Elementy optymalizacji wypukłej

Przypomnijmy, że $f: A \to \mathbb{R}$ jest wypukła, gdy

$\uparrow$ zbiór wypukły

$$\forall_{x,y\in A} \ \forall_{\alpha\in[0,1]} \quad f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y). \qquad (*)$$

Z $(*)$ natychmiast wynika, że dla funkcji wypukłych zachodzi nierówność

$$\forall_{x,y\in A} \quad f(\alpha x + (1-\alpha)y) \leq \max(f(x), f(y)) \qquad (**)$$

Ta nierówność jest spełniona nie tylko dla funkcji wypukłych; od ręki można sprawdzić, że spełnia ją na przykład $f: \mathbb{R}\to\mathbb{R}$ dana wzorem $f(x) = 1 - \chi_{[-1,1]}(x)$, która nie jest wypukła, ani nawet ciągła.

$(**)$ spełniają też wszystkie funkcje monotoniczne (gdy $A \subset \mathbb{R}$).

<u>Definicja</u>: Niech $A \subset \mathbb{R}^n$ będzie zbiorem wypukłym. Funkcję $f: A\to\mathbb{R}$ nazywamy quasiwypukłą (quasiwklęsłą), gdy $\forall_{x,y\in A} \ \forall_{\alpha\in[0,1]}$

$$f(\alpha x + (1-\alpha)y) \leq \max(f(x), f(y))$$

$$\left(f(\alpha x + (1-\alpha)y) \geq \min(f(x), f(y)) \right).$$

Oczywiście funkcje wypukłe są quasiwypukłe, wklęsłe są quasiwklęsłe.

Zadanie: scharakteryzować wszystkie $f: A \to \mathbb{R}$, które są równocześnie quasiwypukłe i quasiwklęsłe. (najpierw dla $A \subset \mathbb{R}$, potem np. dla $A = [0,1]^2$)

Twierdzenie

Funkcja $f: A \to \mathbb{R}$ jest quasiwypukła (quasiwklęsła) wtedy i tylko wtedy, gdy $X_a = \{x \in A : f(x) \leq a\}$ jest wypukły ($Y_a = \{x \in A : f(x) \geq a\}$ jest wypukły).

Dowód:

$\Rightarrow$ niech $x, y \in X_a$ (czyli $f(x) \leq a$, $f(y) \leq a$); ustalmy $\alpha \in [0,1]$

Wtedy $f(\alpha x + (1-\alpha)y) \leq \max(f(x), f(y)) \leq a$,

więc $\alpha x + (1-\alpha) \in X_a \Rightarrow X_a$ jest wypukły.

$\Leftarrow$ Wybierzmy $x, y \in A$ i $\alpha \in [0,1]$ i niech $a = \max(f(x), f(y))$.

Wtedy $x, y \in X_a$, wiemy też, że X_a jest wypukły, więc

$\alpha x + (1-\alpha)y \in X_a \Rightarrow f(\alpha x + (1-\alpha)y) \leq a = \max(f(x), f(y))$.

Dla funkcji quasiwklęsłych dowód jest analogiczny.

Twierdzenia o oddzielaniu

Przypomnijmy ważny, ale prościutki fakt. Niech $A \subset \mathbb{R}^n$ będzie wypukły i domknięty.

Fakt: Funkcja $f: A \to \mathbb{R}$ ściśle wypukła przyjmuje na A minimum w co najwyżej jednym punkcie.

Dowód: Gdyby dla $x \neq y$, $x, y \in A$ było $f(x) = f(y) = \min_A f$,

to ~~$f(x)$~~ $f\left(\frac{x+y}{2}\right) < \frac{f(x)+f(y)}{2} = a$ ↯.

Pierwsze twierdzenie o oddzielaniu

Załóżmy, że $A \subset \mathbb{R}^n$ jest wypukły i domknięty, $y \notin A$. Istnieją wówczas $b \in \mathbb{R}$ i $v \in \mathbb{R}^n \setminus \{0\}$ takie, że

$$\forall_{x \in A} \quad \langle v, y \rangle < b \leq \langle v, x \rangle$$

Co oznaczają nierówności powyżej? Zbiór ~~bz~~ $\{x \in \mathbb{R}^n : b \leq \langle v, x \rangle\}$ to pewna domknięta półprzestrzeń ($b = \langle v, x \rangle$ opisuje hiperpłaszczyznę kowymiaru 1, prostopadłą do v), twierdzenie mówi, że A możemy oddzielić od y hiperpłaszczyzną $\{b = \langle v, x \rangle\}$ tak, że A leży po jednej jej stronie, a y po drugiej.

Dowód: Krok 1 Zauważmy, że w A istnieje punkt z najbliższy punktu y, i to dokładnie 1. Weźmy bowiem dowolne $w \in A$ i rozważmy funkcję wypukłą $f(x) = \|x - y\|^2$ na zbiorze $C = A \cap \overline{B}(y, \|y - w\|)$. Jest to zbiór zwarty i niepusty (bo $w \in C$), więc f przyjmuje i wypukły na C ~~doktao~~ minimum w dokładnie jednym punkcie $z \in C \subset A$.

Krok 2 Niech $v = \frac{z - y}{\cancel{\quad}}$. Zauważmy, że dla każdego $x \in A$ cały odcinek $[x, z]$ leży w A. Funkcja f nie maleje, gdy poruszamy się z z w kierunku x, więc

$$D_{x-z} f(z) = \langle \nabla f(z), x - z \rangle \geq 0$$

$$\langle \nabla f(z), x - z \rangle = \langle 2(z - y), x - z \rangle = 2\langle v, x \rangle - 2\langle v, z \rangle$$

czyli $\langle v, x \rangle \geq \langle v, z \rangle$.

Niech zatem $b = \langle v, z\rangle = \langle z-y, z\rangle$, wiemy już, że

$$\forall_{x\in A} \quad \langle v, x\rangle \geq b.$$

Mamy też $\underset{\langle v,\, z-y\rangle \,=\, \underset{b}{\langle v, z\rangle} \,-\, \langle v, y\rangle}{\|v\|^2} = \|z-y\|^2 > 0$, bo $y\notin A$, $z\in A$.

$$\|v\|^2 = \langle v, z-y\rangle = \underbrace{\langle v, z\rangle}_{b} - \langle v, y\rangle,$$ skąd

$$\langle v, y\rangle < b. \quad \square.$$

Biorąc $c = \frac{1}{2}\left(b + \langle v, y\rangle\right)$ dostajemy od razu

<u>Wniosek</u> (druga wersja pierwszego tw. o oddzielaniu)

$\cancel{\forall_{x\in A}}$ Przy założeniach pierwszego twierdzenia o oddzielaniu istnieje $v\in\mathbb{R}^n$ i $c\in\mathbb{R}$ takie, że

$$\forall_{x\in A} \quad \langle v, y\rangle < c < \langle v, x\rangle.$$

<u>Twierdzenie o hiperpłaszczyźnie podpierającej</u>

Niech $A\subset\mathbb{R}^n$ będzie wypukły i domknięty, $z\in\partial A$ (tzn. w dowolnym otoczeniu z są punkty z A i punkty z $\mathbb{R}^n\setminus A$). Istnieje hiperpłaszczyzna kowymiaru 1, przechodząca przez z, taka, że A leży po jednej jej stronie. Innymi słowy, istnieje $v\in\mathbb{R}^n\setminus\{0\}$ takie, że $\forall_{x\in A} \ \langle v, z\rangle \leq \langle v, x\rangle$.

Dowód: Niech $p_i \in \mathbb{R}^n \setminus A$, $p_i \to z$.

Z ~~pierwszego~~ (dowodu) tw. o oddzielaniu każdy z punktów p_i możemy oddzielić od A hiperpłaszczyzną

$P_i = \{x \in \mathbb{R}^n : \langle v_i, x\rangle = \langle v_i, z_i\rangle\}$

gdzie $v_i = z_i - p_i$, zaś z_i jest punktem A najbliższym p_i.

Zauważmy też, że w warunku na P_i możemy w miejsce v_i wziąć $\tilde{v}_i = \frac{v_i}{\|v_i\|} \in \mathbb{S}^{n-1}$.

Mamy $\|z_i - z\| \leq \|z_i - p_i\| + \|p_i - z\| \leq \|z - p_i\| + \|p_i - z\|$
$= 2\|p_i - z\| \to 0$, więc $z_i \to z$.

Wybierzmy wreszcie podciąg (p_{i_j}) taki, by
$\tilde{v}_{i_j} = \frac{v_{i_j}}{\|v_{i_j}\|}$ zbiegały do pewnego $v \in \mathbb{S}^{n-1}$.

Dla każdego j zachodzi nierówność

$$\forall_{x \in A} \quad \langle \tilde{v}_{i_j}, x\rangle \geq \langle \tilde{v}_{i_j}, z_{i_j}\rangle$$

przechodząc do granicy z $j \to \infty$

$$\forall_{x \in A} \quad \langle v, x\rangle \geq \langle v, z\rangle .$$

Oczywiście ~~płaszczy~~ hiperpłaszczyzna $P = \{y \in \mathbb{R}^n : \langle v, y\rangle = \langle v, z\rangle\}$ przechodzi przez z.

$\square$.

Takie wersje twierdzenia o oddzielaniu nie zachodzą, gdy A nie jest domknięty – nie sposób oddzielić kuli otwartej B od punktu na jej brzegu, choć nie należy on do B. Mamy jednak kolejną wersję:

<u>Wniosek</u>: Niech $\mathcal{U} \subset \mathbb{R}^n$ będzie wypukły, $x \notin \mathcal{U}$. Istnieją wówczas $v \in \mathbb{R}^n \setminus \{0\}$ i $\alpha \in \mathbb{R}$ takie, że

$$\forall_{y \in \mathcal{U}} \quad \langle x, v \rangle \leq \alpha \leq \langle y, v \rangle.$$

<u>Dowód</u>: albo $x \notin \operatorname{cl} \mathcal{U}$ i możemy użyć pierwszego twierdzenia o oddzielaniu, by ~~się~~ x odseparować od $\operatorname{cl} \mathcal{U}$, albo $x \in \operatorname{cl} \mathcal{U}$ i możemy użyć twierdzenia o płaszczyźnie podpierającej $\operatorname{cl} \mathcal{U}$ w punkcie $x \in \partial \operatorname{cl} \mathcal{U}$.

<u>Drugie twierdzenie o oddzielaniu</u>

Załóżmy, że $A, B \subset \mathbb{R}^n$ są wypukłe, domknięte i rozłączne oraz przynajmniej jeden z nich jest ograniczony.

Istnieją wówczas $c \in \mathbb{R}$, $v \in \mathbb{R}^n \setminus \{0\}$ takie, że

$$\forall_{x \in A} \ \forall_{y \in B} \quad \langle v, y \rangle < c < \langle v, x \rangle$$

Czyli A i B można rozdzielić hiperpłaszczyzną

$$P = \{ z \in \mathbb{R}^n : \langle v, z \rangle = c \}.$$

Dowód: potrzebować będziemy sum i różnic Minkowskiego

$$A \pm B = \{x \pm y : x \in A, y \in B\} \quad \text{dla } A, B \subset \mathbb{R}^n.$$

Zadania: ① jeżeli A, B są wypukłe, to $A+B$ i $A-B$ też są wypukłe

② jeżeli A, B są domknięte i przynajmniej jeden z nich jest zwarty, to $A+B$ i $A-B$ są domknięte

③ znajdź przykład A, B domkniętych takich, że $A+B$ nie jest domknięty.

Niech B będzie ograniczony (więc zwarty).

Skoro $A \cap B = \emptyset$, to $0 \notin A - B$ i $A-B$, jak wynika z zadań, jest domknięty i wypukły. Istnieje zatem $\gamma \in \mathbb{R}$ i $v \in \mathbb{R}^n \setminus \{0\}$ takie, że

$$\forall_{z \in A-B} \quad 0 = \langle v, 0 \rangle < \gamma < \langle v, z \rangle$$

oddzielamy 0 od zbioru $A-B$

czyli

$$\forall_{x \in A} \forall_{y \in B} \quad 0 < \gamma < \langle v, x-y \rangle, \text{ czyli}$$

$$\langle v, x \rangle > \langle v, y \rangle + \gamma \quad \text{dla pewnego } \gamma > 0.$$

w szczególności

$$\inf_{x \in A} \langle v, x \rangle \geq \sup_{y \in B} \langle v, y \rangle + \gamma.$$

Biorąc więc $c = \underbrace{\sup_{y \in B} \langle v, y \rangle}_{\text{skończone, bo } B \text{ zwarty}} + \frac{\gamma}{2}$ dostajemy też

zresztą $\sup_{y \in B} \langle v, y \rangle < \inf_{x \in A} \langle v, x \rangle \leq \langle v, x_0 \rangle$ dla dow. $x_0 \in A$.

Analogicznie, korzystając ze słabszej wersji tw. o oddzielaniu punktu od zbioru wypukłego (niekoniecznie domkniętego) dostajemy

<u>Twierdzenie</u>: Niech $U, V \subset \mathbb{R}^n$ będą wypukłe i rozłączne. Istnieje wówczas $v \in \mathbb{R}^n \setminus \{0\}$ i $\alpha \in \mathbb{R}$ tż.

$$\forall_{x \in U} \forall_{y \in V} \quad \langle x, v \rangle \leq \alpha \leq \langle y, v \rangle,$$

czyli U i V można oddzielić hiperpłaszczyzną (choć może ona zawierać punkty U i V).