Орг. вопросы:

- 2 пары в неделю (обе онлайн): общая (время в вскр. определим), индивидуальная по проектам (см. запись в отдельном файле)
- заявление: вскр. 12:00
- форма по проектам: вскр. 20:00
- 2 типа д/з: jupyter notebook и проект
- jn: пт. 23:55

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

x^* – локал. мин. $f(x)$

$$\nabla f(x^*) = 0$$

---

x^*

$$\min_{x \in \mathbb{R}^d} \| x - b \|^2 \qquad x^* = b$$

$$x^0 \qquad \varphi(x^0, \text{par}) = x^1$$

$$x^0, x^1 \dots x^k \qquad \underline{x^k \to x^*}$$

$$\left\{ \begin{array}{l} \|x^k - x^*\|_2^2 < \varepsilon \\ f(x^k) - f(x^*) < \varepsilon \\ \|\nabla f(x^k)\|_2^2 < \varepsilon \end{array} \right.$$

$$\|x_{k+1} - x_k\| \leq$$

$$\underbrace{\|x_k - x^*\|}_{<\varepsilon} + \underbrace{\|x_{k+1} - x^*\|}_{<\varepsilon} \leq 2\varepsilon$$

$$\|x_{k+1} - x_k\| \sim \|\nabla f(x_k)\|$$

Оракулы:

1) нулевого порядка: нулевой оракул (знач. ф) $f(x)$

2) первого порядка: $\nabla f(x)$

3) второго: $\nabla^2 f(x)$

4) тензорные: p-порядок

Скорости сходимости:

1) сублинейная:

$$\|x^k - x^*\| \le \frac{C}{k^\alpha} \qquad \alpha > 0,\ 0 < C < \infty$$

$$\sim \frac{1}{k} \qquad \frac{1}{k^2} \qquad \frac{1}{k^{1/3}}$$

2) линейная

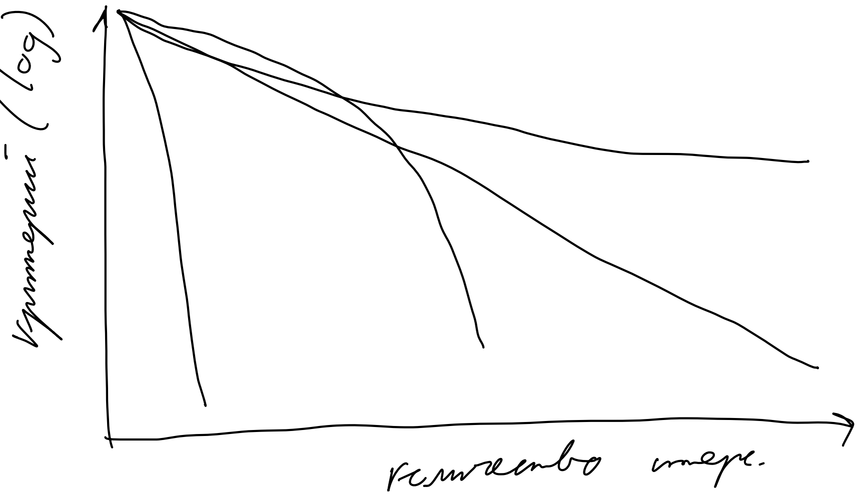
$$\|x^k - x^*\| \le C q^k \qquad 0 < C < \infty,\ 0 < q < 1$$

3) Сверхлинейная

$$\|x^k - x^*\| \le C q^{k^p} \qquad 0 < C < \infty,\ 0 < q < 1,\ p > 1$$

4) Квадратичная

$$\|x^k - x^*\| \le C q^{2^k} \qquad 0 < q < 1,\ 0 < C < \infty$$



$$Ax = b \qquad \underline{Ax - b} = 0$$

$$\min_x \|Ax - b\|_2^2 \qquad x^T \underbrace{A^T A} x - 2A^T b^T x$$

$$\nabla^2 f(x) \succeq 0 \qquad \nabla^2 f(x) \succeq \mu I$$

$$\| \qquad \nabla^2 f(x) - \mu I \succeq 0$$

$$A^T A$$

$$x^k \not\to x^* \qquad x^k \to x^*$$

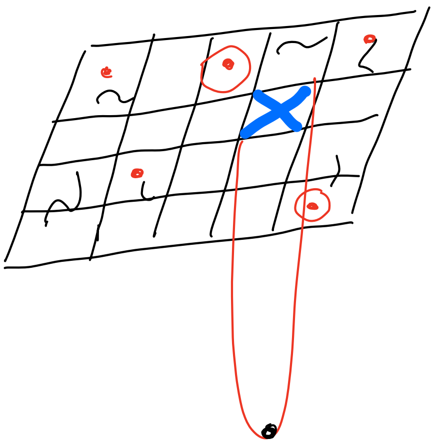
$$\|x^k - x^*\| \not\to 0 \qquad \|x^k - x^*\|^2 \to 0$$

$$f(x^k) - f(x^*) \to 0$$

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

$$f(x_{k+1}) \leq f(x_k) \qquad \gamma_k > 0$$

$\min f(x)$ $\quad k$



$k+1$ квадратик

Выпуклость:

$f(x)$ выпукла на $\mathbb{R}^d$, если $\forall\, x_0, x_1 \in \mathbb{R}^d$

$$f(x_1) \geq f(x_0) + \langle \nabla f(x_0); x_1 - x_0 \rangle$$

Сильная выпуклость

. . . . .

$$f(x_1) \geq f(x_0) + \langle \nabla f(x_0); x_1 - x_0 \rangle + \frac{\mu}{2} \| x_1 - x_0 \|^2$$

Липшицевость градиента (гладкость)

f имеет липшиц град. с конст. L, если $\forall\, x_0; x_1 \in \mathbb{R}^d$

$$\| \nabla f(x_0) - \nabla f(x_1) \| \leq L \| x_0 - x_1 \|$$

Th. $\forall\, x, y \in \mathbb{R}^d$ для ф. с L-липш. град. $\hookrightarrow$

$$| f(y) - f(x) - \langle \nabla f(x); y - x \rangle | \leq \frac{L}{2} \| x - y \|^2$$

Dox. be:

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + \tau(y-x)); y-x \rangle \, d\tau$$

$$= \langle \nabla f(x); y-x \rangle + \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x); y-x \rangle \, d\tau$$

$\leftarrow$

$$\left| f(y) - f(x) - \langle \nabla f(x); y-x \rangle \right| = \left| \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x); y-x \rangle \, d\tau \right|$$

$$\leq \int_0^1 \left| \langle \nabla f(x+\tau(y-x)) - \nabla f(x); y-x \rangle \right| d\tau$$

$$\leq \int_0^1 \| \nabla f(x+\tau(y-x)) - \nabla f(x) \| \, \|y-x\| \, d\tau$$

$$\leq \int_0^1 L \tau \|y-x\| \cdot \|y-x\| \, d\tau$$

$$= L\|y-x\|^2 \int_0^1 \tau \, d\tau = \frac{L\|y-x\|^2}{2}$$

f - выпукла и имеет лип. град.

а) $0 \le f(\tilde{y}) - f(\tilde{x}) - \langle \nabla f(\tilde{x}); \tilde{y} - \tilde{x} \rangle \le \frac{L}{2} \|\tilde{y} - \tilde{x}\|^2$

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

Док-во б:

$$\ni \varphi(y) = f(y) - \langle \nabla f(x); y \rangle$$ – вып. + L-л. гр.

$$y^* = x$$

$$\varphi(x) = \varphi(y^*) \le \varphi\left(y - \frac{1}{L} \nabla \varphi(y)\right) \overset{a)}{\le}$$

$\tilde{y} = y - \frac{1}{L} \nabla \varphi(y)$

$\tilde{x} = y$

$$\varphi\left(y - \frac{1}{L} \nabla \varphi(y)\right) - \varphi(y) - \left\langle \nabla \varphi(y); -\frac{1}{L} \nabla \varphi(y) \right\rangle \le \frac{1}{2L} \|\nabla \varphi(y)\|^2$$

$$\varphi\left(y - \frac{1}{L} \nabla \varphi(y)\right) \le \varphi(y) - \frac{1}{2L} \|\nabla \varphi(y)\|^2$$

$$\le \varphi(y) - \frac{1}{2L} \|\nabla \varphi(y)\|^2$$

$$\varphi(x) \le \varphi(y) - \frac{1}{2L} \|\nabla \varphi(y)\|^2$$

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

в). $$\frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \le \langle \nabla f(x) - \nabla f(y); x-y \rangle$$

Док. во в:

$$+\begin{cases} f(x) + \langle \nabla f(x); y-x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \le f(y) \\ f(y) + \langle \nabla f(y); x-y \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \le f(x) \end{cases}$$

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

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

сильно-выпуклой + L-липшиц град.

$$\leq \|x^k - x^*\|^2 - 2\gamma_k \left(\cancel{\frac{\mu}{2} \|x_k - x^*\|^2} + f(x_k) - f(x^*) \right) + \gamma_k^2 \|\nabla f(x_k)\|^2$$

1)
$$\leq \|x^k - x^*\|^2 - \gamma_k \mu \|x^k - x^*\|^2 + \gamma_k^2 L^2 \|x_k - x^*\|^2$$

$$= \underbrace{\left(1 - \mu\gamma_k + \gamma_k^2 L^2\right)} \|x_k - x^*\|^2$$

$$\gamma_k = \frac{\mu}{2L^2}$$

$$= \left(1 - \frac{\mu^2}{4L^2}\right) \|x_k - x^*\|^2$$

$$\underbrace{\|x^{k+1} - x^*\|^{2}}_{\varepsilon^2} \le \underline{\left(1 - \frac{\mu^2}{4L^2}\right)} \|x^k - x^*\|^2$$

$$\le \left(1 - \frac{\mu^2}{4L^2}\right)^2 \|x^{k-1} - x^*\|^2$$

$$\le \left(1 - \frac{\mu^2}{4L^2}\right)^{k+1} \|x^0 - x^*\|^2$$

$$k = \frac{4L^2}{\mu^2} \log \frac{\|x_0 - x^*\|^2}{\varepsilon^2}$$

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

$$+ \gamma_k^2 \|\nabla f(x_k) - \nabla f(x^*)\|^2$$

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

$$+ \gamma_k^2 \cdot 2L (f(x^k) - f(x^*))$$

$$= (1 - \mu\gamma_k)\|x^k - x^*\|^2 + \underbrace{\underline{2\gamma_k}\left(\gamma_k L - 1\right)\underline{\left(f(x_k) - f(x^*)\right)}}$$

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

$$\le \underline{(1 - \mu\gamma_k)} \|x^k - x^*\|^2 \qquad \gamma_k = \frac{1}{L}$$

$$= \left(1-\frac{\mu}{L}\right)\|x^k - x^*\|^2$$

$$k = \frac{L}{\mu}\log\frac{\|x_0 - x^*\|^2}{\varepsilon^2}$$

---

1) const
2) адаптивно - подбираем как L_k
3) $\frac{1}{k+1}$ $\qquad \frac{1}{\sqrt{k+1}}$
4) $\gamma_k = \arg\min\limits_{\gamma}\left(f(x_k - \gamma \nabla f(x_k))\right)$
5) Polyak $\qquad \gamma_k = \dfrac{f(x_k) - \overset{=0}{f(x^*)}}{\|\nabla f(x_k)\|^2}$
6) Armijo
Wolfe
Goldstein

$\Big\}$ дойдем к д/з