Лекция 4. Методы типа проекции градиента

$$\min_{x \in Q} f(x) \longrightarrow x_* \text{ реш. задачи}$$

$$Q = \mathbb{R}^d_+, \quad Q = B^d_p(R), \quad Q = S_d(1)$$

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

$$\|x^{k+1} - x_*\|_2^2 = \|\pi_Q(x^k - h\nabla f(x^k)) - x_*\|_2^2 =$$

$$= \|\pi_Q(x^k - h\nabla f(x^k) - x_*)\|_2^2 \leq$$

$$\leq \|x^k - h\nabla f(x^k) - x_*\|_2^2 =$$

$$= \|x^k - x_*\|_2^2 - 2h\langle \nabla f(x^k), x^k - x_*\rangle + h^2\|\nabla f(x^k)\|_2^2$$

$$\frac{1}{N}\sum_{k=0}^{N-1} \Bigg| \quad \underbrace{2h\langle \nabla f(x^k), x^k - x_*\rangle}_{\substack{\vee\!/ \\ 2h(f(x^k) - f(x_*))}} \leq \|x^k - x_*\|_2^2 - \|x^{k+1} - x_*\|_2^2 + h^2\underbrace{\|\nabla f(x^k)\|_2^2}_{\substack{\wedge\!| \\ M^2}} \lesssim 2L(f(x^k) - f(x_*))$$

$$2h\frac{1}{N}\sum_{k=0}^{N-1}\left[f(x^k) - f(x_*)\right) \leq \overbrace{\|x^0 - x_*\|_2^2}^{R^2} - \|x^N - x_*\|_2^2 + \frac{1}{N}\sum_{k=0}^{N-1} h^2 \cdot \frac{M^2}{\underline{2L(f(x^k) - f(x_*))}}$$

$$\frac{1}{N}(2h - h^2)\sum_{k=0}^{N-1}\left(f(x^k) - f(x_*)\right) \leq R^2/N$$

$$\frac{1}{N}(2h - h^2)\left(f(\bar{x}^N) - f(x_*)\right) \leq R^2/N \qquad \bar{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k \qquad h = \frac{1}{2L}$$

$$f(\bar{x}^N) - f(x_*) \leq \frac{2LR^2}{N}; \quad \frac{MR}{\sqrt{N}}$$

$$f(\bar{x}^N) - f(x_*) \le \frac{M_2 R_2}{\sqrt{N}}$$

$$\left[\begin{array}{l} M_2 \to \quad |f(y) - f(x)| \le M\|y-x\|_2 \\ R_2 \to \quad \|x^0 - x_*\|_2 \end{array}\right.$$

$M_1 \ll M_2$

$R_1 \to \|x^0 - x_*\|_1 \simeq R_2$

$\frac{1}{p} + \frac{1}{q} = 1$

$M_p = \max\limits_{x \in Q} \|\nabla f(x)\|_q$

$M_1 = \max\limits_{x \in Q} \|\nabla f(x)\|_\infty$

$M_2 = \max\limits_{x \in Q} \|\nabla f(x)\|_2$

$M_1 \sqrt{d} \simeq M_2$ может также быть

$\|(\underbrace{1, \dots, 1}_{d})\|_\infty = 1$

$\|(\underbrace{1, \dots, 1}_{d})\|_2 = \sqrt{d}$

$\|(1, 0, \dots, 0)\|_\infty = 1$

$\|(1, 0, \dots, 0)\|_2 = 1$

$\|\nabla f(y) - \nabla f(x)\|_q \le$

$\le L\|y - x\|_p$

$\frac{1}{p} + \frac{1}{q} = 1$

$p = q = 2$

$p = 1, \quad q = \infty$

$L_p = \max\limits_{x \in Q} \max\limits_{\|h\|_p \le 1} \langle h, \nabla^2 f(x) h \rangle$

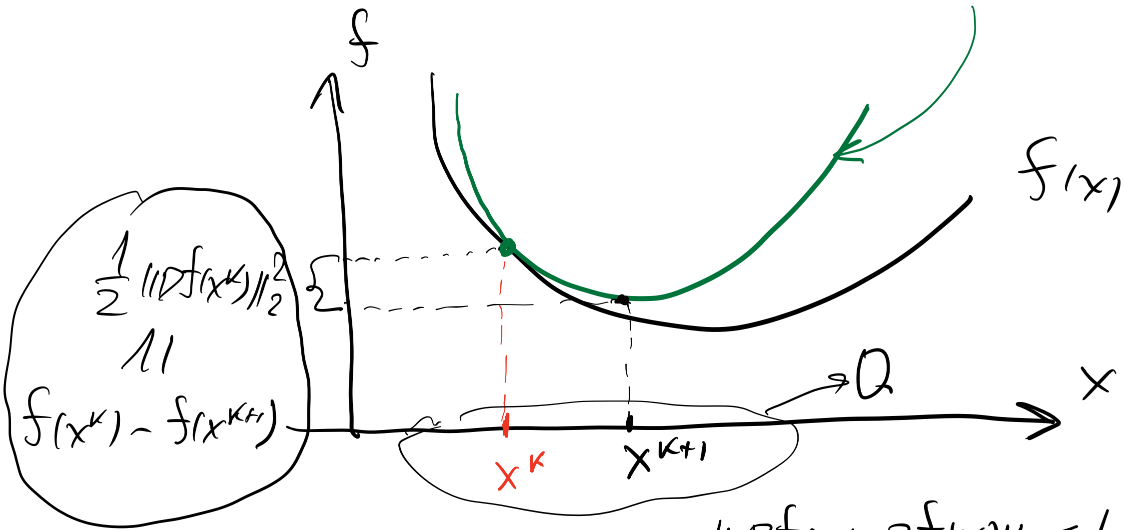
$$\boxed{\frac{L_2}{d} \le L_1 \le L_2}$$

$$\boxed{\frac{M_2}{\sqrt{d}} \le M_1 \le M_2}$$

$f(\bar{x}^N) - f(x_*) \lesssim \frac{L_p R_p^2}{N^\alpha}$ $\quad \alpha = 1, 2$ $\nearrow$ неуск. $\searrow$ уск.

$\frac{M_p R_p}{\sqrt{N}}$

$$x^{k+1} = \underset{x\in Q}{\arg\min} \left\{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2}\|x - x^k\|_2^2 \right\} = \pi_Q\left(x^k - \frac{1}{L}\nabla f(x^k)\right)$$



$$\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y - x\|_2$$

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

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

$$f(x^N) - f(x_*) \le \frac{L_p \tilde{R}_p^2}{N}$$

$$\max_{\{x:\, f(x) \le f(x^0)\}} \|x - x_0\|_p^2$$

А.С. Немировский

конец 70-х

начало 80-х

прокс. ф-ция

$$\frac{1}{2}\|x - x^k\|_2^2 = \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|x^k\|_2^2 - \left\langle \nabla\left(\frac{1}{2}\|x\|_2^2\right)\Big|_{x=x^k}, x - x^k \right\rangle$$

$\frac{1}{2}\|x\|_2^2 \to d(x)$ – 1-сильно вып. отн. p-нормы на Q, $p \in [1,2]$

$$V(x, x^k) = d(x) - d(x^k) - \langle \nabla d(x^k), x - x^k \rangle$$

↘ дивергенция Брэгмана

1-норма, $Q = S_d(1)$

$$d(x) = \sum_{i=1}^{d} x_i \ln x_i \quad \text{— энтропия}$$

↘ 1-сильно вып. ф-ция отн. 1-нормы на $S_d(1)$

$$\mu_p = \max_{x \in Q} \min_{\|h\|_p \le 1} \langle h, \underbrace{\nabla^2 f(x)}_{\begin{bmatrix} 1/x_1 & & 0 \\ & \ddots & \\ 0 & & 1/x_d \end{bmatrix}} h \rangle \qquad p = 1$$

$$KL(x, x^k) = \sum_{k=1}^{d} x_i \ln(x_i / x_i^k)$$

$$x^{k+1} = \operatorname*{argmin}_{x \in S_d(1)} \{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + L_1 \underbrace{V(x, x^k)}_{KL(x, x^k)} \} \rightarrow$$

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

$$\|\nabla f(y) - \nabla f(x)\|_q \le L_p \|y - x\|_p$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$\le f(x) + \langle \nabla f(x), y - x \rangle + L_p V(y, x)$$

$$V \ge \frac{1}{2} \|y - x\|_p^2$$

$$\rightarrow x_i^{k+1} = \frac{\exp\left(-\frac{\partial f(x^k)/\partial x_i}{L}\right)}{\sum_{j=1}^{d} \exp\left(-\frac{\partial f(x^k)/\partial x_j}{L}\right)}$$

$O(d)$

$$f(\bar{x}^N) - f(x_*) \le \frac{M_p \bar{R}_p}{\sqrt{N}}$$

$$\frac{L_p \bar{R}_p^2}{N^\alpha}, \quad \alpha = 1,2$$

$$\underbrace{\bar{R}_p^2 = 2V(x_*, x^0)}$$

$$R_p^2 = \|x^0 - x_*\|_p^2$$

$$\underbrace{\underbrace{O(\ln d)} \cdot \|x^0 - x_*\|_p^2 \ge V(x_*, x^0) \ge \frac{1}{2}\|x^0 - x_*\|_p^2}$$

$$Q = S_d(1)$$

$$\max_{x^0, x_* \in S_d(1)} \|x^0 - x_*\|_1^2 \le 2$$

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$\max_{x_*} KL(x^0, x_*) \le 2\ln d$$

$x^0 = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ центр симплекса

$$\left\| \begin{matrix} x \\ x \\ x \end{matrix} \right\} 1 \Big\|_1^2 = 1$$

$$\left\| \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix} \right\} 1 \Big\|_2^2 = 1$$

$$x^{k+1} = \mathrm{Mirr}_{x^k}^Q\left(\underbrace{h\nabla f(x^k)}\right) =$$

$$= \arg\min_{x \in Q} \left\{ \cancel{f(x^k)} + h\langle \nabla f(x^k), x - x^k \rangle + V(x, x^k) \right\}$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$d(x)$ – 1-сильно-выпукл. в p-норме

$$V(x^{k+1}, x_*) \le V(x^k, x_*) - 2h\langle \nabla f(x^k), x^k - x_* \rangle + h^2 \|\nabla f(x^k)\|_2^2$$

$$\frac{1}{2}\|x^{k+1} - x_*\|_2^2 \le \frac{1}{2}\|x^k - x_*\|_2^2 - 2h\langle \nabla f(x^k), x^k - x_* \rangle + h^2 \|\nabla f(x^k)\|_2^2$$

см. §3 пособия МСиНМО.

Пример (симплекс).

$$\min_{x \in Q} f(x) \to \frac{1}{2}\langle x, Ax \rangle - \langle b, x \rangle$$

$$Q = S_d(1)$$

$$\{x \ge 0 : \sum_{i=1}^{d} x_i = 1\}$$

$$x^{k+1} = \underbrace{\Pi_Q}_{\tilde{O}(d)}(x^k - \overset{\frac{1}{L}}{h}\underbrace{\nabla f(x^k)}_{\underbrace{Ax}_{O(d^2)} - b})$$

$$f(\bar{x}^N) - f(x_*)$$

$$L_2 = \cancel{\max_{x \in Q}} \max_{h: \|h\|_2 \le 1} \langle h, \underbrace{\nabla^2 f(x)}_{A} h \rangle =$$

$$= \lambda_{\max}(A)$$

$$R^2 = \|\underset{\hat{Q}}{x^0} - \underset{\hat{Q}}{x_*}\|_2^2 \le 2$$

$$\underbrace{O(d^2 + d)}_{\text{стоим. итерации}} \qquad \underset{x^N}{f(\bar{x}^N) - f(x_*)} \lesssim \frac{LR^2}{N} \quad \frac{2\lambda_{\max}(A)}{N}$$

$p = 2$ $\boxed{N = \frac{\lambda_{max}(A)\cdot 2}{\varepsilon}, \quad O(d^2)}$

$p = 1$ $\boxed{N = \frac{\max|A_{ij}|\, 2\ln d}{\varepsilon}, \quad O(d^2)}$

$$L_1 = \max_{h:\, \|h\|_1 \le 1} \langle h, Ah \rangle = \max_{i,j} |A_{ij}| \le$$

$$\le \lambda_{max}(A)$$

$R^2 \le 2\ln d$

$M_p R_p$ $\quad 1 \le p \le 2$

$\|\nabla f(x)\|_q$ $\quad \|x^0 - x_*\|_p \ln d$

$p = 1$

$\|\nabla f(x)\|_\infty \le \|\nabla f(x)\|_2$ $\quad \|x^0 - x_*\|_1 \ge \|x^0 - x_*\|_2$

Ускоренный метод

Метод подобных треугольников

$y^0 = z^0, \quad A_0 = \alpha_0 = 0, \quad \alpha_{k+1} = \frac{1}{2L} + \sqrt{\frac{1}{4L^2} + \alpha_k^2}$

$$A_{k+1} = A_k + \alpha_{k+1} \qquad \min_{x \in Q} f(x)$$

$$\begin{cases} x^{k+1} = \dfrac{\alpha_{k+1} z^k + A_k y^k}{A_{k+1}} \\[2ex] z^{k+1} = \operatorname{argmin}_{x \in Q} \{ \alpha_{k+1} \langle \nabla f(x^{k+1}), x - x^{k+1} \rangle + V(x, z^k) \} \\[2ex] y^{k+1} = \dfrac{\alpha_{k+1} z^{k+1} + A_k y^k}{A_{k+1}} \end{cases}$$

$$f(y^N) - f(y_*) \le \frac{4LR^2}{N^2}$$

$$R^2 = V(x_*, x^0) \longrightarrow \text{порожд. } d(x)$$

$\downarrow$

1-сильно вып. отн. p-нормы

$$\|\nabla f(y) - \nabla f(x)\|_q \le L \|y - x\|_p$$

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

$$\cancel{\nabla f(x)} \to \nabla f(x, \xi) \quad [\nabla_x f(x, \xi)]$$

$$\mathbb{E}[\nabla f(x, \xi)] = \nabla f(x) \qquad \forall x \in Q$$

$$\mathbb{E}[\|\nabla f(x, \xi)\|_q^2] \le M^2 \qquad \forall x \in Q$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$p \in [1, 2]$$

$\boxed{h = \frac{\varepsilon}{M^2}}$ $\qquad \underbrace{x^{k+1} = \mathrm{Mirr}^{Q}_{x^k}\left(h \overbrace{\nabla f(x^k, \xi^k)}^{\xi}\right)}$

$$\boxed{V(x_*, x^{k+1}) \le V(x_*, x^k) - 2h \langle \xi, x^k - x_* \rangle + h^2 \underbrace{\|\xi\|_q^2}_{\le M^2}}$$

$$2h \langle \nabla f(x^k, \xi^k), x^k - x_* \rangle \le V(x_*, x^k) - V(x_*, x^{k+1}) + h^2 \|\nabla f(x^k, \xi^k)\|_q^2 \qquad \mathbb{E}_{\xi^k}[\cdot \mid x^k]$$

$x^k(\xi^0, \ldots, \xi^{k-1})$ $\qquad x^k$ не зависит от ξ^k

$$2h \underbrace{\langle \nabla f(x^k), x^k - x_* \rangle}_{\substack{\ge \\ f(x^k) - f(x_*)}} \le V(x_*, x^k) - \mathbb{E}_{\xi^k}[V(x_*, x^{k+1}) \mid x^k] + h^2 M^2 \quad \Big\} \mathbb{E}_{x^k}$$

$$\mathbb{E}\left[2h(f(x^k) - f(x_*))\right] \le \mathbb{E}\, V(x_*, x^k) - \mathbb{E}\, V(x_*, x^{k+1}) + h^2 M^2 \qquad \frac{1}{N}\sum_{k=0}^{N-1}$$

$$2hN\left[\mathbb{E}[f(\bar{x}^N)] - f(x_*)\right] \le V(x_*, x^0) + h^2 M^2 N$$

$$\mathbb{E}\left[f(\bar{x}^N)\right] - f(x_*) \le \underbrace{\frac{R^2}{2hN} + \frac{h}{2}M^2}_{h = \frac{R}{M\sqrt{N}} = \varepsilon / M^2} = \underbrace{\frac{MR}{\sqrt{N}}}_{= \varepsilon}$$