Лекция 3.

Метод сопряженных градиентов

$$\boxed{\min_{x \in \mathbb{R}^d} f(x)} \qquad \underbrace{f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle}$$

$$x^{k+1} \in x^0 + \mathrm{Lin}\{\nabla f(x^0), \nabla f(x^1), \dots, \nabla f(x^k)\}$$

$$x^{k+1} \in \operatorname*{Argmin}_{x \in x^0 + \mathrm{Lin}\{\nabla f(x^0), \dots, \nabla f(x^k)\}} f(x) = \qquad \underbrace{Ax = b}$$

$$= x^k - \alpha_k \underbrace{\nabla f(x^k)}_{= Ax^k - b} + \beta_k (x^k - x^{k-1})$$

$$(\alpha_k, \beta_k) \in \operatorname*{Argmin}_{(\alpha, \beta)} f(x^k - \alpha \nabla f(x^k) + \beta (x^k - x^{k-1}))$$

$$\left.\begin{aligned} x^{k+1} &= x^k - h_k \nabla f(x^k) \\ h_k &\in \operatorname*{Argmin}_h f(x - h \nabla f(x^k)) \end{aligned}\right\} \text{наискор. спуск.}$$

$$x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k), \qquad \|\nabla f(y) - \nabla f(x)\|_2 \le L \|y - x\|_2$$

$$f(x^N) - f(x_*) \le \min\left\{ \frac{LR^2}{2(2N+1)^2}, \ 2LR^2 \left(\frac{\sqrt{\chi} - 1}{\sqrt{\chi} + 1}\right)^{2N}, \ \left(\frac{\lambda_{d-N+1} - \lambda_1}{\lambda_{d-N+1} + \lambda_1}\right)^2 R^2 \right\}$$

$$\underset{\overset{\parallel}{M}}{\lambda_1} \le \lambda_2 \le \dots \le \underset{}{\lambda_d} = L$$

$$\|\nabla f(y) - \nabla f(x)\|_2 \le L \|y - x\|_2, \quad L = \lambda_{max}(A)$$

$\varkappa = L/\mu$, μ – конст. сильн. выпр.

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

$$\mu = \lambda_{\min}(A)$$

$R^2 = \|x^0 - x_*\|_2^2$, x_* – ближ. реш. к x^0

$$\mathrm{Lin}\{\nabla f(x^0), \dots, \nabla f(x^k)\} = \mathrm{Lin}\{Ax^0 - b, Ax^1 - b, \dots, Ax^k - b\}$$

$$\underbrace{Ax_* = b}$$

$$\| $$

$$\mathrm{Lin}\{A(x^0 - x_*), A^2(x^0 - x_*), \dots, A^k(x^0 - x_*)\}$$ – подпр-во Крылова

$k=0$ верно, потому что $Ax_* = b$

шаг. индук.
$$x^{k+1} = x^0 + \sum_{l=0}^{k} A^l (x^0 - x_*)$$

$$\nabla f(x^{k+1}) = A\left(x^0 + \sum_{l=0}^{k} A^l (x^0 - x_*)\right) - \overset{Ax_*}{b} =$$

$$= \underbrace{A(x^0 - x_*) + \sum_{l=1}^{k+1} A^l (x^0 - x_*)} = c + A^{k+1}(x^0 - x_*)$$

$$\mathrm{Lin}\{\nabla f(x^0), \dots, \nabla f(x^k)\} = \mathrm{Lin}\{A(x^0 - x_*), \dots, A^k(x^0 - x_*)\}$$

$\nabla f(x^{k+1})$ $\qquad$ $A^{k+1}(x^0 - x_*)$

$$x^N - x_* = P_N(A)(x^0 - x_*)$$

$$P_N(A) = 1 + a_{1N}A + a_{2N}A^2 + \dots + a_{NN}A^N$$

$$\underbrace{f(x^N) - f(x_*)}_{\min} = \frac{1}{2}\langle x^N, Ax^N\rangle - \langle b, x^N\rangle -$$

$$-\left(\frac{1}{2}\langle x_*, Ax_*\rangle - \langle b, x_*\rangle\right) =$$

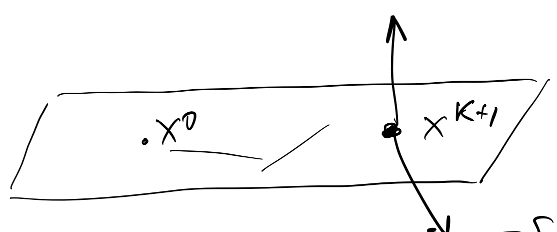
$$= \frac{1}{2}\underbrace{\langle x^N - x_*, A(x^N - x_*)\rangle}_{\min} =$$

$$= \min_{a_{11}, \dots, a_{NN}} \left\{\frac{1}{2}\langle P_N(A)(x^0 - x_*), A(P_N(A)(x^0 - x_*))\rangle\right\} =$$

$$= \min_{a_{11}, \dots, a_{NN}} \left\{\frac{1}{2}\langle AP_N^2(A)(x^0 - x_*), (x^0 - x_*)\rangle\right\} \le$$

$$\le \frac{1}{2}\min_{P_N(\lambda):\ P_N(0)=1}\left[\max_{\underbrace{\mu = \lambda_1 \le \lambda_2 \le \dots \le \lambda_j}= L} \lambda P_N^2(\lambda)\right] R^2$$

$$R^2 = \|x^0 - x_*\|_2^2$$



Λ_k

$\nabla f(x^{k+1}) \perp$ $\mathrm{Lin}\{\nabla f(x^0), \dots, \nabla f(x^k)\}$

1) $\langle \nabla f(x^{k+1}), p\rangle = 0$

$p \in \Lambda_k$

2) $\Lambda_k = \{\delta^0, \delta^1, \dots, \delta^k\}$

$\delta^k = x^{k+1} - x^k$

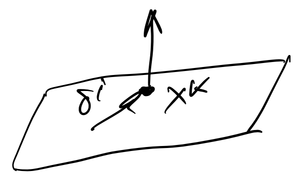
3) $k \neq i$ $\quad \langle A\delta^k, \delta^i \rangle = 0$

$$\langle A\delta^k, \delta^i \rangle = \langle A(x^{k+1} - x^k), \delta^i \rangle$$

$$\langle \nabla f(x^{k+1}) - \nabla f(x^k), \delta^i \rangle = 0$$

$$\nabla f(x^{k+1}) - \nabla f(x^k) = (Ax^{k+1} - b) - (Ax^k - b)$$

$k = i+1$



$$x^{k+1} = x^k - h_k \nabla f(x^k) + \sum_{i=0}^{k-1} \lambda_i \delta^i$$

$A\delta^j$ $\quad \Big| \quad \delta^k = -h_k \nabla f(x^k) + \sum_{i=0}^{k-1} \lambda_i \delta^i$

$\underline{j < k-1}$

$\lambda_j = 0$

$$\langle A\delta^j, \delta^k \rangle = -h_k \langle \nabla f(x^k), A\delta^j \rangle + \sum_{i=0}^{k-1} \lambda_i \langle \delta^i, A\delta^j \rangle =$$

$\langle A\delta^j, \delta^k \rangle = 0$

$$= \underbrace{-h_k \langle \nabla f(x^k), \nabla f(x^{j+1}) - \nabla f(x^j) \rangle}_{=0} + \lambda_j \underbrace{\langle \delta^j, A\delta^j \rangle}_{\neq 0}$$

$$0 = -0 + \lambda_j \underbrace{\langle \delta^j, A\delta^j \rangle}_{\neq 0} \Rightarrow \lambda_j = 0$$

64 Поляк   Метод тяж. шар.

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) + \beta_k (x^k - x^{k-1})$$

75 Немировский

81 Немировский   (вспом. миним.)

3d, 2d, 1d

83 Нестеров $\underbrace{x^1 = x^0 - \frac{1}{L}\nabla f(x^0)}$

$k=1$

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{k-1}{k+2}(x^k - x^{k-1})\right) +$$

$$+ \frac{k-1}{k+2}(x^k - x^{k-1})$$

$$f(x^N) - f(x_*) \le \frac{4LR^2}{N^2} \quad \Big\} \quad \text{град. метод} \quad \frac{LR^2}{N}$$

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1})\right) +$$

$$+ \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1})$$

$$f(x^N) - f(x_*) \le LR^2 \exp\left(-\frac{1}{2}\sqrt{\frac{\mu}{L}}N\right) \quad \Big\} \quad \text{град. метод} \quad \cdot \exp\left(-\frac{\mu}{L}N\right)$$

$$f(x^N) - f(x_*) \le \varepsilon$$

Ускор. метод.

$N \simeq \sqrt{\frac{LR^2}{\varepsilon}}$

$N \simeq \sqrt{\frac{L}{\mu}} \ln\left(\frac{LR^2}{\varepsilon}\right)$

Неускор. метод

$N = \frac{LR^2}{\varepsilon}$

$N = \frac{L}{\mu} \ln\left(\frac{LR^2}{\varepsilon}\right)$

$$\boxed{\mu \simeq \varepsilon / R^2}$$

Регуляризация

$f(x) \to \min \quad (*)$

$f_\mu(x) := f(x) + \frac{\mu}{2}\|x\|_2^2 \to \min\limits_x$

реш. зад. $(*)$

$f_\mu(x^N) - \min\limits_x f_\mu(x) \le \varepsilon/2$

$\mu = \varepsilon/R^2$

$\mu \le \varepsilon/R^2$

$R^2 = \|x_*\|_2^2$

$f(x^N) - \underbrace{\min_x f(x)}_{f(x_*)} \le \varepsilon$

$f_\mu(x^N) - \min\limits_x f_\mu(x) \le \varepsilon/2$

$\vee\!/ \quad \{ f(x^N) \le f_\mu(x^N), \; f_\mu(x_*) \ge \min\limits_x f_\mu(x)$

$f(x^N) - f_\mu(x_*) = f(x^N) - f(x_*) - \frac{\mu}{2}\|x_*\|_2^2 \le \varepsilon/2$

$$f(x^N) - f(x_*) \le \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Рестарты

$$\frac{\mu}{2} \underbrace{\|x^N - x_*\|_2^2}_{\frac{1}{2}\underbrace{\|x^0 - x_*\|^2}_{R^2}} \le f(x^N) - f(x_*) \le \frac{4L\|x^0 - x_*\|_2^2}{N^2}$$

$\to$ выбираем N

$$\frac{\mu}{2}\frac{R^2}{2} \le \frac{4L R^2}{N^2}$$

$$N^2 \simeq 16\frac{L}{\mu}$$

$$\Downarrow$$

$$\underbrace{N = 4\sqrt{\frac{L}{\mu}}}$$

$$\|x^N - x_*\|_2^2 = \frac{1}{2}\|x^0 - x_*\|_2^2$$

Рестартую $\quad x^0 := x^N$

$$\ln\left(\frac{\mu R^2}{\varepsilon}\right)$$

$$N = 4\sqrt{\frac{L}{\mu}}\ln\left(\frac{\mu R^2}{\varepsilon}\right)$$