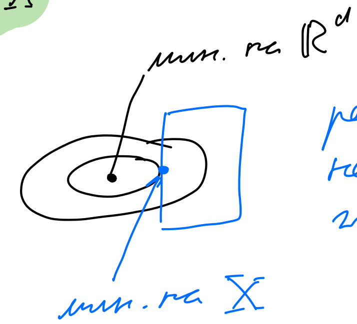
На пред. лекциях $\Longrightarrow$ Сегодня

$$\min_{x\in\mathbb{R}^d} f(x) \qquad \Longrightarrow \qquad \min_{x\in X} f(x)$$

X – „простое" мн-во



решение на X не совпадает с глоб. мин. на $\mathbb{R}^d$

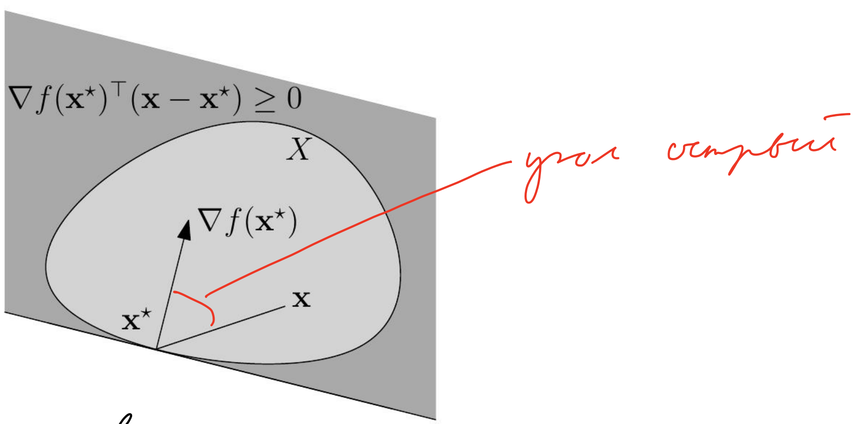
- **Условие оптимальности**
  - f – непрер. дифф. и выпукла на $\mathbb{R}^d$
  - X – выпуклое

$x^* \in X$ – глобальный минимум $\min_{x\in X} f(x)$

$$\Longleftrightarrow$$

$$\langle \nabla f(x^*) ; x - x^* \rangle \geq 0 \qquad \forall x \in X$$

Физический смысл:



Док-во:

- достаточность $\Longleftarrow$

$$\langle \nabla f(x^*) ; x - x^* \rangle \geq 0 \qquad \forall x \in X$$

выпуклость f:

$$f(x) \geq f(x^*) + \underbrace{\langle \nabla f(x^*) ; x - x^* \rangle}_{\geq 0} \geq f(x^*) \qquad \forall x \in X$$

x^* – глоб. минимум на $\overline{X}$

- необходимость $\Rightarrow$

x^* – глоб. минимум на $\overline{X}$

от противного: $\exists\, \tilde{x} \in \overline{X}: \langle \nabla f(x^*); \tilde{x} - x^* \rangle < 0$

$$\text{Рассм.} \quad \underbrace{\tilde{x}_\lambda}_{\in \overline{X}} = \lambda \tilde{x} + (1-\lambda) x^* \qquad \lambda \in [0;1]$$

$$\varphi(\lambda) = f(\tilde{x}_\lambda) = f(\lambda \tilde{x} + (1-\lambda) x^*)$$

$$\frac{d\varphi}{d\lambda} = \frac{d}{d\lambda}\Big(f(\lambda(\tilde{x} - x^*) + x^*)\Big) = \langle \nabla f(\lambda(\tilde{x} - x^*) + x^*); \tilde{x} - x^* \rangle$$

$$\left.\frac{d\varphi}{d\lambda}\right|_{\lambda=0} = \langle \nabla f(x^*); \tilde{x} - x^* \rangle < 0 \text{ по предположению}$$

φ убывает в окр. 0, а значит $\exists\, \hat{\lambda} > 0$:

$$f(\underbrace{x^* + \hat{\lambda}(\tilde{x} - x^*)}_{\neq x^*}) = \varphi(\hat{\lambda}) < \varphi(0) = f(x^*)$$

противоречие с x^* – глоб. минимум

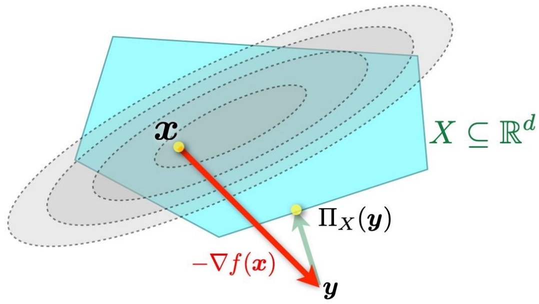
■

- Метод град. спуска с проекцией

$$\underbrace{x^{k+1}}_{\text{не факт } \in \overline{X}} = \underbrace{x^k}_{\in \overline{X}} - \gamma \nabla f(x^k)$$

$$x^{k+1} = \Pi\left(x^k - \gamma \nabla f(x^k)\right)$$

$$\Pi(y) = \arg\min_{x \in X} \|x-y\|_2^2$$ ← евклидова проекция



Св-ва проекции:

1) X – выпуклое, $x \in X$, $y \in \mathbb{R}^d$, тогда

$$\langle x - \Pi(y); y - \Pi(y) \rangle \le 0$$

Док-во: $\Pi(y) = \arg\min_{\underbrace{z \in X}_{\text{выпуклое}}} d(z)$ $\qquad \underbrace{d(z) = \|z-y\|_2^2}_{\text{выпуклая}}$

Условие оптимальности для d, X

$$\langle \nabla d(z^*); z - z^* \rangle \ge 0 \qquad \forall z \in X$$

$z^* = \Pi(y) \qquad z = x$

$$\langle \nabla d(\Pi(y)); x - \Pi(y) \rangle \ge 0$$

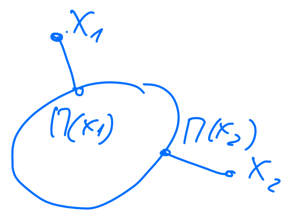
$$\nabla d(z) = 2(z-y)$$

$$2\langle \Pi(y) - y; x - \Pi(y) \rangle \ge 0 \quad \blacksquare$$

2) Нерастяжимость оператора проекции:

$\mathcal{X}$ — выпуклое, $x_1, x_2 \in \mathbb{R}^d$, тогда

$$\|\Pi(x_1) - \Pi(x_2)\|_2 \le \|x_1 - x_2\|_2$$



Док-во: 1) св-во $y = x_1$ $\quad x = \Pi(x_2) \in \mathcal{X}$

$$\langle \Pi(x_2) - \Pi(x_1); x_1 - \Pi(x_1)\rangle \le 0$$

аналогично $y = x_2$, $x = \Pi(x_1)$

$$\langle \Pi(x_1) - \Pi(x_2); x_2 - \Pi(x_2)\rangle \le 0$$

сложим

$$\langle \Pi(x_2) - \Pi(x_1); x_1 - \Pi(x_1) - x_2 + \Pi(x_2)\rangle \le 0$$

$$\underbrace{\langle \Pi(x_2)-\Pi(x_1); \Pi(x_2)-\Pi(x_1)\rangle}_{\|\Pi(x_2)-\Pi(x_1)\|_2^2} \le \underbrace{\langle x_2-x_1; \Pi(x_2)-\Pi(x_1)\rangle}_{\text{КБШ}}$$

$$\|\Pi(x_2)-\Pi(x_1)\|_2^{\cancel{2}} \le \|x_2-x_1\|_2 \cdot \cancel{\|\Pi(x_2)-\Pi(x_1)\|_2}$$

$$\|\Pi(x_2)-\Pi(x_1)\|_2 \le \|x_2-x_1\|_2 \quad \blacksquare$$

3) Проекция однозначна

$\arg\min_{x\in\mathcal{X}} \|x-y\|_2^2 \leftarrow$ сильно выпукла

4) Стац. точка град. спуска с проекцией

$$x^* = \Pi(x^* - \gamma \nabla f(x^*))$$

Док-во:

$$\Pi(x^* - \gamma \nabla f(x^*)) = \arg\min_{x \in X} \left[\|x - x^* + \gamma \nabla f(x^*)\|_2^2 \right]$$

$$= \arg\min_{x \in X} \Big[\underbrace{\|x - x^*\|_2^2}_{\geq 0} + \underbrace{2\gamma \langle \nabla f(x^*); x - x^* \rangle}_{\geq 0 \text{ по условию оптимальности}} + \cancel{\gamma^2 \|\nabla f(x^*)\|_2^2} \Big]$$

≥ 0, 0 достигается $x = x^*$ ∎

Док-во сходимости:

$$\|x^{k+1} - x^*\|_2^2 = \|\Pi(x^k - \gamma \nabla f(x^k)) - x^*\|_2^2$$

4) свойство

$$= \|\Pi(x^k - \gamma \nabla f(x^k)) - \Pi(x^* - \gamma \nabla f(x^*))\|_2^2$$

2) св-во

$$\leq \|x^k - \gamma \nabla f(x^k) - x^* + \gamma \nabla f(x^*)\|_2^2$$

$$= \|x^k - x^*\|_2^2 - 2\gamma \langle \nabla f(x^k) - \nabla f(x^*); x^k - x^* \rangle + \gamma^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2$$

μ-сильная выпуклость для $\langle\ \rangle$ и L-гладкость для $\|\ \|_2^2$

$$\leq \|x^k - x^*\|_2^2 + 2\gamma \langle \nabla f(x^*); x^k - x^* \rangle - 2\gamma \left(\frac{\mu}{2} \|x^k - x^*\|_2^2 + f(x^k) - f(x^*) \right) + 2L\gamma^2 \left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^* \rangle \right)$$

$$= (1 - \gamma\mu) \|x^k - x^*\|_2^2 + 2\gamma(\gamma L - 1) \Big(\overbrace{\underbrace{f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^* \rangle}_{\text{выпуклость } \geq 0}}^{\text{дивергенция Брегмана}} \Big)$$

$\gamma \le \frac{1}{L}$

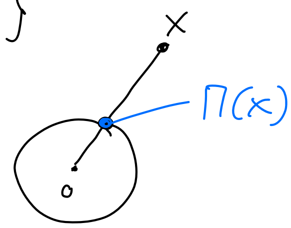
$$\le (1-\gamma\mu)\,\|x^k - x^*\|_2^2 \qquad \blacksquare$$

Та же сходимость, что и у GD.

Проекции:

1) L_2-шар $\quad X = \{x \in \mathbb{R}^d \mid \|x\|_2^2 \le 1\}$

$$\Pi(x) = \min\left\{1;\ \frac{1}{\|x\|_2}\right\} x$$



2) кубик $\quad X = \{x \in \mathbb{R}^d \mid a_i \le x_i \le b_i\}$

$$[\Pi(x)]_i = \begin{cases} a_i & x_i \le a_i \\ x_i & a_i \le x_i \le b_i \\ b_i & x_i \ge b_i \end{cases}$$

3) линейные ограничения $\quad X = \{x \in \mathbb{R}^d \mid Ax = b\}$

$$\Pi(x) = x - A^T (AA^T)^{-1}(Ax - b)$$

• Линейная задача (как альтернатива проекции / кв. задаче)

$$\min_{s \in X} \langle s; g \rangle \quad \text{для некоторого } g \in \mathbb{R}^d$$

← ограничено

1) L_1-шар

$$X = \{x \in \mathbb{R}^d \mid \|x\|_1 \le 1\}$$

$$i = \arg\max_j |g_j| \qquad \begin{pmatrix} 1 \\ 2 \\ -10 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ +1 \end{pmatrix}$$

$s^* = -\operatorname{sign}(g_i) e_i$ ← базисный вектор

2) симплекс $\quad X = \{x \in \mathbb{R}^d \mid \sum_{i=1}^d x_i = 1,\ x_i \ge 0\}$

$$s^* = e_i \qquad i = \arg\min_i g_i$$

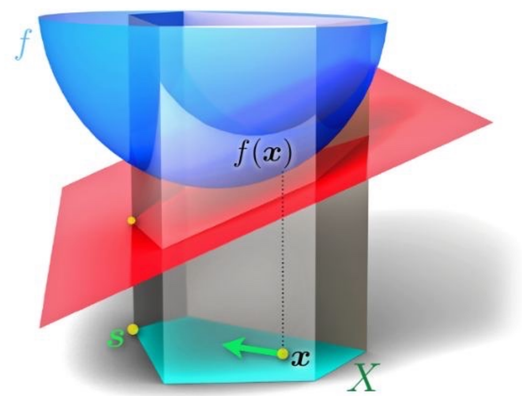
3) L_∞ -шар $\quad X = \{x \in \mathbb{R}^d \mid \|x\|_\infty \le 1\}$

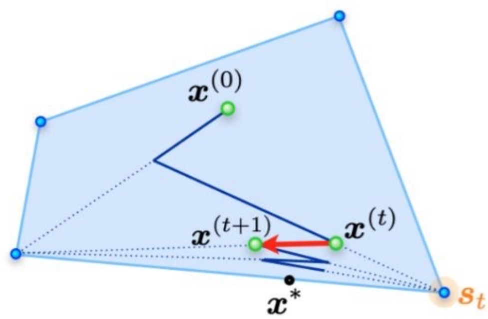
$$s^* = -\sum_{i=1}^{d} \operatorname{sign}(g_i)\, e_i$$

- **Метод Франк-Вульфа (условного градиента)**

$$\boxed{\begin{aligned} s^k &= \operatorname*{argmin}_{s\in X} \langle s; \nabla f(x^k)\rangle \\ x^{k+1} &= (1-\gamma_k)\,x^k + \gamma_k s^k \qquad \gamma_k = \frac{2}{k+2} \end{aligned}}$$

Рисунок:





- $s^k = \operatorname*{argmin}_{s\in X} \langle s; \nabla f(x^k)\rangle = \operatorname*{argmin}_{s\in X} \Big[\underbrace{f(x^k) + \langle \nabla f(x^k); s - x^k\rangle}_{\text{линейная аппроксимация}}\Big]$

находим т. на границе

- $x^{k+1} = \frac{k}{k+1} x^k + \frac{1}{k+1} s^k = \Big(1 - \underbrace{\frac{1}{k+1}}_{\frac{2}{k+2}}\Big) x^k + \underbrace{\frac{1}{k+1}}_{\frac{2}{k+2}} s^k$

усреднение точек с границы

Док-во сходимости:

L – гладкость

$$f(x^{k+1}) \le f(x^k) + \langle \nabla f(x^k); x^{k+1} - x^k \rangle + \frac{L}{2} \|x^k - x^{k+1}\|_2^2$$

$$x^{k+1} = x^k + \gamma_k (s^k - x^k)$$

$$f(x^{k+1}) \le f(x^k) + \gamma_k \langle \nabla f(x^k); s^k - x^k \rangle + \frac{L\gamma_k^2}{2} \|s^k - x^k\|_2^2$$

X – ограниченное, $D = \operatorname{diam} X = \max\limits_{x,y \in X} \|x - y\|_2$

$$f(x^{k+1}) \le f(x^k) + \gamma_k \langle \nabla f(x^k); s^k - x^k \rangle + \gamma_k^2 \cdot \frac{LD^2}{2}$$

$$\langle s^k; \nabla f(x^k) \rangle = \min_{s \in X} \langle s; \nabla f(x^k) \rangle \le \langle x^*; \nabla f(x^k) \rangle$$

$$f(x^{k+1}) \le f(x^k) + \gamma_k \langle \nabla f(x^k); x^* - x^k \rangle + \gamma_k^2 \cdot \frac{LD^2}{2}$$

$- f(x^*)$, выпуклость

$$f(x^{k+1}) - f(x^*) \le f(x^k) - f(x^*) - \gamma_k (f(x^k) - f(x^*)) + \gamma_k^2 \cdot \frac{LD^2}{2}$$

$$f(x^{k+1}) - f(x^*) \le (1 - \gamma_k)(f(x^k) - f(x^*)) + \gamma_k^2 \cdot \frac{LD^2}{2}$$

от k

По индукции: если $\gamma_k = \frac{2}{k+2}$, то

$$f(x^k) - f(x^*) \le \frac{2 \max\{LD^2; f(x^0) - f(x^*)\}}{k+2}$$

БИ: очевидно $k = 0$

ПИ: $f(x^{k+1}) - f(x^*) \le (1 - \gamma_k)(f(x^k) - f(x^*)) + \gamma_k^2 \cdot \frac{LD^2}{2}$

$$\overset{\gamma_k}{=} \frac{k}{k+2}(f(x^k) - f(x^*)) + \frac{\cancel{4}\,2}{(k+2)^2} \cdot \frac{LD^2}{\cancel{2}}$$

$$\overset{\text{предп.}}{\leq} \frac{k}{k+2} \cdot \frac{2\max\{\dots\}}{k+2} + \frac{2LD^2}{(k+2)^2}$$

$$\leq \left(\frac{k}{(k+2)^2} + \frac{1}{(k+2)^2} \right) 2\max\{\dots\}$$

$$\frac{k+1}{(k+2)^2} \leq \frac{1}{k+1+2} = \frac{1}{k+3}$$

$$\leq \frac{2\max\{\dots\}}{k+1+2} \quad \blacksquare$$

Итоги для ФВ:

- сублинейная сходимость для выпуклой задачи (как у GD)
- в сильно выпуклом случае улучшений нет
- как и метод проекции — все накладки для оптимизации на множествах