13.11.17 МОМО $\overline{\underline{X}}$

$\{x_i, y_i\}_{i=1}^N$ Негладкие функции

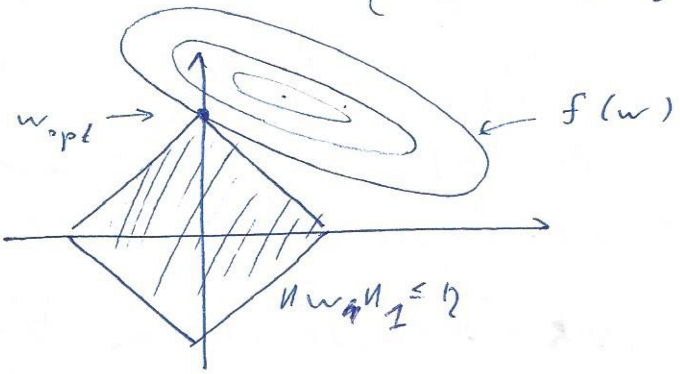
$$F(w) = \underbrace{\frac{1}{N}\sum_{i=1}^N L(y_i, \langle w, x_i\rangle)}_{f(w)} + \lambda \|w\|_1 \to \min_w$$

$$\begin{cases} f(w) \to \min_w \\ \|w\|_1 \le \eta \end{cases}$$

ККТ: $\exists (\hat{w}, \hat{\lambda}):$
$$\begin{cases} \hat{w} = \arg\min_w f(w) + \lambda(\|\hat{w}\|_1 - \eta) \\ \|\hat{w}\|_1 \le \eta \\ \hat{\lambda} \ge 0 \\ \hat{\lambda}(\|\hat{w}\|_1 - \eta) = 0 \end{cases}$$

$\eta = \|w_{opt}\|_1$

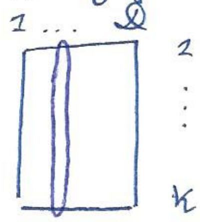
λ обр. пр. η



Групповая разреженность

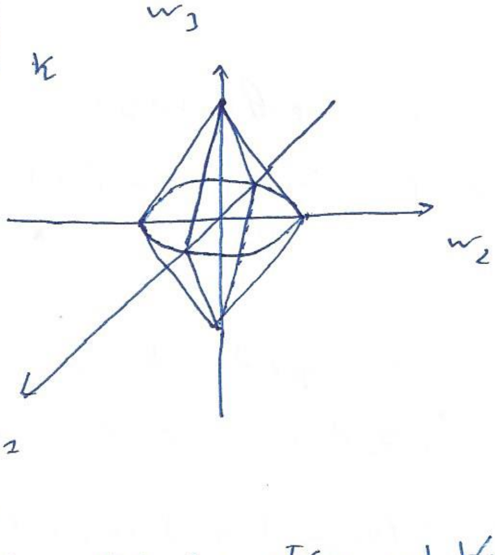
$\{1 \dots D\} = \bigsqcup_{g=1}^{G} A_g$, $w = [w_1 \dots w_G]$

$y(x) = \arg\max_{k \in \{1\dots K\}} w_k^T x$, W:

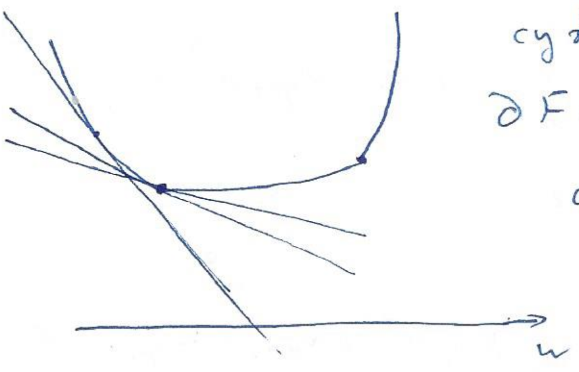


$\|w\|_{1/2} = \sum_{g=1}^{G} \|w_g\|_2$

пример: $\sqrt{w_1^2 + w_2^2} + |w_3| \le \eta$



$F(w) \to \min_w$, F - вып. $\in C \notin C^2$

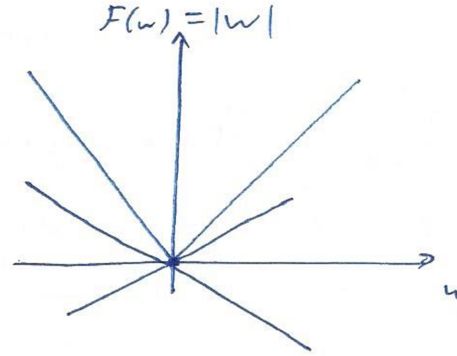


субдифференциал

$\partial F(w) = \{z \mid F(v) \geqslant F(w) + z^T(v - w) \ \forall v\}$

субградиент $z \in \partial F(w)$

утв. $\partial F(w)$ выпукло



$z_1, z_2 \in \partial F(w) \Rightarrow$

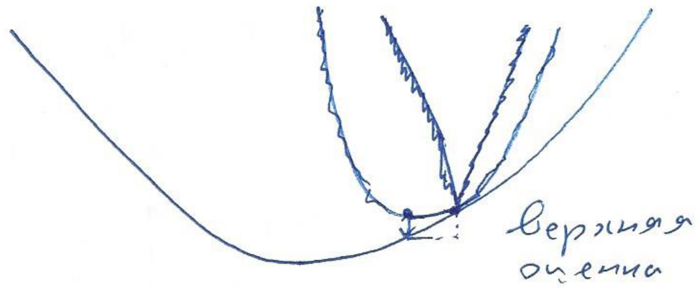
$F(v) \geq F(w) + z_1^T(v-w)$

$F(v) \geq F(w) + z_2^T(v-w)$

$$F(v) \overset{\forall \alpha \in [0,1]}{=} \alpha F(v) + (1-\alpha) F(v) \geq \alpha\,(F(w) + z_1^T(v-w)) +$$

$$+ (1-\alpha)(F(w) + z_2^T(v-w)) = F(w) + (\alpha z_1 + (1-\alpha) z_2)^T (v-w) \Rightarrow$$

$$\Rightarrow \alpha z_1 + (1-\alpha) z_2 \in \partial F(w)$$



пример: $F(w) = |w|, \ w \in \mathbb{R}$

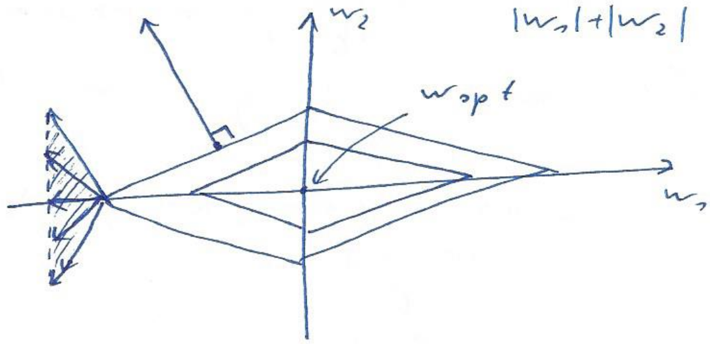
$$\partial F(w) = \begin{cases} w > 0, & \{1\} \\ w < 0, & \{-1\} \\ w = 0, & [-1, 1] \end{cases}$$

утв. w_0 - реш. $F(w) \to \min\limits_w \Leftrightarrow 0 \in \partial F(w_0)$

$\square$ $0 \in \partial F(w_0) \Leftrightarrow F(v) \geq F(w_0) + 0^T(v - w_0) \ \forall v \Leftrightarrow F(v) \geq F(w_0)$

Субградиентный метод

$$\begin{cases} w_{k+1} = w_k - \alpha_k z_k \\ z_k \in \partial F(w_k) \end{cases}$$



Липшицевость

$$F \in C_L^{0,0} \Leftrightarrow |F(v) - F(w)| \leq L \|v - w\| \quad \forall v, w$$

$$\Leftrightarrow z \in \partial F(w), \ \|z\| \leq L$$

$$\|w_{k+1} - w_{opt}\|^2 = \|w_k - \alpha_k z_k - w_{opt}\|^2 = \|w_k - w_{opt}\|^2 - 2\alpha_k z_k^T (w_k - w_{opt}) +$$

$$+ \alpha_k^2 \|z_k\|^2 = \ldots = \|w_0 - w_{opt}\|^2 - 2\sum_{i=0}^{k} \alpha_i \langle z_i, w_i - w_{opt} \rangle + \sum_{i=0}^{k} \alpha_i^2 \|z_i\|^2 \leq$$

$$\leq \left\{ \begin{array}{l} F(w_{opt}) \geq F(w_i) + \langle z_i, w_{opt} - w_i \rangle \\ \langle z_i, w_i - w_{opt} \rangle \geq F(w_i) - F(w_{opt}) \end{array} \right\} \leq \|w_0 - w_{opt}\|^2 -$$

$$- 2\sum_{i=0}^{k} \alpha_i (F(w_i) - F_{opt}) + \sum_{i=0}^{k} \alpha_i^2 \|z_i\|^2$$

$$2\sum_{i=0}^{k} \alpha_i (F(w_i) - F_{opt}) \leq \|w_0 - w_{opt}\|^2 + \sum_{i=0}^{k} \alpha_i^2 \|z_i\|^2 - \|w_{k+1} - w_{opt}\|^2 \leq$$

$$\leq \|w_0 - w_{opt}\|^2 + \sum_{i=0}^{k} \alpha_i^2 \|z_i\|^2$$

$$\bar{w}_k = \underset{0 \le i \le k}{\operatorname{argmin}} \; F(w_i)$$

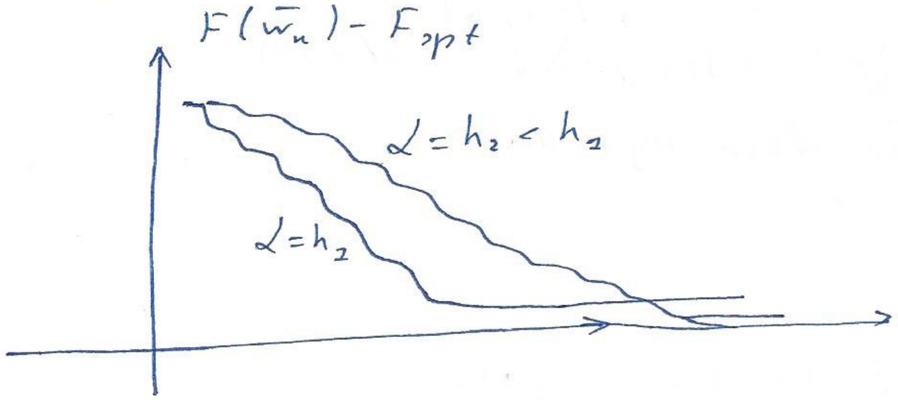
$$2\left(\sum_{i=0}^{k} \alpha_i\right)\left(F(\bar{w}_k) - F_{opt}\right) \le 2\left(\sum_{i=0}^{k} \alpha_i\right)\left(F(w_i) - F_{opt}\right)$$

$$F(\bar{w}_k) - F_{opt} \le \frac{\|w_0 - w_{opt}\|^2 + \sum_{i=0}^{k} \alpha_i^2 \|z_i\|^2}{2\left(\sum_{i=0}^{k} \alpha_i\right)}$$

(1) $\alpha_i = h$

$$\frac{R^2 + (k+1) h^2 L^2}{2(k+1)h} = \frac{R^2}{2h(k+1)} + \frac{hL^2}{2} \xrightarrow[k \to \infty]{} \frac{hL^2}{2}$$

сходится в окрестность оптимального решения.



(2) $\alpha_i = \frac{h}{\|z_i\|} \ge \frac{h}{L}$; $\frac{R^2 + h^2(k+1)}{2(k+1)(h/L)} = \frac{LR^2}{2h(k+1)} + \frac{Lh}{2} \xrightarrow[k \to \infty]{} \frac{hL}{2}$

(3) $\sum_{i=0}^{\infty} \alpha_i = \infty$, $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$

$$\frac{R^2 + L^2\left(\sum_{i=0}^{\infty} \alpha_i^2\right)}{2\left(\sum_{i=0}^{k} \alpha_i\right)} \xrightarrow[k \to \infty]{} 0, \quad \alpha_i = \frac{h}{(i+1)^\tau}, \quad \tau \in \left(\tfrac{1}{2}, 1\right]$$

$\tau = 1$, $\alpha_i = \frac{h}{i+1}$. $\frac{R^2 + L^2 \cdot C}{2 \ln k} = \bar{O}(1/\ln k)$

$$\sum_{i=0}^{k} \frac{1}{(i+1)^\tau} \approx \int_1^k \frac{1}{x^\tau}\, dx$$

$\tau \in (0, \tfrac{1}{2}]$, $\exists\, \tau = \tfrac{1}{2}$

$\alpha_i = \frac{h}{\sqrt{i+1}}$

(4) $\sum_{i=0}^{\infty} \alpha_i = \infty$, $\frac{\sum_{i=0}^{\infty} \alpha_i^2}{\sum_{i=0}^{\infty} \alpha_i} \to 0$ $\quad \frac{R^2 + h^2 L^2 \ln k}{h\sqrt{k}} = \tilde{O}\left(\frac{1}{\sqrt{k}}\right)$

Метод	Ф-ия	Ск. сх-ти
Град. спуск	вып. $f \in C_2^{1,1}$	$O(1/k)$, $O(1/\varepsilon)$
Субгр. метод	вып. $f \in C_2^{0,0}$	$O(1/\sqrt{k})$, $O(1/\varepsilon^2)$

субдифференциал lasso

$$F(w) = \underbrace{\frac{1}{N} \sum_{i=1}^{N} L(y_i, \langle w, x_i \rangle)}_{f(w)} + \lambda \|w\|_1 \to \min_w$$

$$(\partial F(w))_d = (\nabla f(w))_d + \begin{cases} w_d > 0, & \{\lambda\} \\ w_d < 0, & \{-\lambda\} \\ w_d = 0, & [-\lambda, \lambda] \end{cases}$$

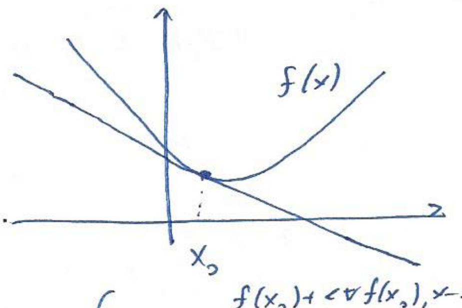
$0 \in \partial F(0) \Leftrightarrow 0 \in [\nabla f(0)_d - \lambda, \nabla f(0)_d + \lambda] \quad \forall d = \overline{1, D}$

$\lambda \geq \|\nabla f(0)\|_\infty \Rightarrow$ все веса нулевые.

семинар

$f(X) = \langle X, uv^T \rangle$, $\nabla f(x) = \frac{1}{2} uv^T + \frac{1}{2} vu^T$

субдифференциальное исчисление



Опр для вып. C^1

$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle \quad \forall x, x_0$

глобальная нижняя оценка ф-ии

$f: E \to \mathbb{R}$, $x_0 \in E$, вектор $g_0 \in V$ наз-ся субградиентом f в т. x_0,

если $f(x) \geq f(x_0) + \langle g_0, x - x_0 \rangle \quad \forall x \in E$

субдиф-л $\partial f(x_0)$ - мн-во всевозможных субгр-в в т. x_0

1) Норма $f: V \to \mathbb{R}$ $f(x) = \|x\|$ - произвольная норма

а) $\partial f(0) = ?$ б) $\partial f(x) = ?$

сопряжённая норма $\|s\|_* = \max\limits_{\|x\| = 1} |\langle s, x \rangle|$

$\forall s, x \quad |\langle s, x \rangle| \leq \|s\|_* \cdot \|x\|$

а) $\|x\| \geq \|0\| + \langle g_0, x\rangle$, $\|x\| \geq \langle g_0, x\rangle \quad \forall x \quad \|x\|=1$

$\Longleftrightarrow \max\limits_{\|x\|=1} \|x\| \geq \max\limits_{\|x\|=1} \langle g_0, x\rangle = \|g_0\|_*$

$\partial f(0) = \{g_0 : \|g_0\|_* \leq 1\} = \overline{B}_{\|\|_*}(0, 1)$

б) $\|x\| \geq \langle g_0, x - x_0\rangle + \|x_0\| \quad \forall\, x \in E \quad \cancel{\|x\| = 1}$

$\cancel{1 \geq \max\limits_{\|x\|=1} \langle g_0, x\rangle - \langle g, x_0\rangle}$,

$-\|x_0\| + \langle g_0, x_0\rangle \geq \langle g_0, x\rangle - \|x\| \quad \Big| \sup\limits_x$

$$\langle g_0, x_0\rangle - \|x_0\| \geq \sup_x (\langle g_0, x\rangle - \|x\|) = \begin{cases} 0, & \|g_0\|_* \leq 1 \\ \infty, & \text{иначе} \end{cases}$$

$\exists\ \|g_0\|_* \leq 1 : \quad \langle g_0, x_0\rangle \geq \|x_0\|$

$\langle g_0, x_0\rangle \leq \|g_0\|_* \|x_0\|$ $\qquad \partial f(x_0) = \begin{cases} ①\ \|g_0\|_* \leq 1 \\ ②\ \langle g_0, x_0\rangle = \|x_0\| \end{cases}$

$0 \leq \langle g_0, x_0\rangle - \|x_0\| \leq \|x_0\| \underbrace{(\|g_0\|_* - 1)}_{\leq 0}$

$\Downarrow = 0$

$\cancel{\langle g_0, x_0\rangle \leq}$ $\quad \|x\| \geq \overbrace{\langle g_0, x - x_0\rangle + \|x_0\|}^{=\|x_0\|}$

$\|x\| \geq \langle g_0, x\rangle$

$\langle g_0, x\rangle \leq \|g_0\|_* \|x\|_* \leq 1$

$f(x) = \|x\| = \langle x, x\rangle^{\frac{1}{2}}$ $\qquad$ ①,② $\underbrace{\langle g_0, x_0\rangle}_{=\|x_0\|} \overset{\text{КБ}}{\leq} \|g_0\| \|x_0\| \overset{①}{\leq} \|x_0\|$

$\nabla f(x) = \dfrac{x}{\|x\|}, \quad x \neq 0$ $\qquad$ $(=) \Rightarrow \boxed{g_0 = x_0}$

Свойства

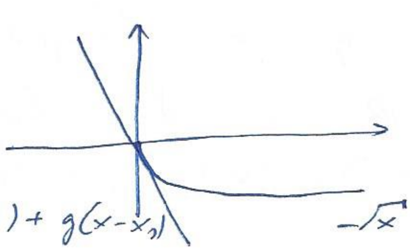
① $f: E \to \mathbb{R}$

$\partial f(x_0)$ выпуклое замкнутое мн-во

$\{g : f(x) \geq f(x_0) + \langle g, x - x_0\rangle \ \forall x\}$ с-ма линейных неравенств

пустой ∂f : $\quad f(x) = -\sqrt{x} \quad [0, \infty)$

$\partial f(0) = \{\}$



св-ва

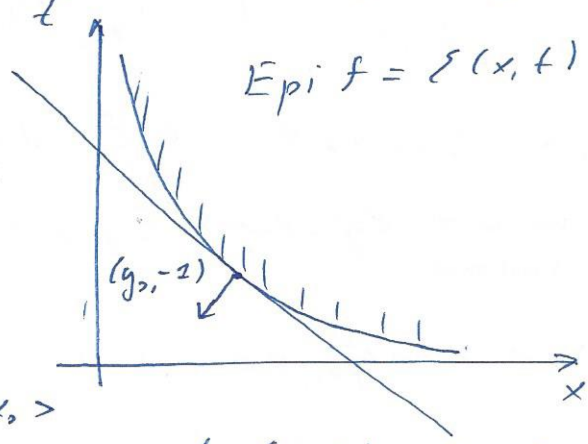
① $f: E \to \mathbb{R}$
$\partial f(x_0)$ – выпуклое замкнутое

② $f: E \to \mathbb{R}$ вып. ф-ия
$x_0 \in \operatorname{int} E \Rightarrow \partial f(x_0)$ – непустое выпуклое, замкнутое, огр.

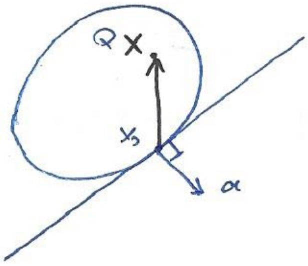
$\overline{B}_{\|\cdot\|_*}(0,1)$ вып. замк. огр.

$Epi\, f = \{(x,t) \in E \times \mathbb{R} : f(x) \le t\}$

касательная гиперпл-ть к надграфику ф-ии



$$t \ge f(x) \ge f(x_0) + \langle g, x - x_0 \rangle$$
$$\langle g, x - x_0 \rangle - (t - f(x_0)) \ge 0 \quad \forall\ (x,t) \in Epi\, f$$



$x_0 \in \partial Q$
$$\langle a, x - x_0 \rangle \underset{\cancel{\ge}}{\le} 0 \quad \forall x \in Q$$
$$\langle a, x \rangle \underset{\cancel{\ge}}{\le} \langle a, x_0 \rangle$$

① Пусть $\partial f(x_0) \ne \emptyset\ \forall x_0 \in E \Rightarrow f$ – выпуклая ф-ия

$f(x) \ge f(x_0) + \langle g_0, x - x_0 \rangle$ $\qquad f(x) = \max\limits_{z} \varphi(x,z)$

$\Downarrow$

$$f(\alpha x + (1-\alpha) y) \ge \alpha f(x) + (1-\alpha) f(y)$$

Правила преобразования, f – вып. ф-ия

$$\lim_{t \to 0+} \frac{f(x_0 + tv) - f(x_0)}{t} = f'(x_0, v) = \max_{g_0 \in \partial f(x_0)} \langle g_0, v \rangle$$

① $x_0 \in \operatorname{int} E$ f дифф-ма в т. $x_0 \Leftrightarrow \partial f(x_0) = \{\nabla f(x_0)\}$

② Аффинное преобразование

$f(x) = g(\underbrace{Lx + b}_{A(x)})$, $g: F \to \mathbb{R}$ вып.

$x_0 \in \operatorname{int} A^{-1}(F) \Rightarrow \partial f(x_0) = L^* \partial g(Lx_0 + b)$

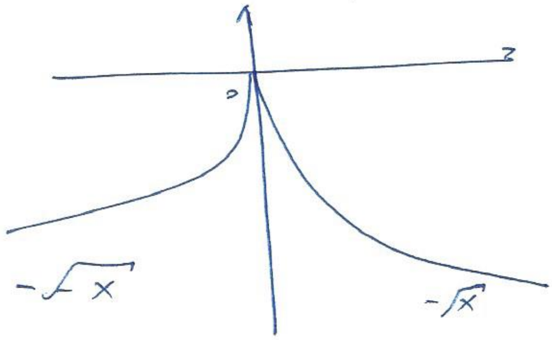
③ Сумма ф-ий, Моро-Рокафеллара

$f_1: E_1 \to \mathbb{R}, \ldots, f_m: E_m \to \mathbb{R}$, $E = \bigcap\limits_{i=1}^{m} E_i$

$\partial f(x_0) = \partial f_1(x_0) + \ldots + \partial f_m(x_0)$ $\forall x_0 \in E$

Если $\operatorname{int} E = \bigcap_{i=1}^{m} \operatorname{int} E_i \neq \emptyset$, то

$\partial f(x) = \sum_{i=1}^{m} \partial f_i(x) \quad \forall x \in E$



$f : \{0\} \to \mathbb{R}$

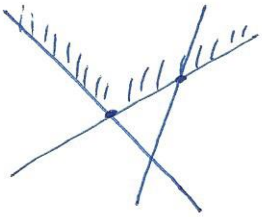
$f(x) = 0 \qquad \partial(-\sqrt{x}) = \emptyset$

$\partial f(0) = \mathbb{R} \qquad \partial(-\sqrt{x}) = \emptyset + = \mathbb{R}$

(4) $f(x) = \max\limits_{1 \le i \le m} f_i(x)$

(4.1)

$x_0 \in \operatorname{int} E \Rightarrow \partial f(x_0) = \operatorname{conv} \bigcup\limits_{i \in I_0} \partial f(x_0)$



$I_0 = \{1 \le i \le m : f_i(x_0) = f(x_0)\}$

(4.2) $f(x) = \max\limits_{\alpha \in I} f_\alpha(x) \quad \{\sup\limits_{\alpha \in I} f_\alpha(x)\}$

произвольное мн-во

$\partial f(x) \supseteq \operatorname{conv} \bigcup\limits_{\alpha \in I_0} \partial f_\alpha(x_0)$, I_0 – мн-во активных индексов

$I_0 = \{\alpha \in I : f_\alpha(x_0) = f(x_0)\}$

Если I – компакт и $\alpha \mapsto f_\alpha(x)$ непр. замкн. $\forall x$

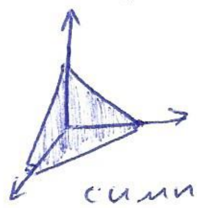
то $\partial f(x_0) = \operatorname{conv} \bigcup\limits_{i \in I_0} \partial f_\alpha(x_0)$ Теорема Данскина

Дубовицкого-Милютина

<u>пример</u>

$f(x) = \max\limits_{1 \le i \le n} \underset{f_i(x)}{x_i}$, $f : \mathbb{R}^n \to \mathbb{R}$, $\nabla f_i(x) = e_i = (0 \dots 0, \underset{i}{1}, 0 \dots 0)$

i-я координата $\qquad x_i = \langle e_i, x \rangle$



$\partial f(x)$? $\quad I(x) = \{1 \le i \le m : f_i(x) = f(x)\}$

$\partial f(x) = \operatorname{conv} \bigcup\limits_{i \in I(x)} \{e_i\} = \operatorname{conv} \{e_i : i \in I(x)\}$

$\partial f(0) = ?$ $\quad I(0) = \{1 \dots n\}$ $\quad \partial f(0) = \operatorname{conv} \{e_1 \dots e_n\}$

(2) $f : S^n \to \mathbb{R}$

$f(x) = \lambda_{max}(X) = \max\limits_{\|v\| = 1} \underbrace{\langle Xv, v \rangle}_{f_v(x)}$, $\partial f(x)$?

$I(x) = \{v \in \mathbb{R}^n : \|v\| = 1, v - \text{с. в. отв. } \lambda_{max}\}$

$\nabla f_v(x) = vv^T$, $\partial f(x) = \operatorname{conv} \left\{ vv^T : v \in \mathbb{R}^n, v - \text{с. в. } \lambda_{max}(x), \|v\| = 1 \right\}$

$\partial f(0) = \operatorname{conv} \{vv^T : \|v\| = 1\} = \{G \in S^n_+, \operatorname{tr} G = 1\}$

$\operatorname{tr} \left(\sum_i \lambda_i v_i v_i^T \right) = \sum_i \lambda_i \|v_i\|^2 = 1$