$$

jest zbiorem kombinacji liniowych $b_1, ..., b_m$ o nieujemnych współczynnikach.

Lemat Farkasa mówi więc, że wektor $v \in \mathbb{R}^n$ tworzy kąt nieostry z dowolnym $x \in Z$ wtedy i tylko wtedy, gdy $v \in Y$.

Wniosek 1 *Dla dowolnej macierzy B o wymiarach $m \times n$ i dowolnego $v \in \mathbb{R}^n$, zachodzi dokładnie jedna z następujących tez:*
(a) istnieje rozwiązanie $x \in \mathbb{R}^n$ układu $Bx \geq 0$, $(v, x) < 0$,
(b) układ $v = B^T u$ ma rozwiązanie $u \in \mathbb{R}^m_+$.

Dowód. Jeśli zachodzi (b), z twierdzenia Farkasa mamy $(v, x) \leq 0$ dla każdego $x \in \mathbb{R}^n$ takiego, że $Bx \leq 0$. Zatem $(v, x) \geq 0$ dla każdego $x \in \mathbb{R}^n$ takiego, że $Bx \geq 0$, czyli (a) nie zachodzi.

Jeśli (b) nie zachodzi, z twierdzenia Farkasa mamy $(v, x) > 0$ dla pewnego $x \in Z$. Kładąc $x_0 = -x$, mamy $Bx_0 = -Bx \geq 0$, $(v, x_0) = -(v, x) < 0$, a zatem zachodzi (a).

Funkcje wypukłe

Definicja 3 *Funkcję $\phi : X \to \mathbb{R}^n$, gdzie $\emptyset \neq X \subseteq \mathbb{R}^n$ jest wypukły, nazywamy wypukłą, jeśli dla dowolnych $x, y \in X$ i $\lambda \in [0, 1]$, mamy*

$$\phi(\lambda x + (1 - \lambda)y) \leq \lambda\phi(x) + (1 - \lambda)\phi(y). \tag{2}$$

Jeśli dla dowolnych $x, y \in X$, $x \neq y$, i $\lambda \in (0, 1)$ nierówność (2) jest ostra, to ϕ nazywamy ściśle wypukłą na X.

Funkcję $\phi : X \to \mathbb{R}^n$ nazywamy wklęsłą (ściśle wklęsłą) na X, jeśli funkcja $-\phi$ jest na X wypukła (ściśle wypukła).

Poznawanie własności funkcji wypukłych rozpoczniemy od następującego prostego, ale użytecznego i ważnego, twierdzenia.

Twierdzenie 3 (Nierówność Jensena) *Niech $\emptyset \neq X \subseteq \mathbb{R}^n$ będzie zbiorem wypukłym i niech $\phi : X \to \mathbb{R}^n$ będzie funkcją wypukłą. Niech $m \in \mathbb{N}$, $x_1, ..., x_m \in X$, $\alpha_1, ..., \alpha_m \geq 0$, $\sum_{i=1}^m \alpha_i = 1$. Wtedy*

$$\phi\left(\sum_{i=1}^m \alpha_i x_i\right) \;\leq\; \sum_{i=1}^m \alpha_i \; \phi(x_i). \tag{3}$$

Dowód twierdzenia 3 jest podobny do dowodu faktu, że zbiór wypukły zawiera wszystkie kombinacje wypukłe swoich elementów. Stosujemy indukcję względem m.

$\mathbf{m} = 1$. Mamy $\alpha_1 = 1$ i (3) sprowadza się do oczywistej nierówności $\phi(x_1) \leq \phi(x_1)$.

Krok indukcyjny. Załóżmy, że (3) zachodzi dla $m-1$. Niech $z = \sum_{i=1}^{m} \alpha_i x_i$, $x_i \in X$, $\alpha_i \geq 0$, $\sum_{i=1}^{m} \alpha_i = 1$. Rozważymy dwa przypadki.

a) $\alpha_m = 1$. W tym przypadku $\alpha_1 = ... = \alpha_{m-1} = 0$, $z = x_m$ i (3) sprowadza się do oczywistej nierówności $\phi(x_m) \leq \phi(x_m)$.

b) $\alpha_m < 1$. W tym przypadku, kładąc $y = \sum_{i=1}^{m-1} \frac{\alpha_i}{1-\alpha_m} x_i$ i używając założenia indukcyjnego, dostajemy

$$\begin{aligned}
\phi(z) &= \phi\left((1-\alpha_m)\sum_{i=1}^{m-1}\frac{\alpha_i}{1-\alpha_m}x_i + \alpha_m x_m\right) = \phi((1-\alpha_m)y + \alpha_m x_m) \\
&\leq (1-\alpha_m)\phi\left(\sum_{i=1}^{m-1}\frac{\alpha_i}{1-\alpha_m}x_i\right) + \alpha_m\phi(x_m) \\
&\leq (1-\alpha_m)\sum_{i=1}^{m-1}\frac{\alpha_i}{1-\alpha_m}\phi(x_i) + \alpha_m\phi(x_m) = \sum_{i=1}^{m}\alpha_i\,\phi(x_i).
\end{aligned}$$

□

W teorii programowania wypukłego będziemy często używać następującego prostego lematu.

Lemat 1 *Jeśli X jest wypukły i $\phi : X \to \mathbb{R}^n$ jest wypukła (wklęsła), to dla dowolnego $\lambda > 0$ zbiór $Z = \{x \in X : \phi(x) \leq (\geq)\lambda\}$ jest wypukły.*

Dowód. Jeśli ϕ jest wypukła, $x, y \in Z$ i $\alpha \in [0,1]$, to

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y) \leq \alpha\lambda + (1-\alpha)\lambda = \lambda,$$

więc $\alpha x + (1-\alpha)y \in Z$. Drugiej części lematu dowodzimy podobnie. □

Następujące twierdzenie podamy bez dowodu (można go znaleźć, podobnie jaki dowody wielu innych ciekawych rezultatów, np. w klasycznej monografii Rockafellara "Convex Analysis").

Twierdzenie 4 *Niech $\emptyset \neq X \subseteq \mathbb{R}^n$ będzie wypukły i niech $\phi : X \to \mathbb{R}^n$ będzie funkcją wypukłą. Wtedy ϕ jest ciągła w $int(X)$. Co więcej, dla każdego $x \in int(X)$ i $s \in \mathbb{R}^n$ takiego, że $||s|| = 1$, istnieje pochodna jednostronna*

$$\frac{\partial\phi}{\partial s}(x) = \lim_{h\downarrow 0}\frac{\phi(x+hs)-\phi(x)}{h}.$$

Uwaga. Nie znaczy to, że funkcja wypukła ϕ jest różniczkowalna w $int(X)$! Rozważmy na przykład funkcję $\phi : \mathbb{R} \to \mathbb{R}$ daną wzorem $\phi(x) = |x|$. Łatwo sprawdzić, że ϕ jest wypukła w $\mathbb{R}$ i $\frac{\partial\phi}{\partial 1}(0) = 1 \neq -1 = \frac{\partial\phi}{\partial(-1)}(0)$, a zatem ϕ nie jest różniczkowalna w zerze.

Twierdzenie 4 można znacznie wzmocnić. Na przykład przy założeniach tego twierdzenia ϕ jest ciągła w sensie Lipschitza w $int(X)$. Jeśli dodatkowo X jest domknięty i ma gładki (klasy C^2) brzeg, to na mocy twierdzenia Aleksandrowa ϕ jest dwukrotnie różniczkowalna w prawie każdym punkcie X (względem miary Lebesgue'a, będącej naturalnym uogólnieniem objętości w $\mathbb{R}^n$).