Twierdzenie (warunki pierwszego rzędu na wypukłość)

Niech $\mathcal{U} \subset \mathbb{R}^n$ będzie otwarty, $A \subset \mathcal{U}$ wypukły.

Funkcja $f : \mathcal{U} \to \mathbb{R}$, różniczkowalna na $\mathcal{U}$, jest wypukła na A (tzn. $f|_A$ jest wypukła) wtedy i tylko wtedy, gdy

$$\forall_{x,y \in A} \quad f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle; \qquad (*)$$

f jest ściśle wypukła na A wtedy i tylko wtedy, gdy

$$\forall_{\substack{x,y \in A \\ x \neq y}} \quad f(y) > f(x) + \langle \nabla f(x), y-x \rangle.$$

Dowód: Załóżmy, że $f|_A$ jest wypukła. Dla każdego $t \in [0,1]$ $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$

$$\iff \quad f(y) \geq f(x) + \frac{f(x+t(y-x)) - f(x)}{t}$$

i biorąc $t \to 0^+$ dostajemy

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

Załóżmy teraz, że f spełnia warunek $(*)$.

Niech $z = \alpha x + (1-\alpha) y \in A$. Mamy $\alpha \in [0,1]$

$$f(x) \geq f(z) + \langle \nabla f(z), x-z \rangle$$
$$f(y) \geq f(z) + \langle \nabla f(z), y-z \rangle$$

Mnożymy pierwszą ~~równanie~~ nierówność przez α, drugą przez $(1-\alpha)$ i dodajemy stronami:

$$\alpha f(x) + (1-\alpha) f(y) \geqslant \underset{\substack{\| \\ 1}}{(\alpha + (1-\alpha))} f(z) + \langle \nabla f(z), \underbrace{\alpha(x-z) + (1-\alpha)(y-z)}_{\substack{\| \\ 0}} \rangle$$

i stąd $f(\alpha x + (1-\alpha) y) = f(z) \leqslant \alpha f(x) + (1-\alpha) f(y)$,

cyli f jest wypukła.

Przypadek f ściśle wypukłej zostawiam jako zadanie, proszę przypomnieć sobie postać ~~tw~~ (i dowód) twierdzenia o monotoniczności ilorazów różnicowych dla f ściśle wypukłych. $\square$.

Uwaga: funkcje wklęsłe i ściśle wklęsłe spełniają analogiczne warunki, z nierównościami w przeciwną stronę.

Wniosek: Niech U, A, f będą jak w warunkach I rzędu na wypukłość. Jeżeli dla pewnego $x_0 \in A$ zachodzi $\nabla f(x_0) = 0$, to $f(x_0) = \min_A f$, (a więc warunek konieczny Fermata jest warunkiem dostatecznym!)

Dowód: od razu z war. I rzędu:

$$\forall_{y \in A} \quad f(y) \geqslant f(x_0) + \langle \nabla f(x_0), y - x_0 \rangle = f(x_0) \quad \square$$

Jeżeli f jest ściśle wypukła, to to minimum jest właściwe.

Będziemy chcieli osłabić nieco te warunki na f.

Definicja: Niech $f: A \to \mathbb{R}$. Mówimy, że $v \in \mathbb{R}^n$ jest **subgradientem** f w $x \in A$, gdy dla każdego $y \in A$ zachodzi $f(y) \geq f(x) + \langle v, y-x \rangle$. (*)

Uwaga: pojęcie subgradientu można definiować dla dowolnej funkcji f, czasem żąda się, by (*) nie zachodziła dla wszystkich $y \in A$, ale tylko dla y dost. bliskich x itp. Prawdziwy pożytek z tego pojęcia jest jednak dla funkcji wypukłych.

Def: **Subróżniczka** funkcji wypukłej $f: A \to \mathbb{R}$ w punkcie $x \in A$ to zbiór wszystkich subgradientów f w x. Oznaczany jest $\partial f(x)$.

Przykład: Niech $f: \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$.

$$\partial f(0) = [-\tfrac{1}{2}, \tfrac{1}{2}].$$

Fakcik: Dla każdego $x \in A$ subróżniczka $\partial f(x)$ jest zbiorem wypukłym

Dowód: Załóżmy, że $v, w \in \partial f(x)$ i niech $\alpha \in [0,1]$.

Dla każdego $y \in A$

$$f(x) + \langle \alpha v + (1-\alpha) w, y-x \rangle = \alpha \big(\overbrace{f(x) + \langle v, y-x \rangle}^{\leq f(y),\ \text{bo}\ v \in \partial f(x)}\big) + (1-\alpha)\big(\overbrace{f(x) + \langle w, y-x \rangle}^{\leq f(y),\ \text{bo}\ w \in \partial f(x)}\big)$$

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

zatem $\alpha v + (1-\alpha) w \in \partial f(x)$ $\square$.

<u>Tw.</u>: Jeżeli $f: A \to \mathbb{R}$ jest wypukła, to dla każdego $x \in \operatorname{int} A$ $\partial f(x) \neq \emptyset$ (czyli istnieje subgradient f w x.

<u>Dowód:</u>

<u>Zadanko:</u> Jeżeli f jest wypukła na $A \subset \mathbb{R}^n$, to nadwykres $\operatorname{epi} f \subset \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$, $\operatorname{epi} f = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in A,\ y \geq f(x)\}$ jest zbiorem wypukłym, a punkty wykresu $\operatorname{graph} f = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in A,\ y = f(x)\}$ leżą na brzegu $\operatorname{epi} f$.

Jak już zrobimy zadanko, to wiemy, że ~~epi~~ dla $x \in A$ możemy znaleźć ~~H~~ hiperpłaszczyznę podpierającą $\operatorname{epi} f$ w punkcie $(x, f(x))$: istnieje $0 \neq v = (\alpha, \beta) \in \mathbb{R}^{n+1}$ taki, że

$$\forall_{(y,z) \in \operatorname{epi} f} \quad \langle (y,z), v \rangle \leq \langle (x, f(x)), v \rangle, \text{ czyli}$$

$$\langle \alpha, (y-x) \rangle + \beta (z - f(x)) \leq 0 \qquad (*)$$

Zauważmy, że jeżeli $(y,z) \in \operatorname{epi} f$, to ~~$(y,z') \in$~~ dla każdego $z' > z$ $(y,z') \in \operatorname{epi} f$, więc w $(*)$ możemy wziąć dowolnie duże z. To oznacza, że, aby $(*)$ mogła zachodzić, β musi być ≤ 0.

Załóżmy, że $\beta = 0$. Wtedy $\forall_{y \in A}$ mamy $\langle \alpha, (y-x) \rangle \leq 0$ czyli $\alpha \in N_A(x)$, ale $x \in \operatorname{int} A \Rightarrow N_A(x) = \{0\}$.
Sprzeczność z tym, że $v = (\alpha, \beta) \neq 0$.
Stąd $\beta < 0$.

Podzielmy obie strony (*) przez $|\beta|$.

$$\left\langle \frac{\alpha}{|\beta|}, y-x \right\rangle - z + f(x) \leq 0 \qquad \forall_{(y,z)\in \operatorname{epi} f}$$

w szczególności dla $(y,z)=(y,f(y))$:

$$\left\langle \frac{\alpha}{|\beta|}, y-x \right\rangle - f(y) + f(x) \leq 0 \iff f(y) \geq f(x) + \left\langle \frac{\alpha}{|\beta|}, y-x \right\rangle$$

czyli $\alpha/|\beta| \in \partial f(x)$. $\square$.

<u>Zadanko</u>: Wykaż, że w $x \in \partial A$ może się zdarzyć, że $\partial f(x) = \emptyset$.

<u>Twierdzenie</u>: Niech A będzie wypukły i otwarty, $f: A \to \mathbb{R}$. Jeżeli dla każdego $x \in A$ f ma w x subgradient, to f jest wypukła w A.

<u>Dowód</u>: powtarzamy część dowodu warunków I rzędu na wypukłość:
niech $x, y \in A$, $\alpha \in [0,1]$ i niech $z = \alpha x + (1-\alpha) y$
Niech v będzie subgradientem f w z.

$$\begin{aligned} & f(x) \geq f(z) + \langle v, x-z \rangle && \cdot\ \alpha \\ + \quad & f(y) \geq f(z) + \langle v, y-z \rangle && \cdot\ (1-\alpha) \end{aligned}$$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(z) + \langle v, \alpha(x-z) + (1-\alpha) y - z \rangle = f(z). \ \square.$$

Z warunków <u>I</u> rzędu wynika, że jeżeli f jest wypukła na A i różniczkowalna <u>we wszystkich punktach A</u>, to $\forall_{x\in A}\ \nabla f(x) \in \partial f(x)$.

Wykażemy znacznie mocniejszy fakt

<u>Twierdzenie</u> Niech $A \subset \mathbb{R}^n$ będzie wypukły i otwarty, $f : A \to \mathbb{R}$ niech będzie wypukła. Jeżeli f jest różniczkowalna w $x \in A$, to $\partial f(x) = \{\nabla f(x)\}$.

<u>Dowód</u>: ~~Niech $\varphi(t) = f$~~

Ustalmy $v \in \mathbb{R}^n$ i niech $\varphi(t) = f(x + tv)$.

Funkcja φ jest różniczkowalna w $t = 0$,

$$\varphi(t) = \varphi(0) + t\varphi'(0) + o(t) = f(x) + t\langle \nabla f(x), v\rangle + o(t)$$

$$\varphi(t) = f(x+tv)$$

Wiemy też, że f ma w x jakiś subgradient w.

$$f(x+tv) \geq f(x) + \langle w, tv\rangle = f(x) + t\langle w, v\rangle$$

Porządkując, dostajemy $\langle \nabla f(x) - w, v\rangle + \frac{o(t)}{t} \geq 0$

~~i mier~~

$$\downarrow t \to 0$$

$$\langle \nabla f(x) - w, v\rangle \geq 0$$

i zachodzi to dla każdego $v \in \mathbb{R}^n$, w szczególności dla $v = -(\nabla f(x) - w)$. Stąd $\|\nabla f(x) - w\|^2 = 0 \Rightarrow w = \nabla f(x)$

□.

Twierdzenie: Niech $A \subset \mathbb{R}^n$ będzie otwarty i wypukły. Wówczas f przyjmuje w $x_0 \in A$ minimum (globalne na A) wtedy i tylko wtedy, gdy $0 \in \partial f(x_0)$.

Dowód:

Jeżeli $0 \in \partial f(x_0)$, to $\forall_{y \in A}\ f(y) \geq f(x_0) + \langle 0, y - x_0 \rangle$
$$= f(x_0).$$
czyli $f(x_0) = \min_A f$

Jeżeli $0 \notin \partial f(x_0)$, to $\sim \left(\forall_{y \in A}\ f(y) \geq f(x_0) + \langle 0, y - x_0 \rangle \right)$
$$\Downarrow$$
$$\exists_{y \in A}\ f(y) < f(x_0)$$
czyli $f(x_0) \neq \min_A f$. $\square$.

A co, jeżeli A nie jest otwarty?

Tu sprawę komplikuje trochę fakt, że to, jaka jest subróżniczka f w x_0 zależy od dziedziny f: niech $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1$. Wtedy $\partial f(1) = \{0\}$.

Rozważmy jednak $g = f|_{[-1,1]}$. Łatwo można sprawdzić, że $\partial g(1) = [0, \infty)$.

Dlatego następne twierdzenie występuje w dwóch wersjach; o tej samej treści, ale różnych dowodach.

<u>Twierdzenie 1</u>

Niech $A \subset \mathbb{R}^n$ będzie wypukły, $f : A \to \mathbb{R}$ wypukła. Dla $x_0 \in A$, $f(x_0) = \min_A f$ wtedy i tylko wtedy, gdy $-\partial f(x_0) \cap N_A(x_0) \neq \emptyset$.

<u>Twierdzenie 2</u>

Niech $\mathcal{U} \subset \mathbb{R}^n$ będzie wypukły, $A \subset \operatorname{int} \mathcal{U}$ wypukły, $f : \mathcal{U} \to \mathbb{R}$ wypukła. Dla $x_0 \in A$, $f(x_0) = \min_A f$ wtedy i tylko wtedy, gdy $-\partial f(x_0) \cap N_A(x_0) \neq \emptyset$.

tylko tym razem subróżniczka jest wyznaczona dla $f : \mathcal{U} \to \mathbb{R}$, nie $f : A \to \mathbb{R}$, jest <u>więc mniejsza</u> (a w każdym razie niewiększa)

Dowód:

W pierwszą stronę dowód obu twierdzeń jest taki sam.

Załóżmy, że $-\partial f(x_0) \cap N_A(x_0) \neq \emptyset$, a więc istnieje $v \in \partial f(x_0)$ t.ż. $\forall_{y \in A} \ \langle y - x_0, v \rangle \geq 0$

Stąd $\forall_{y \in A} \ f(y) \underset{\substack{\uparrow \\ \text{bo } v \in \partial f(x_0)}}{\geq} f(x_0) + \underbrace{\langle y - x_0, v \rangle}_{\geqslant 0} \geqslant f(x_0)$,

czyli $f(x_0) = \min\limits_A f$.

W drugą stronę: załóżmy, że $f(x_0) = \min\limits_A f$

Tw. 1: Skoro $\forall_{y \in A} \ f(y) \geq f(x_0) = f(x_0) + \langle 0, y - x_0 \rangle$, to $0 \in \partial f(x_0)$, oczywiście $0 \in N_A(x_0)$, więc $0 \in -\partial f(x_0) \cap N_A(x_0)$.

Tw. 2: Tu znacznie więcej roboty. Zdefiniujmy

$$B_1 = \{(y - x_0, z) \in \mathbb{R}^n \times \mathbb{R} : y \in \mathcal{U}, \ z > f(y) - f(x_0)\}$$

$$B_2 = \{(y - x_0, z) \in \mathbb{R}^n \times \mathbb{R} : y \in A, \ z \leq 0\}$$

łatwo możemy sprawdzić, że B_1 i B_2 są wypukłe i, jeżeli $f(x_0) = \min\limits_A f$, rozłączne. Możemy je zatem rozdzielić hiperpłaszczyzną w $\mathbb{R}^n \times \mathbb{R}$: istnieje niezerowy wektor $(\xi, \mu) \in \mathbb{R}^n \times \mathbb{R}$ oraz stała $\alpha \in \mathbb{R}$ takie, że

(1) $\forall_{y \in \mathcal{U}} \ \forall_{z > f(y) - f(x_0)} \ \langle \xi, y - x_0 \rangle + z\mu \leq \alpha$

(2) $\forall_{y \in A} \ \forall_{z \leq 0} \ \langle \xi, y - x_0 \rangle + z\mu \geq \alpha$.

Weźmy w (2) $y = x_0$ i $z = 0 \Longrightarrow \alpha \le 0$

weźmy w (1) $y = x_0 \Longrightarrow \forall_{z>0} \quad z\mu \le \alpha \le 0$

ta nierówność może zachodzić tylko wtedy, gdy $\alpha = 0$, $\mu \le 0$.

Wykażemy, że $\alpha = 0$, $\mu < 0$. Gdyby bowiem $\mu = 0$, to

- wiedzielibyśmy, że $\xi \neq 0$, bo wektor (ξ, μ) jest niezerowy
- z (1) $\forall_{y \in \mathcal{U}} \quad \langle \xi, y - x_0 \rangle \le \alpha = 0$,

  co oznacza, że $\xi \in N_{\mathcal{U}}(x_0)$. Wiemy jednak, że $x_0 \in A \subset \operatorname{int} \mathcal{U}$, więc $N_{\mathcal{U}}(x_0) = \{0\}$. sprzeczność.

Stąd $\alpha = 0$, $\mu < 0$. Podzielmy nierówności (1) i (2) stronami przez $-\mu > 0$.

(1') $\forall_{y \in \mathcal{U}} \ \forall_{z > f(y) - f(x_0)} \quad z \ge \langle -\frac{\xi}{\mu}, y - x_0 \rangle$,

więc $\forall_{\varepsilon > 0} \quad f(y) - f(x_0) + \varepsilon \ge \langle -\frac{\xi}{\mu}, y - x_0 \rangle$

$$\downarrow \varepsilon \to 0$$

$$f(y) - f(x_0) \ge \langle -\frac{\xi}{\mu}, y - x_0 \rangle$$

czyli $-\frac{\xi}{\mu} \in \partial f(x_0)$

(2') $\forall_{y \in A} \ \forall_{z \le 0} \quad \langle -\frac{\xi}{\mu}, y - x_0 \rangle \ge z$, biorąc $z = 0$

$$\forall_{y \in A} \quad \langle \frac{\xi}{\mu}, y - x_0 \rangle \le 0$$

czyli $\xi/\mu \in N_A(x_0)$.

$\square$.

ale na pewno $\nabla g_i(x_0) \in N_A(x_0)$ dla $i \in I(x_0)$ i $\nabla h_j(x_0) \in N_A(x_0)$,
więc warunek Lagrange'a rzeczywiście daje $-\nabla f(x_0) \in N_A(x_0)$)

To twierdzenie można wzmocnić, osłabiając założenia na f.

<u>Zadanie</u>: Załóżmy, że zachodzą założenia (1)–(5), ~~zbiór~~ funkcje g_i, h_j spełniają w x_0 warunek Mangasariana-Fromovitza, a funkcja f jest quasiwypukła i $\nabla f(x_0) \neq 0$. Wówczas f ma w x_0 minimum lokalne na S.
Czy wynika stąd, że f ma w x_0 minimum globalne na S?

Obliczanie subróżniczek jest niestety, w ogólności, dość trudne.

- jeżeli $U \subset \mathbb{R}^n$ jest wypukły i otwarty, $g, h : U \to \mathbb{R}$ są wypukłe, to

$$\partial \max(g,h)(x) = \begin{cases} \partial g(x) & \text{gdy } g(x) > h(x) \\ \partial h(x) & \text{gdy } h(x) < g(x) \\ \operatorname{conv}(\partial h(x) \cup \partial g(x)) & \text{gdy } h(x) = g(x) \end{cases}$$

- tw. Rockafellara-Moreau: przy powyższych założeniach

$$\partial(g+h)(x) = \partial g(x) + \partial h(x)$$

W szczególności

<u>Wniosek</u>: Jeżeli f jest wypukła na A, różniczkowalna w $x_0 \in A$ i $-\nabla f(x_0) \in N_A(x_0)$, to $f(x_0) = \min_A f$.

<u>Wniosek</u>:

<u>Tw.</u> (warunki dostateczne KKT)

Załóżmy, że $U \subset \mathbb{R}^n$ jest otwarty,

(1) $f, g_i, h_j : U \to \mathbb{R}$ dla $i = 1, \dots, k$, $j = 1, \dots, l$
są różniczkowalne,

(2) $S = \{x \in U : g_i(x) \leq 0,\ h_j(x) = 0 \quad i = 1, \dots, k,\ j = 1, \dots, l\}$,

(3) ~~spełni~~ w $x_0 \in S$ spełniony jest warunek Lagrange'a,
tzn. istnieją $\lambda_1, \dots, \lambda_k \geq 0$, $\mu_1, \dots, \mu_l \in \mathbb{R}$ tż.

$$\begin{cases} \nabla f(x_0) = -\sum_{i=1}^{k} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{l} \mu_j \nabla h_j(x_0), \\ \lambda_i g_i(x_0) = 0 \quad \text{dla } i = 1, \dots, k \end{cases}$$

<u>i dodatkowo</u>: (4) funkcje h_j są afiniczne
(5) funkcje g_i są quasiwypukłe

i (6) funkcja f jest wypukła

to $f(x_0) = \min_S f$.

<u>Dowód</u>: Przy powyższych założeniach S jest wypukły, a warunek Lagrange'a gwarantuje, że $-\nabla f(x_0) \in N_A(x_0)$

(bez warunków jakości więzów nie mamy pewności, czy cały stożek $N_A(x_0)$ jest postaci $\{\sum_{i \in I(x_0)} \lambda_i' \nabla g_i(x_0) + \sum_j \mu_j' \nabla h_j(x_0)\}$, Zadanie: Udowodnić twierdzenie Moreau-Rockafellara.

Szkic / wskazówka: niech $x_0 \in U$.

1. Zawieranie $\partial f(x_0) + \partial g(x_0) \subset \partial(f+g)(x_0)$ jest trywialne.
2. Niech $\xi \in \partial(f+g)(x_0)$.

Rozważmy zbiory $A = \{(x-x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \langle \xi, x-x_0 \rangle\}$

$B = \{(x-x_0, \eta) \in \mathbb{R}^n \times \mathbb{R} : -y \geq g(x) - g(x_0)\}$

- wykazać, że A i B są wypukłe i niepuste
- zauważyć, że $\xi \in \partial(f+g)(x_0) \Longrightarrow A \cap B = \emptyset$
- rozdzielić A i B hiperpłaszczyzną i dalej postępować jak w dowodzie Twierdzenia 2 z poprzedniego wykładu. ∎ (5 stron wcześniej).