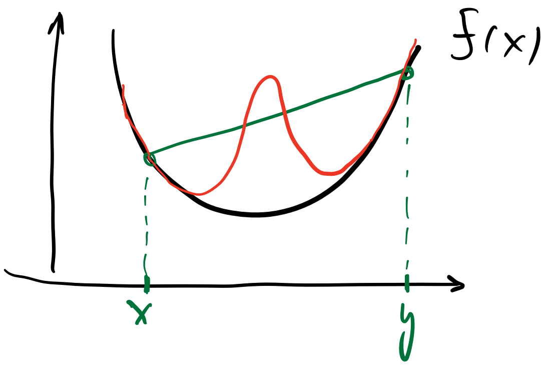
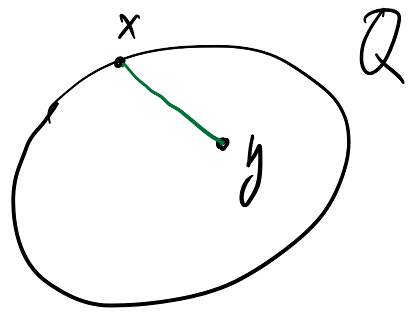
Лекция 1. Методы градиентного типа

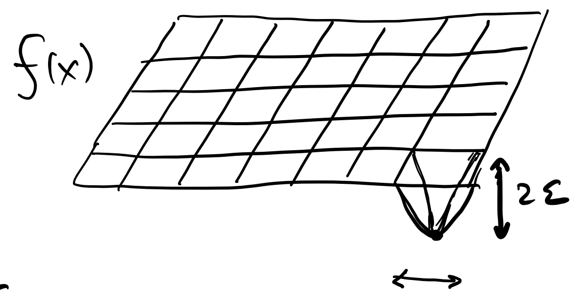
$$\boxed{\min_{x \in Q} f(x)}$$





$f(x)$ - невыпукл.

$d=2$



$\varepsilon > 0$ — точность реш. задачи по ф-ции

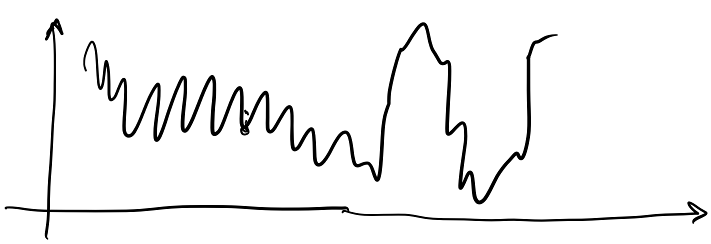
$f(x^N) - f(x_*) \le \varepsilon$

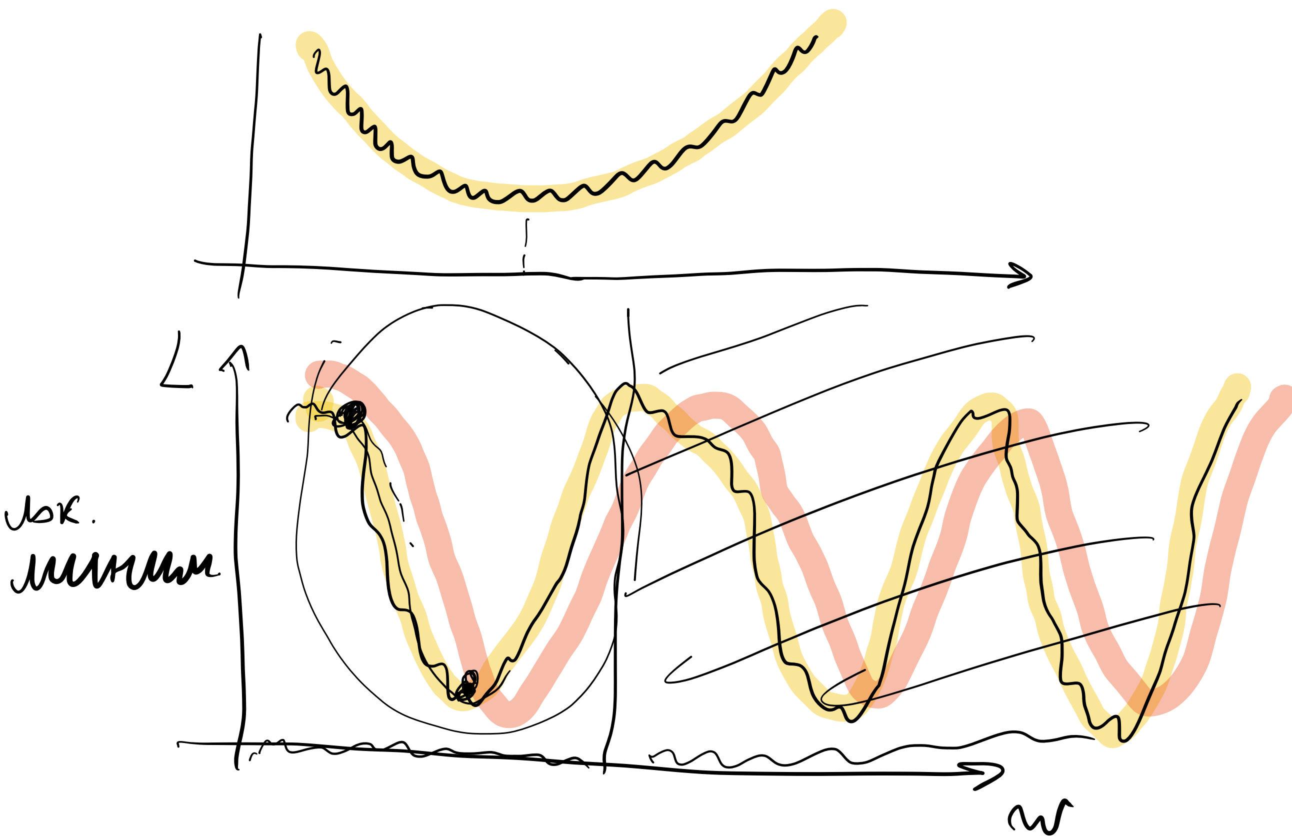
x_* - реш. зад.

1) $|f(y) - f(x)| \le M\|y-x\|_2$ $\quad \cdot \sim \varepsilon$

2) $\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y-x\|_2$ $\quad \cdots \sim \sqrt{\varepsilon}$

$\left(\frac{1}{\varepsilon}\right)^d$ - число точек.





Выпуклая оптимизация

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

Q - простая структура

$$Q = \{x\in\mathbb{R}^d \mid x \geq 0\}$$

$$x^{k+1} = x^k - h\nabla f(x^k) \qquad Q=\mathbb{R}^d$$

$$x^{k+1} = \underbrace{\Pi_Q (x^k -}_{\text{чем проект.}} h\underbrace{\nabla f(x^k)}_{\text{дороже}}) \qquad A = d\left[\;\right]^d$$

$$f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle$$

$\nabla f(x)$ $\quad O(d^2)$

Если $Q = \mathbb{R}^d_+$

$$x^{k+1} = [x^k - h\nabla f(x^k)]_+$$

покомп. срезка $O(d)$

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

$f(x)$ – вып., Q – вып.

1) $|f(y) - f(x)| \le M\|y-x\|_2$

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

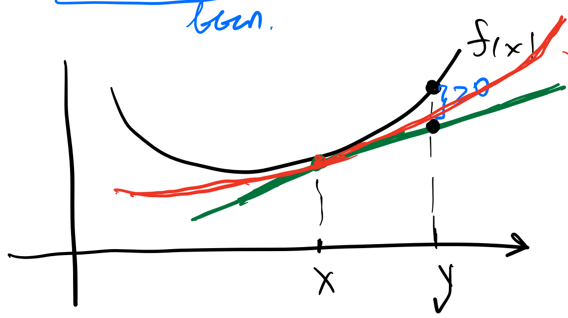
точка старта

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

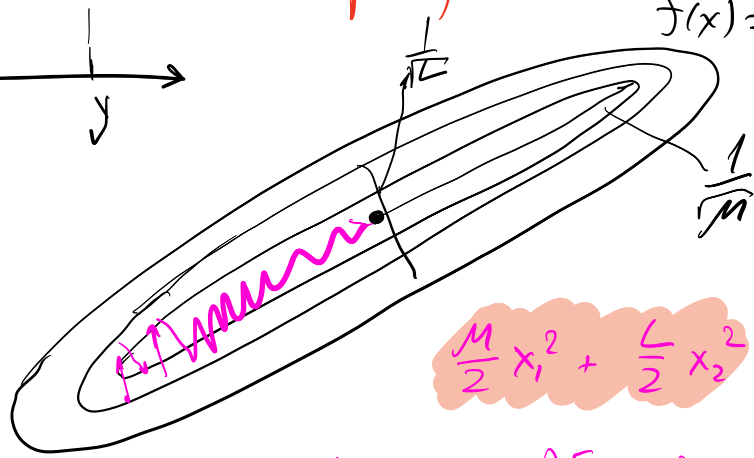
решение задачи $(*)$ наиболее близк. к x^0

$f(x)$ – μ-сильно вып. ф-ция.

$$\underbrace{f(y) \ge f(x) + \langle \nabla f(x), y-x\rangle}_{\text{вып.}} + \frac{\mu}{2}\|y-x\|_2^2$$



параб. с крив. μ.



$$\frac{\mu}{2}x_1^2 + \frac{L}{2}x_2^2$$

$$L = \max_x \lambda_{max}(\nabla^2 f(x))$$

$$\mu = \min_x \lambda_{min}(\nabla^2 f(x))$$

Классифик. по А.С. Немировскому '79

1) $\underline{N \le d}$ $\quad \varepsilon/M^2 \quad \vee \quad |x| \quad M=1$

Субградиент. метод. $\quad x^{k+1} = x^k - h\,\nabla f(x^k)$ $\quad \} \; Q = \mathbb{R}^d$

Поляк-Шор $\quad \bar{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k$

$$f(\bar{x}^N) - f(x_*) \lesssim \frac{MR}{\underbrace{\sqrt{N}}} = \varepsilon \quad (\square)$$

$$N \simeq \frac{M^2R^2}{\varepsilon^2} \quad \} \quad \text{для невып} \sim \left(\frac{1}{\varepsilon}\right)^d$$

$$\underbrace{x^{k+1} = x^0 + \mathrm{Lin}\{\nabla f(x^0), \nabla f(x^1), \ldots, \nabla f(x^k)\}}$$

f [руб] $\quad$ M [руб/кг]

x [кг] $\quad$ R [кг]

ε [руб]

$$N = \left(\frac{MR}{\varepsilon}\right)^\alpha, \quad \alpha = 2$$

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

$$\text{кг} = \text{кг} - \underbrace{[?] \cdot \frac{\text{руб}}{\text{кг}}}_{\text{кг}}$$

$$h \left[\frac{\text{кг}^2}{\text{руб}}\right] \to \frac{\varepsilon}{M^2}, \quad \cancel{h = \frac{R}{M\sqrt{N}}}$$

2) $$f(x^N) - f(x_*) \le \frac{LR^2}{N^2}$$

быстр. град. метод Нестерова (уск., импульсн.)

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

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

$f(x)$ - μ-сильно вып.

1) $$N = \frac{M^2}{\mu\varepsilon}$$

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

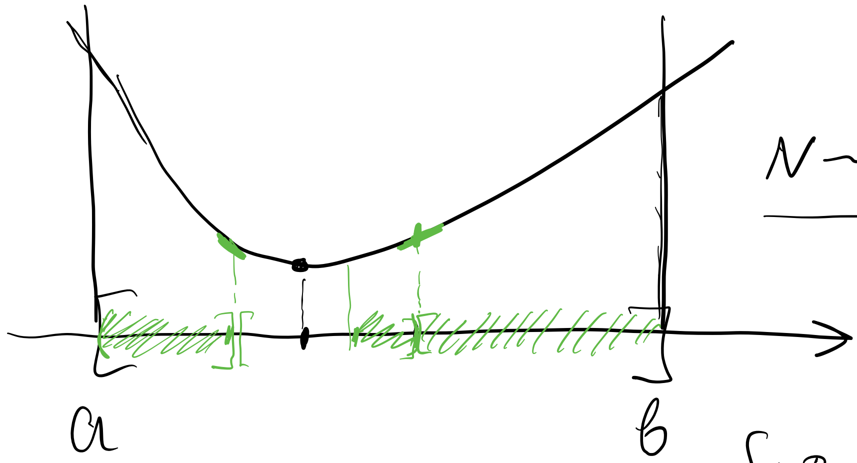
$\nabla f \to$ N $\boxed{N \le d}$	$\lvert f(y) - f(x)\rvert \le M\lVert y-x\rVert_2$	$\lVert \nabla f(y) - \nabla f(x)\rVert_2 \le L\lVert y-x\rVert_2$
$f(x)$ - вып.	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}}$
$f(x)$ - μ-сильн. вып.	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}} \ln \frac{\mu R^2}{\varepsilon}$

$N \le d$

А если $N \ge d$

$$\min_{x\in[a,b]} f(x)$$

$\underline{d = 1}$



$$N \sim \log_2 \frac{(b-a)}{\varepsilon}$$

$$N \cong d \cdot \ln\left(\frac{\Delta f}{\varepsilon}\right), \quad \Delta f \to f(x^0) - f(x_*)$$

$$\underbrace{\nabla f(x), \quad O(d^3)}$$

Докажем ($\square$)

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

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

$V(x) = \frac{1}{2}\|x - x_*\|_2^2$ – ф-ция Ляпунова

системы вида: $\frac{dx}{dt} = -\nabla f(x)$

Если $f(x)$ – вып. $\quad \frac{dV}{dt} = \langle \nabla V(x), \frac{dx}{dt} \rangle =$

$$= \langle \nabla f(x), x_* - x \rangle \overset{\text{по вып.}}{\leq}$$

$$\leq f(x_*) - f(x) \leq 0$$

$$\|x^{k+1} - x_*\|_2^2 \overset{x_* \in Q}{=} \|\pi_Q(x^k - h\nabla f(x^k) - x_*)\|_2^2 \overset{\text{св-ва проекции}}{\leq}$$

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

$$= \|x^k - x_*\|_2^2 - 2h\langle \nabla f(x^k), x^k - x_*\rangle + h^2 \underbrace{\|\nabla f(x^k)\|_2^2}_{\le M^2}$$

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

$$2h(f(x^k) - f(x_*)) \le 2h\langle \nabla f(x^k), x^k - x_*\rangle \le$$

вып.

$$\boxed{f(x_*) \ge f(x_k) + \langle \nabla f(x_k), x_* - x_k\rangle}$$

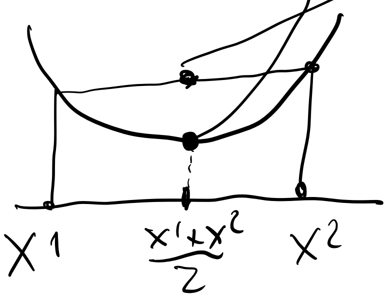
$$\le \|x^k - x_*\|_2^2 - \|x^{k+1} - x_*\|_2^2 + h^2 M^2 \quad \Big| \sum_{k=0}^{N-1}$$

$$0 \le \underbrace{2h \sum_{k=0}^{N-1} (f(x^k) - f(x_*))}_{} \le \overbrace{\|x^0 - x_*\|_2^2}^{R^2} - \|x^N - x_*\|_2^2 + Nh^2M^2 \quad \Big| \frac{1}{N}$$

$$\frac{1}{N}\sum_{k=0}^{N-1} f(x^k)$$

$\vee$ н-во Иенсена

$$f\left(\frac{1}{N}\sum_{k=0}^{N-1} x^k\right)$$



$$2h(f(\bar{x}^N) - f(x_*)) \le \frac{R^2}{N} + h^2M^2$$

$$\|x^N - x_*\|_2^2 \le R^2 + \underbrace{Nh^2M^2}_{R^2} \quad h \text{ выб}$$

$$\|x^N - x_*\|_2^2 \le 2R^2$$

$$f(\bar{x}^N) - f(x_*) \le \frac{R^2}{2Nh} + \frac{hM^2}{2} \quad \forall h \ge 0$$

$$\min_h \left(\frac{R^2}{2Nh} + \frac{hM^2}{2}\right) \to \psi(h)$$

$$\psi'(h) = 0 \qquad -\frac{R^2}{2Nh^2} + \frac{M^2}{2} \Rightarrow h = \frac{R}{M\sqrt{N}}$$

$$\frac{R^2 M\sqrt{N}}{2NR} + \frac{M^2 R}{2M\sqrt{N}}$$

$$\downarrow$$

$$\frac{MR}{\sqrt{N}}$$

$$\frac{MR}{\sqrt{N+1}} \leq f(\bar{x}^N) - f(x_*) \leq \frac{MR}{\sqrt{N}} = \varepsilon \Rightarrow$$

$$\Rightarrow N = \frac{M^2R^2}{\varepsilon^2}$$

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

$$h = \frac{R}{M\sqrt{N}}$$

$$\boxed{h = \frac{\varepsilon}{M^2}}$$

$$M_k \simeq \|\nabla f(x^k)\|_2$$

$$x^{k+1} = x^k - \frac{\varepsilon}{\|\nabla f(x^k)\|_2} \nabla f(x^k)$$

$$|f(y) - f(x)| \leq M\|y - x\|_2 \; ; \quad M = \max_{x \in Q} \|\nabla f(x)\|$$