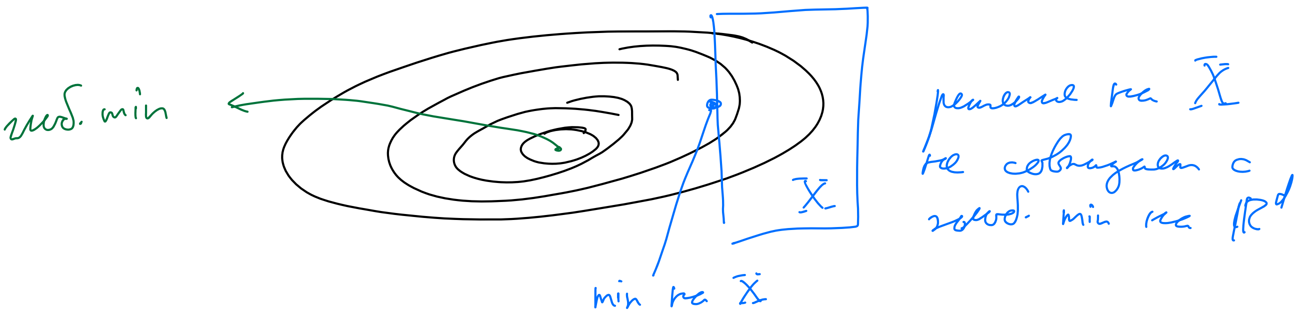
На пред. занятиях $\Longrightarrow$ Сегодня

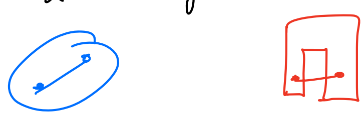
$$\min_{x \in \mathbb{R}^d} f(x) \qquad \Longrightarrow \qquad \min_{x \in \bar{X}} f(x)$$

$\bar{X}$ - „простое" мн-во



Условие оптимальности

- f - непр. дифф. и выпукла на $\mathbb{R}^d$
- $\bar{X}$ - выпукло

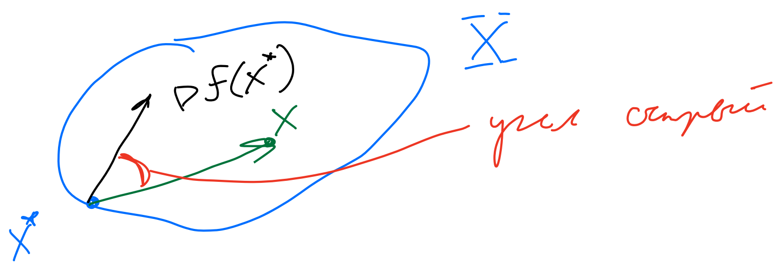


$x^* \in \bar{X}$ - глоб. минимум $f(x)$ на $\bar{X}$

$$\Longleftrightarrow$$

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

Физический смысл:



Док-во:

- достаточность $\Longleftarrow$

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

выпуклость f:

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

x^* - глоб. минимум на X

- необходимость $\Rightarrow$

x^* - глоб. мин. на X

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

$\triangle$ $\underbrace{\tilde{x}_\lambda}_{\in X \text{ (в силу вып. мн-ва } X)} = \lambda \tilde{x} + (1-\lambda) x^* \quad \lambda \in [0;1]$

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

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

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

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

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

с x^* - глоб. минимум

Метод град. спуска с проекцией

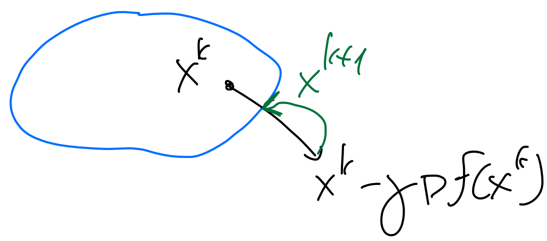
$$x^{k+1} = \underbrace{x^k - \gamma \nabla f(x^k)}_{\in \mathcal{X}}$$

не прин. $\in \mathcal{X}$

$$\boxed{\begin{aligned} x^{k+1} &= \Pi\left(x^k - \gamma \nabla f(x^k)\right) \\ \Pi(y) &= \arg\min_{x \in \mathcal{X}} \|x - y\|_2^2 \end{aligned}}$$

$\leftarrow$ евклидова проекция

Физический смысл:



Св-ва проекции:

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

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

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

Условие оптимальности для d, $\mathcal{X}$

$$\langle \nabla d(z^*); z - z^* \rangle \geq 0 \qquad \forall z \in \mathcal{X}$$

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

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

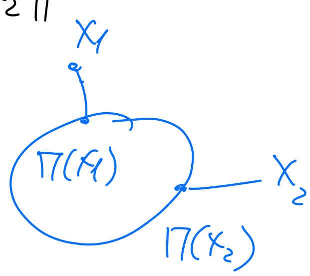
$$\nabla d(\Pi(y)) = 2(\Pi(y) - y)$$

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

2) Нерастяжимость оператора проекции

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

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



Док-во: св-во 1)

$y = x_1 \qquad x = \Pi(x_2) \in X$

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

аналогично $y = x_2 \qquad x = \Pi(x_1)$

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

сложим

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

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

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

$$\|\Pi(x_2)-\Pi(x_1)\| \le \|x_2-x_1\| \quad ■$$

3) Проекция однозначна

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

4) Стационарная точка град. спуска с проекц.:

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

Док-во:

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

$$= \arg\min_{x\in X}\Big[\underbrace{\underbrace{\|x-x^*\|^2}_{\ge 0} + 2\gamma\underbrace{\langle\nabla f(x^*);\, x-x^*\rangle}_{\substack{\ge 0 \\ \text{по усл. оптим}}}}_{\ge 0} + \cancel{\gamma^2\|\nabla f(x^*)\|^2}\Big]$$

не влияет на argmin

0 достигается в x^* ■

Док-во сходимости:

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

св-ва 4)

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

Д-во 2)

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

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

— по μ-сильной выпуклости

〰 по L-гладкости

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

$$= (1 - \gamma\mu) \|x^k - x^*\|^2 + 2\gamma(\gamma L - 1)\left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^* \rangle \right)$$

≥ 0 по выпуклости f

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

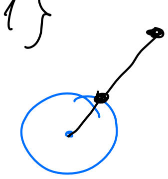
$$\le (1 - \gamma\mu)\|x^k - x^*\|^2 \quad \blacksquare$$

Та же, что и у град спуска.

Примеры

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 $\quad \mathcal{X} = \{x \in \mathbb{R}^d \mid a_i \le x \le b_i\}$

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

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

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

Линейная задача (альтернатива для задачи)

$$\min_{s \in \mathcal{X}} \langle s, g \rangle$$

фиксирован

ограничено

1) L_1-шар $\quad \mathcal{X} = \{x \in \mathbb{R}^d \mid \|x\|_1 \le 1\}$

$$i = \arg\max_j |g_j|$$

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

базис. вектор

$$\underset{g}{\begin{pmatrix} 1 \\ 2 \\ -10 \end{pmatrix}} \quad \underset{s}{\begin{pmatrix} 0 \\ 0 \\ +1 \end{pmatrix}}$$

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

$$s^* = e_i \qquad i = \arg\min_j g_j$$

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

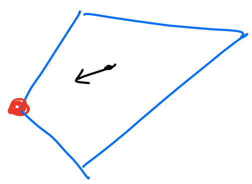
$$s^* = -\sum_{i=1}^d \operatorname{sign}(g_i)\, e_i$$

Метод Франк-Вульфа (условного градиента)

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

Рисунок

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



s^k находится на границе мн-ва

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

усреднение точек с границы

Сходимость

- f - L-гладкая
- f - выпуклая

$$f(x^{k+1}) - f^* \le \frac{2 L D^2}{k+1}$$

$\max\limits_{x,y\in X} \|x-y\|_2$

Итоги по проекциям и ФВ:

⊕ лин. сход. у проекции для μ-сильно вып. ф. (как у GD)

⊕ сублин. сход. у проек. и у ФВ для вып. задач (как у GD)

⊕ есть физика (например, разреж. решение)

⊖ „простые" мн-ва