16. 10. 17 МОМО VII

$$\begin{cases} f(x) \to \min\limits_x \\ g_i(x) \le 0, \ i=\overline{1,m} \\ h_j(x)=0, \ j=\overline{1,p}\end{cases}$$

алгебраическое описание неоднозначное

$h_j(x)=0 \iff \hat{h}_j(x) = h_j^2(x)=0$

задача условной оптимизации

Пример: SVM

$\{x_i, y_i\}_{i=1}^N$, $x_i \in \mathbb{R}^D$, $y_i \in \{-1, 1\}$

$\hat{y}(x) = \mathrm{sgn}(w^T x + b)$

гладкая задача условной минимизации

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^N \xi_i \to \min\limits_{w,\xi,b} \\ y_i(w^T x_i + b) \ge 1 - \xi_i \quad \forall i \\ \xi_i \ge 0 \ \forall i \end{cases} \Rightarrow \begin{cases} \xi_i \ge 1 - y_i(w^T x_i + b) \\ \xi_i \ge 0 \end{cases}$$

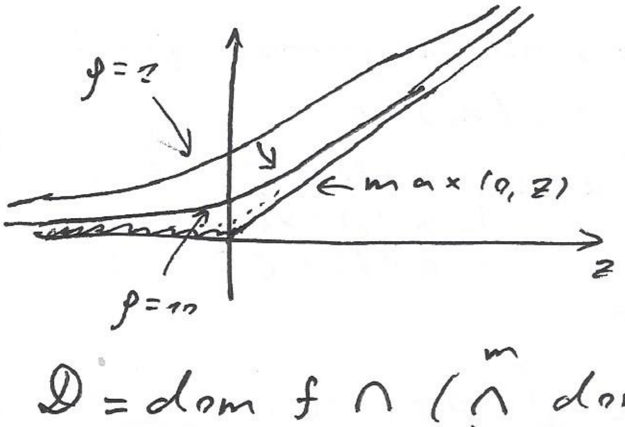
$\Rightarrow \xi_i \ge \max(0, 1 - y_i(w^T x_i + b))$

негладкая экв-я задача безусловной опт-и

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^N \max(0, 1 - y_i(w^T x_i + b)) \to \min_{w,b}$$

$$\max(z_1 \ldots z_n) \le \log\left(\sum_{i=1}^n \exp(z_i)\right)$$

$$\max(0, z) \overset{\rho>0}{=} \frac{1}{\rho}\max(0, \rho z) \le \frac{1}{\rho}\log(1 + \exp(\rho z))$$



$$\frac{1}{2}\|w\|^2 + \frac{C}{\rho}\sum_{i=1}^N \log(1 + \exp(\rho(1 - y_i(w^T x_i + b)))) \to \min_{w,b}$$

$$\mathcal{D} = \mathrm{dom}\, f \cap \left(\bigcap_{i=1}^m \mathrm{dom}\, g_i\right) \cap \left(\bigcap_{j=1}^p \mathrm{dom}\, h_j\right)$$

$$F = \{x \in \mathcal{D} \mid g_i(x) \le 0 \ \forall i, \ h_j(x) = 0 \ \forall j\}$$

$$Active(x) = \{1..p\} \cup \{i \mid g_i(x) = 0\}$$

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Т ККТ

$f, g_i, h_j \in C^1$, x_* – т. лок. экстр., выполнены условия рег-ти для $\{g_i(x_*), h_j(x_*) \mid i, j \in Active(x_*)\}$

тогда $\exists\, \lambda_*, \mu_*$:

1) $\nabla_x L(x_*, \lambda_*, \mu_*) = \nabla f(x_*) + \sum_{i=1}^{m} \lambda_{i,*} \nabla g_i(x_*) + \sum_{j=1}^{p} \mu_{j,*} \nabla h_j(x_*) = 0$

2) $x_* \in F$

3) $\lambda_{i,*} \geq 0 \quad \forall i$

4) $\lambda_{i,*}\, g_i(x_*) = 0 \quad \forall i$

в форме Джона:

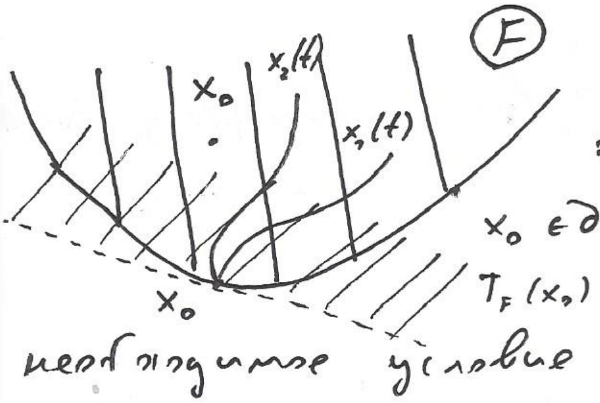
$$\tilde{L}(x, \lambda, \mu) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Т John

$f, g_i, h_j \in C^1$, x_* – лок. экстр. $\Rightarrow \exists\, (\lambda_*, \mu_*) \neq 0$:

выполнены 1), 2), 3), 4)

$\lambda_0 \geq 0$



$x_0 \in int\, F \Rightarrow$ необходимое усл-е экстремума $\nabla f(x_0) = 0$

$$x_0 \in \partial F, \quad S(x_0) = \left\{ x(t) \;\middle|\; \begin{array}{l} x(t) \in F \ \forall t \\ x(0) = x_0 \\ x(t) \in C^1 \end{array} \right\}$$

необходимое условие экстремума:

$$\frac{d}{dt} f(x(t))\Big|_{t=0} \geq 0, \quad \frac{d}{dt} f(x(t)) = \sum_{i=1}^{n} \frac{\partial f(x(t))}{\partial x_i(t)} \frac{d x_i(t)}{dt} =$$

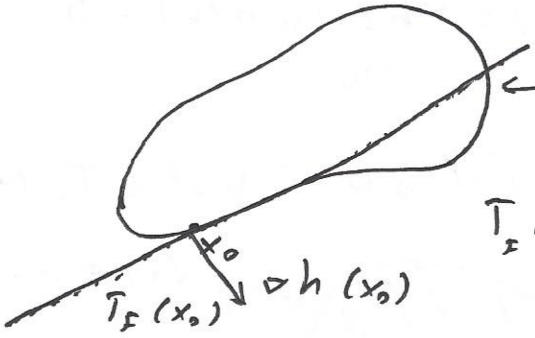
$$= \nabla_x f(x(t))^T \frac{dx}{dt}, \quad T_F(x_0) = \left\{ d \;\middle|\; \begin{array}{l} \exists\, x(t) \in S(x_0) \\ d = \dfrac{dx(t)}{dt}\Big|_{t=0} \end{array} \right\}$$

$\nabla f(x_0)^T d \geq 0 \quad \forall d \in T_F(x_0)$

необходимое условие экстремума

Случай 1

$$\begin{cases} f(x) \to \min\limits_x \\ h(x) = 0 \end{cases}$$



$T_F(x_0) = \{d \mid \nabla h(x_0)^T d = 0\}$

$\nabla f(x_0)^T d \geq 0 \quad \forall d \in T_F(x_0) \Leftrightarrow \nabla f(x_0) \parallel \nabla h(x_0)$

$\Leftrightarrow \exists \mu : \nabla f(x_0) = -\mu \nabla h(x_0)$

Два случая:

① $\nabla f(x_0) = 0$, $\nabla f(x_0)^T d = 0$

для $\mu = 0$ $\nabla f(x_0) = -\mu \nabla h(x_0)$

② $\nabla h(x_0) = 0$, невозможно описать $T_F(x_0)$ через $\nabla h(x_0)$

условия регулярности запрещают эту ситуацию

constraint qualification, CQ

① LCQ

Все $g_i(x), h_j(x)$ яв-ся аффинными

② LICQ

$\{\nabla g_i(x), \nabla h_j(x)\}$ л.н.з.

③ Slater

для выпуклых задач оптимизации $\operatorname{int} F \neq \varnothing$

$\Leftrightarrow \exists \tilde{x} : h_j(\tilde{x}) = 0, \quad g_i(\tilde{x}) < 0$

④ weak Slater

для вып. з. $\exists \tilde{x} : h_j(\tilde{x}) = 0$, $g_i(\tilde{x}) \leq 0$ для аффинных $g_i(x)$

$g_i(\tilde{x}) < 0$ для остальных

$\lambda_0 \nabla f(x_0) + \mu \nabla \underset{\neq 0}{h(x_0)} = 0$ т Джона

$\lambda_0 = 0,\ \mu \neq 0 \implies \nabla h(x_0) = 0$

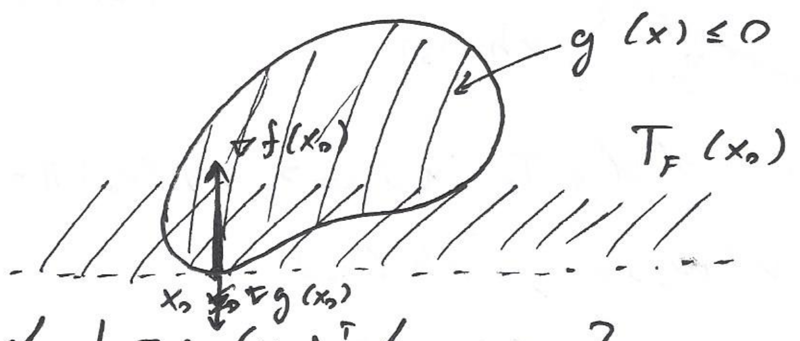
$\forall x_0$ виполнені необходимие условия

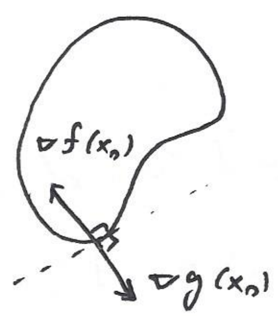
$\tilde{h}(x) = h^2(x), \quad \nabla \tilde{h}(x) = 2\underset{=0}{h(x)}\nabla h(x) = 0$

<u>Случай 2</u>

$\nabla f(x_0)^T d \geq 0 \quad \forall d \in T_F(x_0)$

$$\begin{cases} f(x) \to \min\limits_x \\ g(x) \leq 0 \end{cases}$$





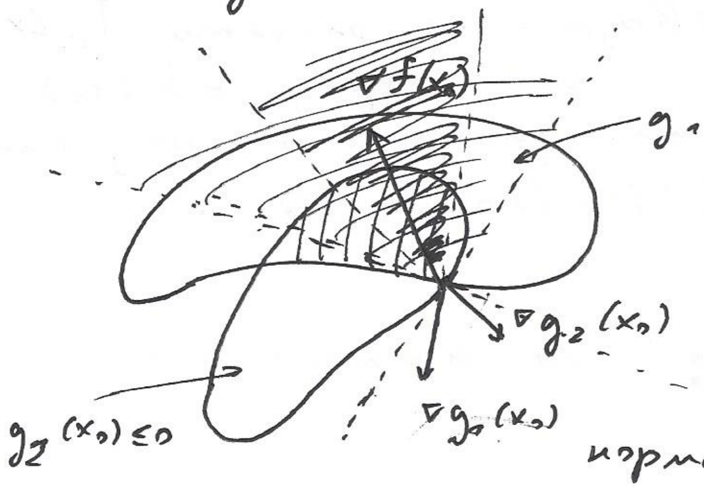
$T_F(x_0) = \{ d \mid \nabla g(x_0)^T d \leq 0 \}$

$\exists \lambda \geq 0: \nabla f(x_0) = -\lambda \nabla g(x_0)$

$$\lambda g(x_0) = 0 \iff \left[\begin{array}{ll} \lambda = 0 & (x_0 \in \operatorname{int} F) \\ g(x_0) = 0 & (x_0 \in \partial F) \end{array} \right.$$

<u>Случай 3</u>

$$\begin{cases} f(x) \to \min\limits_x \\ g_1(x) \leq 0 \\ g_2(x) \leq 0 \end{cases}$$



$\exists \lambda_1, \lambda_2 \geq 0$

$\nabla f(x_0) = -\lambda_1 g_1(x_0) - \lambda_2 \nabla g_2(x_0)$

нормальний конус

Обиций случай

$$T_F(x_0) = \left\{ d \;\middle|\; \begin{array}{ll} \nabla g_i(x_0) d \leq 0 & \forall i \\ \nabla h_j(x_0) d = 0 & \forall j \end{array} \right\}$$

$\nabla f(x_0)^T d \geq 0 \quad \forall d \in T_F(x_0) \iff \exists \lambda \geq 0, \mu:$

Farcas lemme

$\nabla f(x_0) = -\sum_i \lambda_i \nabla g_i(x_0) - \sum_j \mu_j \nabla h_j(x_0)$

Достаточные условия:

$$\nabla f(x_0)^T d = 0$$

$$\| $$

$$-\underbrace{\sum_i \lambda_i}_{\geq 0} \underbrace{\nabla g_i (x_0)^T d}_{\leq 0} - \sum_j \mu_j \underbrace{\nabla h_j (x_0)^T d}_{=0}$$

$$C(x_*, \lambda_*) = \{ d \in T_F (x_*) \mid \nabla g_i (x_*)^T d = 0 \text{ если } \lambda_{i*} > 0 \}$$

критический конус

Т ККТ

$f, g_i, h_j \in C^2$, x_* - лок. мин., вып. усл. рег-ти для $\{ g_i(x_*), h_j(x_*) \mid i, j \in Active(x_*) \}$

$\Rightarrow \lambda_*, \mu_*$: 1), 2), 3), 4)

5) $d^T \nabla^2_{xx} L(x, \lambda, \mu) d \geq 0 \quad \forall d \in C(x_*, \lambda_*)$

Если для (x_*, λ_*, μ_*) вып. усл-я рег-ти 1), 2), 3), 4) и 5) $d^T \nabla^2_{xx} L(x_*, \lambda_*, \mu_*) d > 0 \quad \forall d \in C(x_*, \lambda_*)$

Тогда x_* - лок. минимум

семинар

GD $\quad T(\varepsilon) \approx \varkappa \ln(\frac{1}{\varepsilon})$ для квадратичной

CG $\quad T(\varepsilon) \approx \sqrt{\varkappa} \ln(\frac{1}{\varepsilon})$ ф-ии

$$\begin{cases} \min\limits_{x \in V} f(x) \\ s.t. \ Ax = b \end{cases} \underline{(p)} \qquad f: V \to \mathbb{R}, \quad A: V \to V \text{ лин. оп.} \qquad \mathcal{L}(x, \lambda) = f(x) - \langle \lambda, Ax - b \rangle$$

$$\nabla f(x_0) - A^T \lambda = 0$$

Пусть x_0 - реш. (p). Тогда $\exists \lambda : \nabla f(x_0) = A^T \lambda$

$$\{ x \in V : Ax = b \} = x_0 + \underbrace{\{ x \in V : Ax = 0 \}}_{\ker A}$$

$$x_0 + \sum_{i=1}^m y_i u_i, \quad y \in \mathbb{R}^m \qquad u_1 \ldots u_m \text{ базис}$$

$y_i \in \mathbb{R}$, $u_i \in \mathbb{R}^m$

$U = [u_1 \dots u_m] \in \mathbb{R}^{m \times m}$

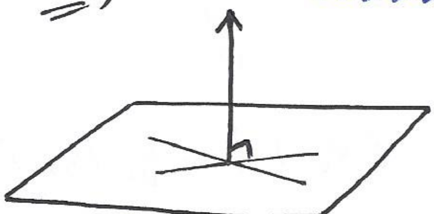
$f(x_0 + Uy) = g(y)$, $\nabla g(y) = U^T \nabla f(x_0 + Uy)$

верно для $\forall y$, тогда выберем $y = 0$

$\Rightarrow \underbrace{\{U^T \nabla f(x_0) = 0\}}$, $\langle u_i, \nabla f(x_0) \rangle = 0$

$\nabla g(y) = 0 \Rightarrow$ $\quad 1 \le i \le m$

Экстремум, градиент равен нулю



$\cancel{\nabla g(y) = U^T \nabla f(x_0 + Uy)}$

$\nabla f(x_0) \in (\ker A)^\perp = \operatorname{im} A^* = \{A^* \lambda : \lambda \in V\}$

$A : V \to V$

① $x \in \operatorname{im} A^* \Leftrightarrow x = A^* \lambda$

$\forall y \in \ker A$, $\langle x, y \rangle = \langle A^* \lambda, y \rangle = \langle \lambda, \underset{0}{\underbrace{Ay}} \rangle = 0$

тогда $x \in (\ker A)^\perp$

$\operatorname{im} A^* \subseteq (\ker A)^\perp$

② $y \in (\ker A)^\perp$

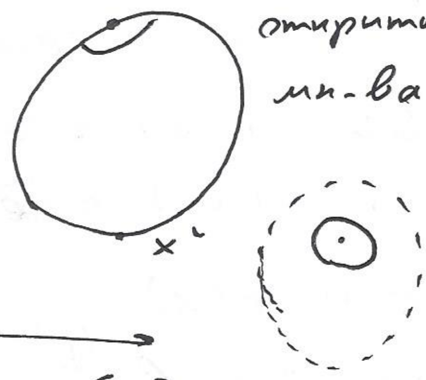
П $\exists\, y \notin \operatorname{im} A^* \Rightarrow y \in (\operatorname{im} A^*)^\perp \Rightarrow$

Возьмём X из $\operatorname{im} A^*$: $\langle A^* x, y \rangle = 0$

$\langle A^* \lambda, y \rangle = \langle \lambda, Ay \rangle = 0 \;\; \forall \lambda \Rightarrow Ay = 0 \Rightarrow y \in \ker A$

правило множителей Лагранжа

открытые мн-ва



$$\begin{cases} \min f(x) \\ x \in E \text{ открытое мн-во} \quad (p) \\ \text{s.t. } h_j(x) = 0, \quad 1 \le j \le k \end{cases}$$

$\min\limits_{x \ge 0} x = 0$

Пусть x_0 - реш. (p) и пусть $\nabla h_1(x_0), \dots, \nabla h_k(x_0)$ л.н.з.

$\min\limits_{x > 0} = \{\varnothing\}$

Тогда $\exists \lambda : \nabla f(x_0) = \sum\limits_{i=1}^{k} \lambda_i \nabla h_j(x_0) = 0$

Т ККТ

$\min\limits_{x \in E} f(x)$, s.t. $g_i(x) \le 0 \quad 1 \le i \le m$, $\lambda_1 \dots \lambda_m \ge 0$

$(p) \qquad h_j(x) = 0 \quad 1 \le j \le k$, $\mu_1 \dots \mu_k \in \mathbb{R}$

Пусть x_0 - реш. (p), и пусть выполнены усл-я рег. Тогда $\exists \lambda_1 \dots \lambda_m \geq 0$, $\exists \mu_1 \dots \mu_k \in \mathbb{R}$:

① $\nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) - \sum_{j=1}^{k} \mu_j \nabla h_j(x_0) = 0$

② $\lambda_i \nabla g_i(x_0) = 0$ доп. неж.

ограничения либо внутри, либо на границе.

условия регулярности

⓪ g_i, h_j - аффинные

① лин. нез. гр. $\nabla h_1(x_0), \dots, \nabla h_k(x_0)$

$(\nabla g_i(x_0))$ $i \in A(x_0) = \{ i \in [1, m] . g_i(x_0) = 0 \}$

② E - выпуклое мн-во

$g_1 \dots g_m$ - вып., $h_1 \dots h_k$ - аффинные

\+ условие Слейтера: $\exists \bar{x} \in E$: $g_1(\bar{x}), \dots, g_m(\bar{x}) < 0$

Примеры:

проекция на евклидов шар

$\min_{x \in V} \| x - x_0 \|^2$ s.t. $\|x\|^2 \leq 1$

$\min_{x \in \mathbb{R}} x$, s.t. $x^2 \leq 0$ $\quad x^* = 0$ $\quad$ не выполнены условия рег-ти

$\mathcal{L}(x, \lambda) = x + \lambda x^2$, $\lambda \geq 0$

$\begin{cases} 1 + 2x\lambda = 0 \\ \lambda x^2 = 0 \end{cases} \Bigg|_{x_* = 0} \quad 1 = 0$

Достаточное условие на минимум

Если (p) - выпуклая и $\exists \lambda_1 \dots \lambda_m \geq 0, \mu_1 \dots \mu_k \in \mathbb{R}$:

условие стационарности ① и доп. нежёсткости ②

x_0 - глобальное решение задачи

$\min_{x \in V} \langle c, x \rangle$, s.t. $\langle a, x \rangle \leq b$

выпуклая задача, но решения может не быть

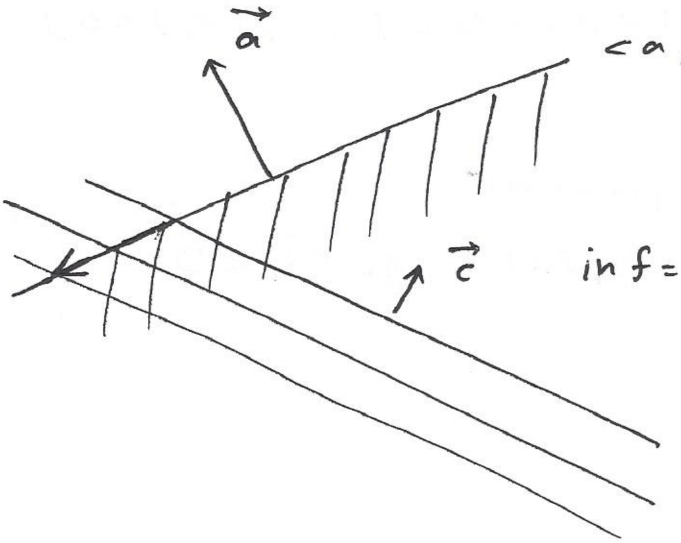
$$\mathcal{L}(x,\lambda) = \langle c, x\rangle + \lambda(\langle a,x\rangle - b)$$

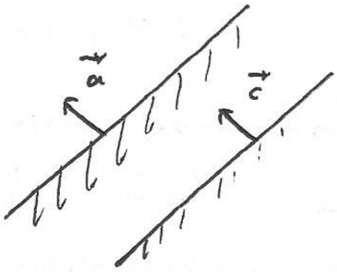
$$\begin{cases} \nabla \mathcal{L}(x,\lambda) = c + \lambda a = 0 \\ \lambda(\langle a,x\rangle - b) = 0 \end{cases}$$

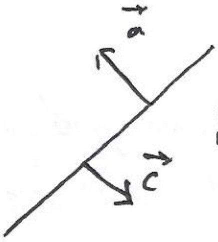
$\lambda = 0$, ~~$\langle a,x\rangle = b$~~, $c = 0$

$\lambda > 0$, $\langle a,x\rangle = b$, $c = -\lambda a$, $c \uparrow\downarrow a$

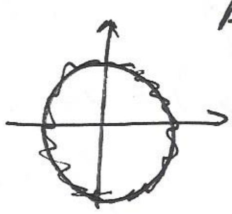
Ответ: $\begin{cases} c \uparrow\downarrow a \Rightarrow \{x : \langle a,x\rangle = b\} \\ \quad c \neq 0 \\ c = 0 \Rightarrow \forall x \\ \text{иначе } \varnothing \end{cases}$ выпуклость и достаточные условия







$\min\limits_{x \in \mathbb{R}^n} \|x\|^2$ s.t. $\langle Ax,x\rangle = 1$, $A \in S^n$ $\quad A = I$



$\nabla \|x\|^2 = 2x$, $\nabla \langle Ax,x\rangle = 2Ax \neq 0$

л.н.з. решения

$\mathcal{L}(x,\lambda) = \|x\|^2 - \lambda'(\langle Ax,x\rangle - 1)$

$$\begin{cases} \nabla_x \mathcal{L}(x,\lambda) = 2x - 2\lambda' Ax = 0, \\ \langle Ax,x\rangle = 1 \end{cases}, \quad \begin{cases} Ax = \frac{1}{\lambda'} x \\ \langle Ax,x\rangle = 1 \end{cases}, \quad \begin{cases} Ax = \lambda x \\ \langle Ax,x\rangle = 1 \end{cases}$$

$\lambda \|x\|^2 = 1$, $\|x\|^2 = \frac{1}{\lambda}$ Ответ: мн-во собств. в. x:

$\lambda_{с.зн.} = \max \lambda$, причём отнормирование: $\|x\|^2 = \frac{1}{\lambda}$