Metody optymalizacji. Wykład 2

Zbiory wypukłe w przestrzeni euklidesowej

Dla $x, y \in \mathbb{R}^n$, $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$, niech $(x, y) = \sum_{i=1}^n x_i y_i$ oznacza iloczyn skalarny x i y, a $||x|| = \sqrt{(x, x)}$ normę euklidesową (długość) wektora x. Przy x, y jak wyżej, będziemy pisać $x \geq y$, jeśli $x_1 \geq y_1$, $x_2 \geq y_2$, ... $x_n \geq y_n$ (nierówność między wektorami oznacza analogiczne nierówności na wszystkich ich współrzędnych).

Definicja 1 *Zbiór $A \subseteq \mathbb{R}^n$ nazywamy wypukłym, jeśli dla dowolnych $x, y \in A$, $\alpha \in [0, 1]$, mamy $\alpha x + (1 - \alpha) y \in A$.*

Innymi słowy, zbiór jest wypukły jeśli wraz z każdymi swoimi dwoma punktami zawiera łączący je odcinek. Wektor $\alpha x + (1 - \alpha) y$, $\alpha \in [0, 1]$, występujący w definicji 1 nazywamy *kombinacją wypukłą wektorów x* i y.

Nietrudno sprawdzić, że kula (otwarta lub domknięta) w $\mathbb{R}^n$ jest wypukła.

Lemat 1 *Niech A i B będą macierzami o wymiarach odpowiednio $m \times n$ i $k \times n$ dla pewnych $k, m, n \in \mathbb{N}$ i niech $a \in \mathbb{R}^m$, $b \in \mathbb{R}^k$. Wtedy zbiór*

$$X = \{x \in \mathbb{R}^n : Ax \geq a, Bx = b\}$$

jest wypukły.

Znaczy to, że zbiór w $\mathbb{R}^n$ zadany dowolną ilością ograniczeń (tzn. równań czy nierówności) liniowych jest wypukły.

Dowód. Dla $x, y \in X$, $\alpha \in [0, 1]$, mamy

$$\begin{aligned} A(\alpha x + (1 - \alpha) y) &= \alpha A x + (1 - \alpha) A y \geq \alpha a + (1 - \alpha) a = a, \\ B(\alpha x + (1 - \alpha) y) &= \alpha B x + (1 - \alpha) B y = \alpha b + (1 - \alpha) b = b. \end{aligned}$$

$\square$

Lemat 2 *Niech C_α, $\alpha \in I$, będzie rodziną zbiorów wypukłych w $\mathbb{R}^n$ indeksowaną przez pewien zbiór I. Wtedy zbiór $\bigcap_{\alpha \in I} C_\alpha$ jest też wypukły.*

Dowód - ćwiczenie.

Definicja 2 *Kombinacją wypukłą wektorów $x_1, ..., x_m \in \mathbb{R}^n$ nazywamy wektor postaci $x = \sum_{i=1}^m \alpha_i x_i$, gdzie $\alpha_i \geq 0$, $i = 1, ..., m$, $\sum_{i=1}^m \alpha_i = 1$. Liczbę m nazywamy długością tej kombinacji liniowej.*

Zauważmy, że w definicji 1 występowały kombinacje wypukłe o długości 2.

Lemat 3 *Zbiór $C \subseteq \mathbb{R}^n$ jest wypukły wtedy i tylko wtedy, gdy zawiera wszystkie kombinacje wypukłe swoich elementów.*

Dowód.
$\Leftarrow$ Oczywiste (wystarczy, że zawiera kombinacje o długości 2).

$\Rightarrow$ $x = 1 \cdot x$, więc z definicji wypukłości C zawiera wszystkie kombinacje wypukłe swoich elementów o długościach 1 i 2. Załóżmy, że C zawiera wszystkie kombinacje wypukłe swoich elementów o długości $m-1$, gdzie $m \geq 3$. Niech $x = \sum_{i=1}^m \alpha_i x_i$, $\alpha_i \geq 0$, $x_i \in C$, $i = 1, ..., m$, $\sum_{i=1}^m \alpha_i = 1$. Jeśli $\alpha_1 = 1$, x jest kombinacją wypukłą długości 1, więc z założenia $x \in C$. Jeśli $\alpha_1 \neq 1$, połóżmy $\beta_i = \alpha_i/(1-\alpha_1)$, $i = 2, ..., n$. Wtedy

$$x = \alpha_1 x_1 + (1-\alpha_1)(\beta_2 x_2 + ... + \beta_m x_m). \tag{1}$$

Ale $\beta_i \geq 0$ i

$$\sum_{i=2}^m \beta_i = \frac{1}{1-\alpha_1} \sum_{i=2}^m \alpha_i = \frac{1-\alpha_1}{1-\alpha_1} = 1,$$

a zatem z założenia indukcyjnego $\beta_2 x_2 + ... + \beta_m x_m \in C$. To zaś w połączeniu z (1) i wypukłością C daje $x \in C$. $\square$

Rzutowanie (projekcja) na zbiór wypukły

Definicja 3 *Odległością punktu $x \in \mathbb{R}^n$ od zbioru $Y \subseteq \mathbb{R}^n$ nazywamy liczbę*

$$\rho(x, Y) := \inf_{y \in Y} ||x - y||.$$

Projekcją (rzutem) punktu $x \in \mathbb{R}^n$ na zbiór $Y \subseteq \mathbb{R}^n$ nazywamy punkt $p \in Y$ taki, że $||x - p|| = \rho(x, Y)$.

W ogólnym przypadku projekcja punktu na zbiór może nie istnieć, a jeśli istnieje, może nie być wyznaczona jednoznacznie. Problemy te ilustrują następujące przykłady.

1. Niech $Y = B(0,1) = \{y \in \mathbb{R}^n : ||y|| < 1\}$ i niech $x = (3, 0, ..., 0)$. Nietrudno sprawdzić, że $\rho(x, Y) = 2$ (np. $y_n = (1 - 1/n, 0, ..., 0) \in Y$, $||x - y_n|| = 2 + 1/n \to 2$ przy $n \to \infty$), ale projekcja x na Y nie istnieje.

2. Rozważmy zbiór $Y = (-\infty, -1] \cup [1, \infty)$ w $\mathbb{R}^1$ i niech $x = 0$. Wtedy $\rho(x, Y) = 1$ i punkty $-1, 1$ są projekcjami x na Y.

Następujące twierdzenie podaje wygodne warunki dostateczne na istnienie i jednoznaczność projekcji.

Twierdzenie 1 *Jeśli zbiór $Y \subseteq \mathbb{R}^n$ jest domknięty, to rzut dowolnego punktu $x \in \mathbb{R}^n$ na Y istnieje. Jeśli dodatkowo Y jest wypukły, rzut ten jest wyznaczony jednoznacznie.*

Powyższe przykłady pokazują, że założeń twierdzenia 1 nie da się pominąć.

Dowód. Wybierzmy ciąg $y_n \in Y$ taki, że

$$||x - y_n|| < \rho(x, Y) + \frac{1}{n}, \qquad n = 1, 2, ...$$

Ciąg $\{y_n\}$ jest ograniczony, zatem istnieje jego podciąg $\{y_{n_k}\}$ zbieżny do pewnego $p \in \mathbb{R}^n$. Ale Y jest domknięty, zatem $p \in Y$. Ponadto z ciągłości normy

$$||x - p|| = ||x - \lim_{k\to\infty} y_{n_k}|| = \lim_{k\to\infty} ||x - y_{n_k}|| \leq \rho(x, Y).$$

Oczywiście $||x - p|| \geq \rho(x, Y)$, gdyż $p \in Y$, a zatem p jest rzutem x na Y.

Załóżmy dodatkowo, że Y jest wypukły, i że istnieją $p', p'' \in Y$, $p' \neq p''$, takie, że $||x-p'|| = ||x-p''|| = \rho(x, Y)$. Z wypukłości Y, $p := \frac{1}{2}(p'+p'') \in Y$. Łatwo widać, że $||x-p|| < ||x-p'||$. Rzeczywiście, z twierdzenia Pitagorasa,

$$||x - p||^2 + \left(\frac{||p' - p''||}{2}\right)^2 = ||x - p||^2 + ||p - p'||^2 = ||x - p'||^2.$$

Dostaliśmy sprzeczność, ponieważ $||x - p'|| = \rho(x, Y)$. $\square$

W dalszym ciągu potrzebna nam będzie następująca geometryczna charakteryzacja projekcji.

Twierdzenie 2 *Niech $x \in \mathbb{R}^n$ i niech Y będzie domkniętym i wypukłym podzbiorem $\mathbb{R}^n$. Punkt $p \in Y$ jest rzutem x na Y wtedy i tylko wtedy, gdy dla każdego $y \in Y$*

$$(y - p, x - p) \leq 0. \tag{2}$$

Intuicja geometryczna: jeśli $x \neq p \neq y$, to

$$\frac{(y - p, x - p)}{||y - p||\,||x - p||}$$

jest cosinusem kąta między wektorami $y - p$ i $x - p$. Twierdzenie 2 mówi, że aby p było projekcją x na Y potrzeba i wystarcza, by dla dowolnego $y \in Y$, $y \neq p$, kąt między $y - p$ i $x - p$ był rozwarty lub prosty (ale nigdy ostry).

Dowód.

$\Rightarrow$ Jeśli p jest rzutem x na Y, to na mocy wypukłości Y dla dowolnego $y \in Y$ i $\alpha \in (0,1)$ wektor $z = \alpha y + (1-\alpha)p \in Y$, a zatem z własności iloczynu skalarnego

$$\begin{aligned}
||x-p||^2 &\le ||x-z||^2 = ||x-\alpha y-(1-\alpha)p||^2 = ||(x-p)-\alpha(y-p)||^2 \\
&= ((x-p)-\alpha(y-p),(x-p)-\alpha(y-p)) \\
&= (x-p,x-p)-2\alpha(y-p,x-p)+\alpha^2(y-p,y-p) \\
&= ||x-p||^2-2\alpha(y-p,x-p)+\alpha^2||y-p||^2,
\end{aligned}$$

a zatem

$$0 \le \alpha^2||y-p||^2-2\alpha(y-p,x-p) = \alpha(\alpha||y-p||^2-2(y-p,x-p)),$$

co dla małych $\alpha > 0$ jest możliwe tylko gdy zachodzi (2).

$\Leftarrow$ Dla dowolnego $y \in Y$, na mocy warunku (2),

$$\begin{aligned}
||y-x||^2 &= ||(y-p)+(p-x)||^2 = ((y-p)+(p-x),(y-p)+(p-x)) \\
&= (y-p,y-p)+2(y-p,p-x)+(p-x,p-x) \\
&= ||y-p||^2+2(y-p,p-x)+||p-x||^2 \\
&\ge ||p-x||^2,
\end{aligned}$$

a zatem p jest projekcją x na Y. $\square$

Twierdzenia o oddzielaniu

Definicja 4 *Niech $c \in \mathbb{R}^n \setminus \{0\}$, $\lambda \in \mathbb{R}$. Zbiory*

$$\{x \in \mathbb{R}^n : (c,x) \le \lambda\}, \qquad \{x \in \mathbb{R}^n : (c,x) \ge \lambda\}$$

nazywamy półprzestrzeniami, a zbiór

$$\Pi = \{x \in \mathbb{R}^n : (c,x) = \lambda\}$$

hiperpłaszczyzną ($n-1$-wymiarową w $\mathbb{R}^n$). c jest wektorem normalnym (tzn. prostopadłym) do hiperpłaszczyzny Π.

W analizie wypukłej i jej zastosowaniach wielką rolę odgrywają twierdzenia o oddzielaniu. Podamy tu trzy tego typu rezultaty.

Twierdzenie 3 (Oddzielanie punktu od zbioru wypukłego i domkniętego) *Niech $Y \subseteq \mathbb{R}^n$ będzie domknięty i wypukły i niech $x \in \mathbb{R}^n \setminus Y$. Istnieje wtedy hiperpłaszczyzna*

$$\Pi = \{y \in \mathbb{R}^n : (c,y) = \lambda\} \tag{3}$$

taka, że $x \in \Pi$ i $(c,y) < \lambda$ dla każdego $y \in Y$.

Uwaga. Hiperpłaszczyzna taka nie jest na ogół wyznaczona jednoznacznie.

Dowód. Niech p będzie rzutem x na Y i niech $c = x - p$, $\lambda = (c, x)$. (Zauważmy, że $c \neq 0$, bo $x \notin Y$.) Z definicji λ, x należy do hiperpłaszczyzny Π danej przez (3). Z twierdzenia 2 wynika, że dla każdego $y \in Y$

$$(y, x - p) \leq (p, x - p).$$

Z drugiej strony, $0 < ||x - p||^2 = (x - p, x - p)$, więc

$$(p, x - p) < (x, x - p).$$

Stąd dla każdego $y \in Y$

$$(c, y) = (x - p, y) = (y, x - p) < (x, x - p) = (c, x) = \lambda.$$

$\square$

Niech $A \subseteq \mathbb{R}^n$. W dalszym ciągu przez $int(A)$ będziemy oznaczać wnętrze A, przez $\overline{A}$ jego domknięcie, a przez ∂A jego brzeg.

Twierdzenie 4 (O hiperpłaszczyźnie podpierającej) *Niech $X \subseteq \mathbb{R}^n$ będzie zbiorem domkniętym i wypukłym i niech $x_0 \in \partial X$. Istnieje wtedy hiperpłaszczyzna $\Pi = \{x \in \mathbb{R}^n : (c, x) = \lambda\}$ podpierająca zbiór X w x_0, tzn. taka, że $x_0 \in \Pi$ i dla wszystkich $x \in X$ mamy $(c, x) \leq \lambda$.*

Dowód. Ponieważ $x_0 \in \partial X$, istnieje ciąg wektorów $v_n \in \mathbb{R}^n \setminus X$ taki, że $v_n \to x_0$. Na mocy twierdzenia 3 istnieją hiperpłaszczyzny $\Pi_n = \{x \in \mathbb{R}^n : (c_n, x) = \lambda_n\}$ takie, że $\lambda_n = (c_n, v_n)$ i

$$(c_n, x) < \lambda_n, \qquad x \in X, \quad n \in \mathbb{N}. \tag{4}$$

Możemy założyć, że $||c_n|| = 1$ dla każdego n (jeśli tak nie jest, weźmy $c'_n = c_n/||c_n||$ i $\lambda'_n = \lambda_n/||c_n||$, wtedy $||c'_n|| = 1$ i $\Pi_n = \{x \in \mathbb{R}^n : (c'_n, x) = \lambda'_n\}$). Zatem dla pewnego podciągu indeksów $\{n_k\}$ i pewnego wektora $c \in \mathbb{R}^n$ mamy $c_{n_k} \to c$ przy $k \to \infty$. Zauważmy, że

$$||c|| = ||\lim_{k\to\infty} c_{n_k}|| = \lim_{k\to\infty} ||c_{n_k}|| = 1,$$

więc $c \neq 0$. Z ciągłości iloczynu skalarnego,

$$(c, x_0) = \left(\lim_{k\to\infty} c_{n_k}, \lim_{k\to\infty} v_{n_k}\right) = \lim_{k\to\infty} (c_{n_k}, v_{n_k}) = \lim_{k\to\infty} \lambda_{n_k}.$$

Połóżmy $\lambda = \lim_{k\to\infty} \lambda_{n_k}$. Wtedy $(c, x_0) = \lambda$, zatem $x_0 \in \Pi = \{x \in \mathbb{R}^n : (c, x) = \lambda\}$. Ponadto dla $x \in X$, na mocy (4),

$$(c, x) = \lim_{k\to\infty} (c_{n_k}, x) \leq \lim_{k\to\infty} \lambda_{n_k} = \lambda,$$

a zatem Π jest hiperpłaszczyzną podpierającą X w x_0. $\square$

W dowodzie następnego twierdzenia będzie nam potrzebny

Lemat 4 *Niech $X \subseteq \mathbb{R}^n$ będzie wypukły. Wtedy*
(a) jeśli $int(X) \neq \emptyset$, to $int(X)$ jest zbiorem wypukłym;
(b) zbiór $\overline{X}$ jest wypukły.

Dowód.

(a) Niech $x, y \in int(X)$. Istnieje wtedy $\epsilon > 0$ taki, że kule otwarte $B(x, \epsilon) = \{z \in \mathbb{R}^n : ||z - x|| < \epsilon\} \subseteq X$ i $B(y, \epsilon) \subseteq X$. Z wypukłości X,

$$Z = \{z = \alpha x' + (1 - \alpha)y' : x' \in B(x, \epsilon), y' \in B(y, \epsilon), \alpha \in [0, 1]\} \subseteq X.$$

Nietrudno sprawdzić, że zbiór Z jest otwarty, zatem $Z \subseteq int(X)$. Co więcej, $\alpha x + (1 - \alpha)y \in Z$ dla każdego $\alpha \in [0, 1]$.

(b) Niech $x, y \in \overline{X}$, $\alpha \in [0, 1]$. Istnieją wtedy wektory $x_n, y_n \in X$ takie, że przy $n \to \infty$ mamy $x_n \to x$, $y_n \to y$. Z wypukłości X, $\alpha x_n + (1-\alpha)y_n \in X$ i

$$\alpha x_n + (1 - \alpha)y_n \to \alpha x + (1 - \alpha)y,$$

a zatem $\alpha x + (1 - \alpha)y \in \overline{X}$. $\square$

Twierdzenie 5 (O rozdzielaniu zbiorów wypukłych hiperpłaszczyzną) *Niech X i Y będą wypukłymi i domkniętymi podzbiorami $\mathbb{R}^n$ takimi, że $int(X) \neq \emptyset$ i $int(X) \cap Y = \emptyset$. Istnieje wtedy hiperpłaszczyzna Π rozdzielająca X i Y, tzn. istnieje wektor $c \in \mathbb{R}^n \setminus \{0\}$ taki, że*

$$(c, y) \leq (c, x) \qquad x \in X, \;\; y \in Y. \tag{5}$$

Dowód. Na mocy lematu 4 zbiór $int(X)$ jest wypukły. Łatwo sprawdzić, że

$$Z = \{z \in \mathbb{R}^n : z = y - x, y \in Y, x \in int(X)\}$$

jest zbiorem wypukłym. Z założenia $0 \notin Z$ (gdyby $0 \in Z$, mielibyśmy $x = y$ dla pewnych $x \in int(X)$, $y \in Y$, co przeczy założeniu). Istnieje zatem wektor $c \neq 0$ taki, że dla dowolnego $z \in Z$

$$(c, z) = (c, y - x) \leq (c, 0) = 0.$$

Rzeczywiście, jeśli $0 \notin \overline{Z}$, wynika to z twierdzenia 3 o oddzielaniu punktu (0) od zbioru wypukłego i domkniętego ($\overline{Z}$) hiperpłaszczyzną. Jeśli zaś $0 \in \overline{Z}$, powyższy fakt wynika z twierdzenia (4) o istnieniu hiperpłaszczyzny podpierającej ($\overline{Z}$ w 0). Stąd dla każdego $x \in int(X)$, $y \in Y$, mamy $(c, y) \leq (c, x)$. Jeśli $x \in X \setminus int(X) = \partial X$, istnieją punkty $x_n \in int(X)$ takie, że $x_n \to x$, a zatem $(c, y) \leq (c, x_n) \to (c, x)$ dla dowolnego $y \in Y$. W ten sposób wykazaliśmy (5). $\square$