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

выпуклое

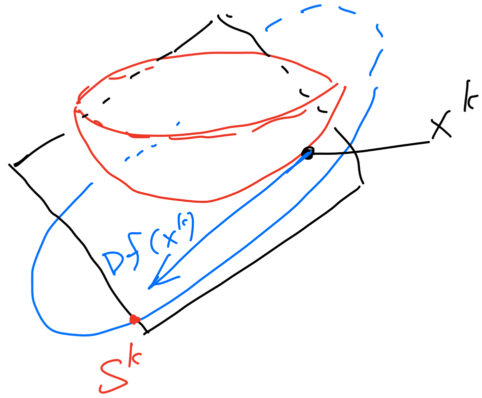
„простое"

Франк – Вульф (C – ограниченное)

$$s^k = \operatorname*{argmin}_{s\in C} \langle \nabla f(x^k);\, s - x^k\rangle$$ ← min линейную аппроксимацию ∇f на C

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

$$\gamma_k = \frac{2}{k+2}$$



L – гладкая, выпуклая

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

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

$(\le)$

$$s^k = \operatorname*{argmin}_{s\in C} \langle \nabla f(x^k);\, s - x^k\rangle$$

$x^* \in C$: $f(x^*) \le f(x)$ $\forall x \in C$ $\qquad \nabla f(x^*) \ne 0$

$$\langle \nabla f(x^k);\, s^k - x^k\rangle \le \langle \nabla f(x^k);\, x^* - x^k\rangle$$

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

$$\le f(x^k) + \gamma_k\left(f(x^*) - f(x^k)\right) + \frac{L\gamma_k^2}{2}\|s^k - x^k\|^2$$

доб. $-f(x^*)$

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

C - ограниченное $\quad D = \max\limits_{x,y\in C} \|x-y\| \quad x^k, s^k \in C$

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

$C_k = \frac{1}{k+2}\max\left(f(x^0) - f(x^*); \frac{4LD^2}{2}\right)$

$f(x^k) - f(x^*) \le C_k$ ← хотим доказать по индукции

индукция

$$f(x^{k+1}) - f(x^*) \le \left(1 - \frac{2}{k+2}\right) C_k + \frac{4}{(k+2)^2}\cdot\frac{LD^2}{2}$$

$$= \frac{k}{(k+2)^2}\cdot\max(\ldots) + \frac{1}{(k+2)^2}\cdot\underbrace{\frac{4LD^2}{2}}_{\le\max(\ldots)}$$

$$\le \frac{k+1}{(k+2)^2}\cdot\max(\ldots)$$

$$\le \frac{1}{k+3}\cdot\max(\ldots)$$

$$\frac{1}{k+3} \overset{?}{\ge} \frac{k+1}{(k+2)^2}$$

$$k^2+4k+4 \overset{?}{\ge} k^2+4k+3$$

$$\underbrace{f(x^k) - f(x^*)}_{\varepsilon} \le \frac{4LD^2}{k+2} \qquad O\left(\frac{1}{\varepsilon}\right)$$

оптимальный $(\nabla f(x))$ (Nesterov) $\sim \frac{1}{k^2}$ $\qquad O\left(\frac{1}{\sqrt{\varepsilon}}\right)$

с точки зрения мн. мин.: ФВ неулучш.

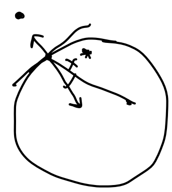
с точки зрения $\nabla f(x)$: можно улучшить и получить оценку Нестерова

$\nabla f(x^*) \neq 0$ (может быть) $\quad f(x)$ - выпуклая

C - выпуклое, замкнутое, то

x^* - решение $\min\limits_{x\in C} f(x) \iff \langle \nabla f(x^*); x - x^* \rangle \geq 0$, $\quad x \in C$

$\Leftarrow$



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

$$\geq f(x^*) \Rightarrow x^* - \text{решение}$$

$\Rightarrow$ от противного: $\exists\, x$:

$\langle \nabla f(x^*); x - x^* \rangle < 0$

$$\varphi(a) = f(x^* + a(x - x^*))$$

$$\varphi(0) = f(x^*) \qquad \varphi'(0) = \langle \nabla f(x^* + a(x - x^*)); x - x^* \rangle \Big|_{a=0}$$

$$= \langle \nabla f(x^*); x - x^* \rangle$$

$$< 0$$

в некотор окр в окр 0 $\varphi(a)$ убывает

а значит $\tilde{a} \in$ этой окрестности $(\min(R; 1))$

$\uparrow$ радиус окр.

$$\varphi(\tilde{a}) < \varphi(0)$$

$$f(\underbrace{x^* + \tilde{a}(x - x^*)}_{(1-\tilde{a})x^* + \tilde{a}x \,\in\, C}) < f(x^*) \rightarrow \text{противоречие, а значит}$$

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

Метод зеркального спуска

$$x \in \Delta = \{x \mid x_i \geq 0, \ \sum x_i = 1\}$$

Дивергенция Брэгмана, порожденная $\psi(x)$ − 1 сильно выпуклой $\|\cdot\|_p$ на C

$p \leq 2$

в определ. сильной вып. переходим $\| \|_2 \to \| \|_p$

$$V(x,y) = \psi(x) - \psi(y) - \langle \nabla\psi(y); x-y \rangle$$

1) $\psi(x) = \frac{1}{2}\|x\|^2$

$$V(x,y) = \frac{1}{2}\|x\|^2 - \frac{1}{2}\|y\|^2 - \langle y; x-y \rangle$$
$$= \frac{1}{2}\|x-y\|^2$$

2) $\psi(x) = \sum_{i=1}^{d} x_i \log x_i \qquad C = \Delta_d$

$$V(x,y) = \sum x_i \log x_i - \sum y_i \log y_i -$$
$$- \sum (\log y_i + 1)(x_i - y_i)$$
$$= \sum x_i \log x_i - \sum x_i \log y_i$$
$$- \underbrace{\sum x_i}_{=1} + \underbrace{\sum y_i}_{=1}$$
$$= \sum_{i=1}^{d} x_i \log \frac{x_i}{y_i} = KL(x \| y)$$

Св-во: $V(x,y) \geq \frac{1}{2}\|x-y\|_p^2$ из сильной выпуклости

Упр. Доказать $KL(x\|y)$ — 1-сильно вып. $(\| \ \|_1)$ на симплексе

Св-ва:

$$V(z;x)+V(x;y)-V(z;y)$$

$$= \cancel{\psi(z)} - \cancel{\psi(x)} - \langle \nabla\psi(x); z-x \rangle$$

$$\cancel{\psi(x)} - \cancel{\psi(y)} - \langle \nabla\psi(y); x-y \rangle$$

$$- \left(\cancel{\psi(z)} - \cancel{\psi(y)} - \langle \nabla\psi(y); z-y \rangle \right)$$

$$= \langle \nabla\psi(y); y-x+z-y \rangle + \langle \nabla\psi(x); x-z \rangle$$

$$= \langle \nabla\psi(y)-\nabla\psi(x); z-x \rangle$$

Метод ЗС:

$$x^{k+1} = \arg\min_{x\in C} \left(\gamma\langle \nabla f(x^k); x \rangle + \frac{1}{2}\|x-x^k\|_2^2 \right) \quad - \ GD$$

$$\| \ \arg\min \frac{1}{2}\|x-(x^k-\gamma\nabla f(x^k))\|^2$$

$$\Downarrow \text{ ЗС}$$

$$x^{k+1} = \arg\min_{x\in C} \left(\langle \nabla f(x^k); x \rangle + V(x, x^k) \right)$$

Док-во сход. (L-гладкий, выпуклый)

$\| \ \|_p$ $\rightarrow$ $f(x) \le f(y) + \langle \nabla f(y); x-y \rangle + \frac{L}{2}\|x-y\|_p^2$

$$\langle \nabla g(\tilde{x}^*); x - \tilde{x}^* \rangle \geq 0$$
$$x \in C$$

$$g(\tilde{x}^*) = \langle \nabla f(x^k); \tilde{x}^* \rangle + V(\tilde{x}^*, x^k)$$
$$\tilde{x}^* = x^{k+1}$$
$x = x^* \leftarrow$ решение $\min\limits_{C} f(x)$

$$\langle x^* - x^{k+1}; \gamma \nabla f(x^k) + \nabla\psi(x^{k+1}) - \nabla\psi(x^k) \rangle \geq 0$$
$$\langle x^* - x^{k+1}; \gamma \nabla f(x^k) \rangle + \langle x^* - x^{k+1}; \nabla\psi(x^{k+1}) - \nabla\psi(x^k) \rangle \geq 0$$

$$V(z;x) + V(x;y) - V(z;y) = \langle \nabla\psi(y) - \nabla\psi(x); z - x \rangle$$

$x = x^{k+1}$, $z = x^*$, $y = x^k$

$$\langle x^* - x^{k+1}; \gamma \nabla f(x^k) \rangle + \underbrace{V(z,y) - V(x,y) - V(z,x)}_{V(x^*;x^k) - V(x^{k+1};x^k) - V(x^*;x^{k+1})} \geq 0$$

$$\underbrace{\langle \gamma \nabla f(x^k); x^* - x^k \rangle}_{\text{выпуклость}} + \gamma \underbrace{\langle \nabla f(x^k); x^k - x^{k+1} \rangle}_{L\text{-гладкость } (\|\cdot\|_p) + \text{выпуклость}} + V(x^*; x^k) - \underbrace{V(x^{k+1}; x^k)}_{\leq \frac{1}{2}\|x^{k+1} - x^k\|_p^2} - V(x^*; x^{k+1}) \geq 0$$

$$\underbrace{\gamma\left(f(x^*) - f(x^k)\right)}_{\text{вып.}} + \gamma\underbrace{\left(f(x^k) - f(x^{k+1})\right) + \frac{\gamma L}{2}\|x^{k+1} - x^k\|_p^2}_{\text{гладк}}$$

$$+ V(x^*; x^k) - \underbrace{\frac{1}{2}\|x^{k+1} - x^k\|_p^2}_{V \text{ сильн. вып}} - V(x^*; x^{k+1}) \geq 0$$

$$\gamma\left(f(x^{k+1}) - f(x^*)\right) \leq V(x^*; x^k) - V(x^*; x^{k+1})$$

$$+ \underbrace{\left(\frac{L\gamma}{2} - \frac{1}{2}\right)}_{\leq 0}\|x^{k+1} - x^k\|_p^2$$

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

$$\gamma\left(f(x^{k+1}) - f(x^*)\right) \leq V(x^*; x^k) - V(x^*; x^{k+1})$$

далее как в док-ве град спуска

$\sum$ $k = 0$ до K, Йенсен f

$$f(\bar{x}^K) - f(x^*) \leq \frac{V(x^*; x^0)}{\gamma K+1} = \frac{L V(\ldots)}{K+1}$$