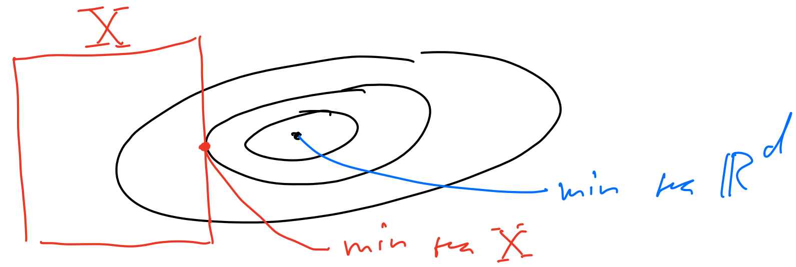
На пред. лекциях

$$\min_{x \in \mathbb{R}^d} f(x) \Rightarrow$$

Сегодня

$$\min_{x \in X} f(x)$$

X – „простое" мн-во



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

Условие оптимальности

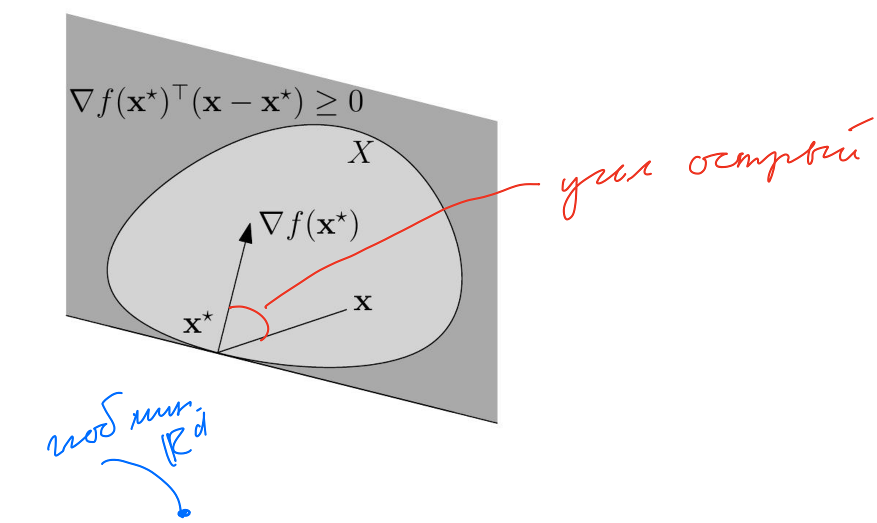
- f – непр. дифф. и выпукла на $\mathbb{R}^d$
- X – выпукло

$x^* \in X$ – глоб. минимум f на X

$$\Longleftrightarrow$$

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

Физический смысл:



Док-во:

- достаточность $\Leftarrow$

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

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. минимум на X

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

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

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

$$\frac{d\varphi}{d\lambda} = \frac{d}{d\lambda}\left(f(x^* + \lambda(\tilde{x} - x^*))\right)$$

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

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

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

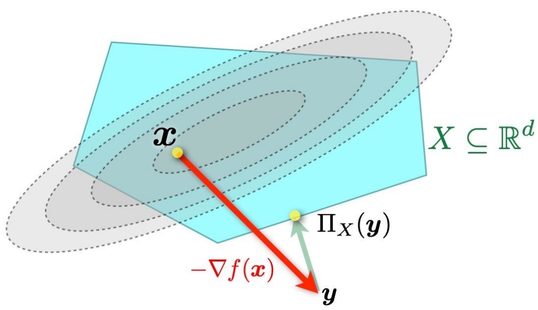
$$f(\underbrace{x^* + \hat{\lambda}(\tilde{x} - x^*)}_{\in X \quad \neq x^*}) = \varphi(\hat{\lambda}) < \varphi(0) = f(x^*)$$

противоречие с x^* – г.м. на X ■

Метод град. спуска с проекцией

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

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



св-ва проекции:

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

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

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

Условие оптимальности для d, X

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

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

$$< \nabla d(\Pi(y)) ; x-\Pi(y) > \geq 0$$

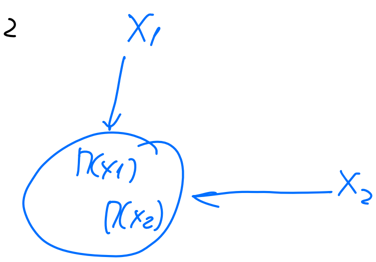
$$\nabla d(z) = 2(z-y)$$

$$2 < \Pi(y)-y ; x-\Pi(y) > \geq 0 \quad \blacksquare$$

2) Нерасширяемость оператора проекции

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

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



Док-во: св-во 1)

$y = x_1 \quad x=\Pi(x_2) \in \bar{X}$

$$\underbrace{< \Pi(x_2)-\Pi(x_1)}_{} ; x_1-\Pi(x_1) > \leq 0$$

аналогично $\quad y = x_2 \quad x=\Pi(x_1) \in \bar{X}$

$$\underbrace{< \Pi(x_1)-\Pi(x_2)}_{} ; x_2-\Pi(x_2) > \leq 0$$

сложим

$$< \Pi(x_2)-\Pi(x_1) ; x_1-\Pi(x_1)-x_2+\Pi(x_2) > \leq 0$$

$$\underbrace{< \Pi(x_2)-\Pi(x_1) ; \Pi(x_2)-\Pi(x_1) >}_{\|\Pi(x_2)-\Pi(x_1)\|_2^2} \leq \underbrace{< x_2-x_1 ; \Pi(x_2)-\Pi(x_1) >}_{\text{КБШ}}$$

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

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

3) Проекция однозначная

$$\underset{x \in X}{\operatorname{argmin}} \|x-y\|_2^2 \leftarrow \text{сильно выпуклая}$$

4) Стационарная точка град. спуска с проекцией

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

Док-во:

$$\Pi(x^* - \gamma \nabla f(x^*)) = \underset{x \in X}{\operatorname{argmin}} \|x - x^* + \gamma \nabla f(x^*)\|_2^2$$

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

≥ 0, 0 достигается $x = x^*$ ■

Док-во сходимости

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

св-во 4)

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

св-во 2)

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

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

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

— μ-сильной выпуклости, или — L-гладкости

$$\leq \|x^k - x^*\|_2^2 + 2\gamma\langle \nabla f(x^*); x^k - x^*\rangle$$
$$- 2\gamma\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right)$$
$$+ 2L\gamma^2\left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^*\rangle\right)$$
$$= (1-\gamma\mu)\|x^k - x^*\|_2^2$$
$$+\underbrace{(2\gamma^2 L - 2\gamma)}_{\leq 0}\underbrace{\left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^*\rangle\right)}_{\geq 0}$$

по выпуклости f

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

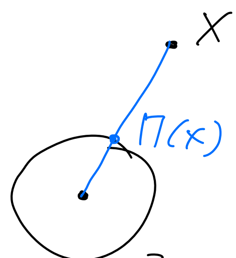
$$\leq (1-\gamma\mu)\|x^k - x^*\|_2^2 \quad \blacksquare$$

Так же, как и град. спуск.

Проекции

1) l_2 − шар $\quad X = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$

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



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

$$[\Pi(x)]_i = \begin{cases} a_i & x_i < a_i \\ x_i & a_i \leq x_i \leq b_i \\ b_i & x_i > b_i \end{cases}$$

$a_i \in [-\infty; +\infty)$
$b_i \in (-\infty; +\infty]$

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

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

Итог по град. спуску с проекцией

⊕ решает задачи на „простых" мн. вах

⊕ сходимость, как у град. спуска на $\mathbb{R}^d$

⊖ проекция считается трудно для большого числа множеств

Линейная задача (как альтернатива проекции)

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

ограничено

1) L_1 - шар

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

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

$$s^* = -\operatorname{sign}(g_j)\, e_j \quad \leftarrow \text{баз. вектор}$$

$$\begin{array}{cc} g & s \\ \begin{pmatrix} 1 \\ 2 \\ +10 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{array}$$

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

$$i = \arg\min_j g_j$$

$$s^* = e_i$$

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

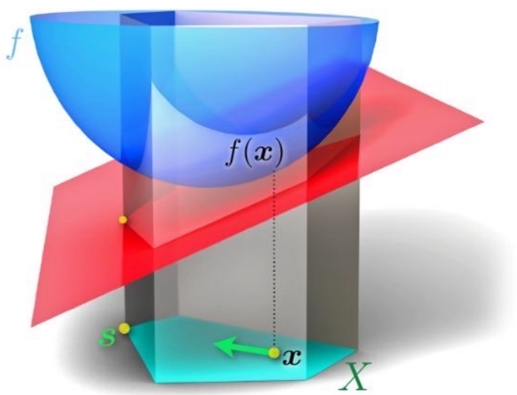
$$s^* = -\sum_{i=1}^{d} \operatorname{sign}(g_i)\, e_i$$

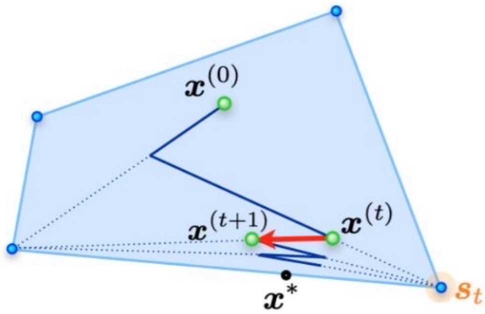
Метод Франк-Вульфа (условного градиента)

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

$$x^{k+1} = (1-\gamma_k)x^k + \gamma_k s^k \qquad \gamma_k = \frac{2}{k+2}$$

Рисунки





- $s^k = \arg\min_{s\in X} \langle s; \nabla f(x^k)\rangle = \arg\min_{s\in X} \Big[\underbrace{f(x^k) + \langle \nabla f(x^k); s - x^k\rangle}_{\text{линейная аппрокс.}}\Big]$

на границе мн-ва

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

усреднение точек на границе

Док-во сходимости

- f – L-гладкая
- f – выпуклая

L-гладкость

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

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

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

$\mathcal{X}$ – ограничено, $D = \dim \mathcal{X} = \max\limits_{x,y \in \mathcal{X}} \|x-y\|_2$

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

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

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

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

$$- f(x^*)$$

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

от k

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

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

БИ: $k = 0$ $\quad f(x^0) - f(x^*) \le \max\{f(x^0) - f(x^*); LD^2\}$

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

$$\overset{\gamma_k}{=} \left(1 - \frac{2}{k+2}\right)\left(f(x^k) - f(x^*)\right) + \frac{\cancel{4}^{\,2}}{(k+2)^2} \frac{LD^2}{\cancel{2}}$$

$$= \frac{k}{k+2}\left(f(x^k) - f(x^*)\right) + \frac{2LD^2}{(k+2)^2}$$

предп. И

$$\le \frac{k}{k+2} \cdot \frac{2}{k+2} \max\{\ \} + \frac{2LD^2}{(k+2)^2}$$

$$\le \frac{2k}{(k+2)^2} \max\{\ \} + \frac{2\max\{\ \}}{(k+2)^2}$$

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

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

$$\le \frac{2\max\{\ \}}{(k+1)+2} \qquad \blacksquare$$

Итоги по ФВ

(+) сублин. сход. $\frac{1}{k}$ для L-гладких выпуклых функций (как у град. спуска)

(+) решает задачу на „простых" мн-вах

(−) в сильно выпуклом случае сход. не улучшается

(−) требуется „простота" мн-ва