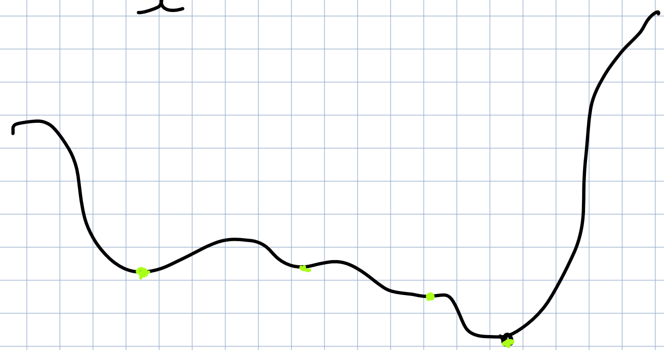
# Лекция 7. Градиентный спуск

Гасников Александр
@gasnikov
gasnikov@yandex.ru

## Θ. Коши

$$\min_x f(x)$$



$$\frac{dx}{dt} = -\nabla f(x)$$ система Коши

$x(t) \to x_{loc}$ – лок. мин. (типично)

Пример.

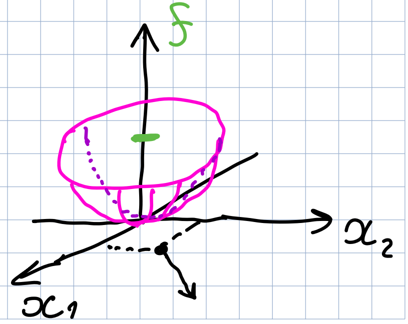
$$f(x_1, x_2) = x_1^2 + x_2^2$$

$\boxed{f(x_1, x_2) = const}$

$x_1^2 + x_2^2 = const$

если $x_1 = const$

$f(const, x_2) =$
$= const^2 + x_2^2$



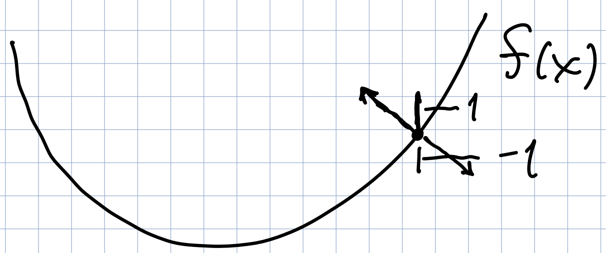
$$\nabla f(x_1, x_2) = (2x_1, 2x_2)^T \quad \left\{ \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^T \right.$$

$$\frac{dx}{dt} = l \Rightarrow \quad \frac{dx_1}{dt} = l_1, \quad \frac{dx_2}{dt} = l_2$$

$$\frac{df}{dt}(x_1, x_2) = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} = \langle \nabla f(x), \underset{\nabla f(x)}{\underbrace{l}} \rangle = \|\nabla f(x)\|_2^2$$

$$\frac{df}{dt}(x_1, x_2) = -\langle \nabla f(x), \nabla f(x) \rangle = -\|\nabla f(x)\|_2^2$$

$$l = -\nabla f(x)$$




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

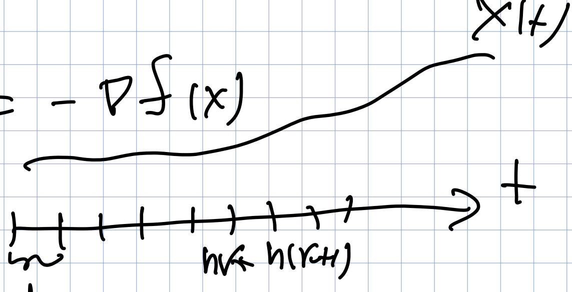



схема Эйлера

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

Градиентный спуск

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

$[?]\ \frac{руб}{кг} = \; = кг.$

$\Downarrow$

$[?] = \frac{кг^2}{руб}$

$$\min_x f(x)$$

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

$x_*$ — реш.

$x^0$ — точка старта

$\forall\ x, y$

$|x|,\ M = 1$

$$\boxed{|f(y) - f(x)| \le M\|y - x\|_2};$$

$$\|\nabla f(x)\|_2 \le M$$

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

$\left[\frac{руб}{кг.}\right]$

$|x|,\ L = \infty$

$x$ [кг]

$f$ [руб]

$\cancel{R\ [кг]}$

$M$ [руб/кг]

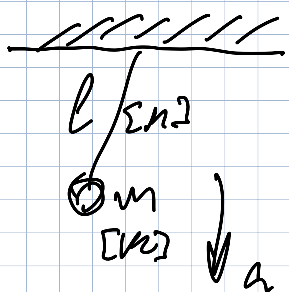
$L$ [руб/кг²]

$\varepsilon$ [руб]

$h$ [кг²/руб]

$$f(x^N) - f(x_*) \le \varepsilon$$ — точность по ф-ции

$\boxed{\text{П-теорема теории размерностей}}$




Задача о маятнике (Галилей)

$T$ [сек]

$g$ [м/сек²]

$$T / \sqrt{l/g}$$

$$\frac{[сек]}{\sqrt{м/(м/сек^2)}} = \frac{[сек]}{[сек]} = [\varnothing]$$

П-теорема

$$F\left(T/\sqrt{l/g}\right) = 1$$

$$T/\sqrt{l/g} = F^{-1}(1) = const$$

$$\boxed{T = const\sqrt{\frac{l}{g}}}$$

$\varepsilon$ [руб]

$M$ [руб/м]

$h$ [м$^2$/руб]

$$\frac{h}{\varepsilon/M^2} = \frac{м^2/руб}{руб/(руб/м)^2} = \frac{м^2/руб}{м^2/руб} = 1$$

Const = 1

$$\boxed{h = const\,\frac{\varepsilon}{M^2}}$$ субградиентный метод

$L$ [руб/м$^2$]

Const = 1

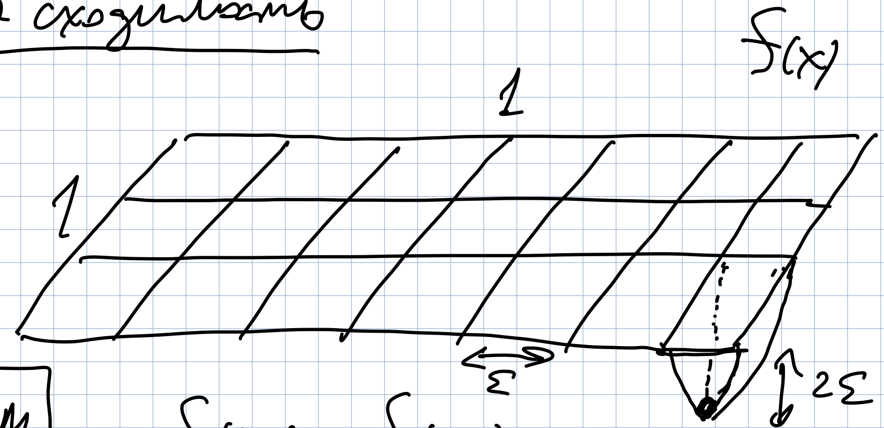
$$\boxed{h = const \cdot \frac{1}{L}}$$

$$x^{k+1} = x^k - \frac{\varepsilon}{M^2} \nabla f(x^k)$$

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

---

## Плохая сходимость

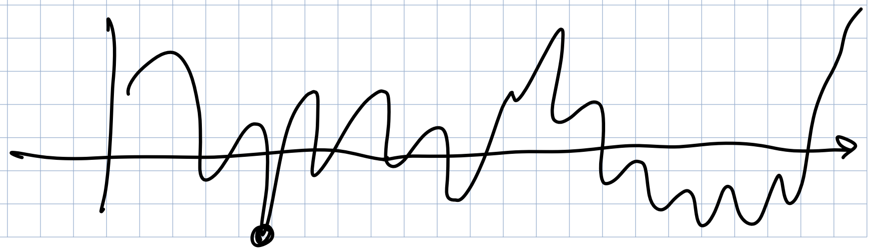



$$\boxed{\|\nabla f(x)\|_2 \le M}$$

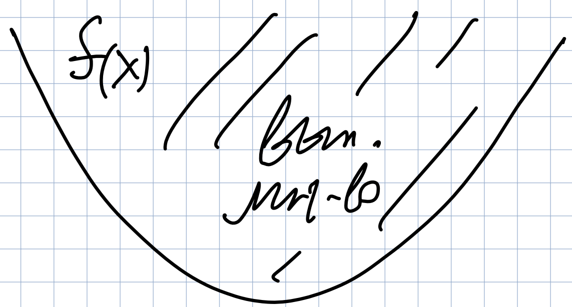
$$\underbrace{f(x_\varepsilon) - f(x_*) \le \varepsilon}$$

итераций $\sim \frac{1}{\varepsilon^2}$ $\qquad \min\limits_{x \in \mathbb{R}^2} f(x)$

итераций $\sim \frac{1}{\varepsilon^d}$ $\qquad \min\limits_{x \in \mathbb{R}^d} f(x)$

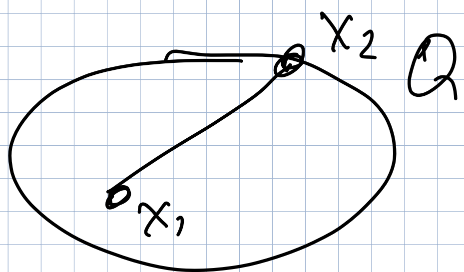


# Выпуклая оптимиз.




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

вып. вып.




из $x_1, x_2 \in Q \Rightarrow$
$\Rightarrow [x_1, x_2] \in Q$

$\forall x \in Q$

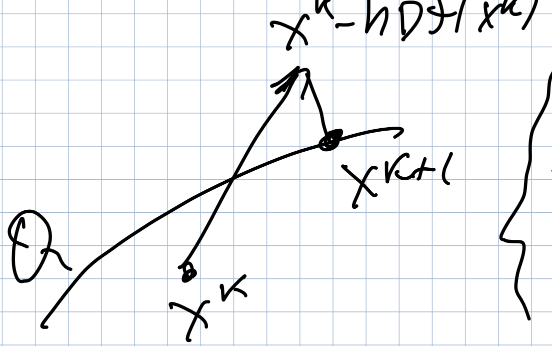
$\|\nabla f(x)\|_2 \le M$

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

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

$h \to \frac{\varepsilon}{M^2}$

проекция на мн-во Q




$$= \operatorname*{argmin}_{x \in Q} \{ h\langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2}\|x - x^k\|_2^2 \}$$

$$\frac{dx}{dt} = -\nabla f(x) \longrightarrow x(t)$$

$$V(x) = \frac{1}{2}\|x - x_*\|_2^2 \quad \Big\{ \; V(x) = f(x)$$

$$\frac{dV(x(t))}{dt} \le 0$$

$=0$ только в $x_*$

$$\frac{dV(x(t))}{dt} = \left\langle \nabla V(x), \frac{dx}{dt} \right\rangle =$$

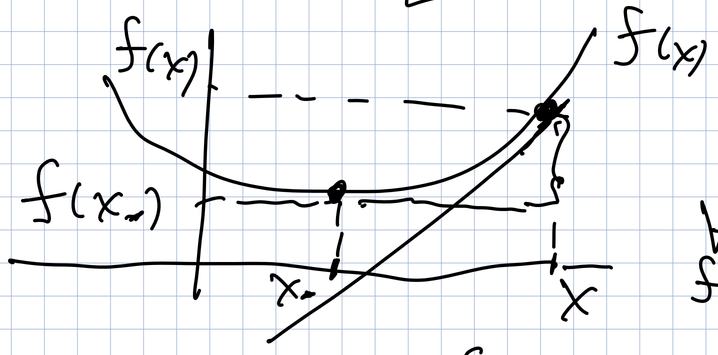
$$= \langle \nabla V(x), -\nabla f(x) \rangle =$$

$$= -\langle \nabla f(x), x - x_* \rangle =$$

$$= \underbrace{\langle \nabla f(x), x_* - x \rangle} \le 0$$

$= 0$

$x = x_*$




$\forall\, y, x$

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$$

$$\underbrace{\langle \nabla f(x), x_* - x \rangle \le f(x_*) - f(x)}_{0}$$

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

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

$$\pi_Q(x^k - h\nabla f(x^k)) - x_*$$

$$\qquad\qquad\qquad \| \qquad\qquad \overset{\Uparrow}{Q}$$

$$\pi_Q(x^k - h\nabla f(x^k) - x_*)$$

$$\|\pi_Q\|_2 \le 1$$

$$\|\pi_Q(y)\|_2 \le \|y\|_2 \qquad (*)$$

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

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

$$\underbrace{\langle \nabla f(x^k), x^k - x_* \rangle}_{\substack{\geqslant \text{ вып.} \\ f(x^k) - f(x_*)}} \leq \frac{1}{2h}\|x^k - x_*\|^2 - \frac{1}{2h}\|x^{k+1} - x_*\|_2^2 + \frac{h}{2}M^2$$

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

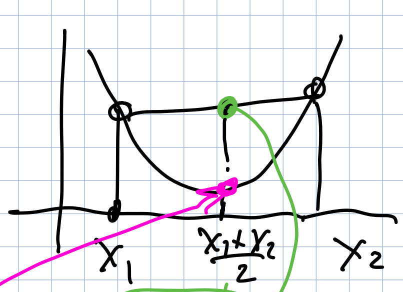
$$\underbrace{\frac{1}{N}\sum_{k=0}^{N-1} f(x^k)}_{\substack{\geqslant \text{ н-во} \\ \text{Иенсена}}} - f(x_*) \leq \frac{1}{2hN} R^2 - \cancel{\frac{1}{2h}\|x^N - x_*\|_2^2} + \frac{hM^2}{2}$$

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

$f(\bar{x}^N)$

$\|$

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

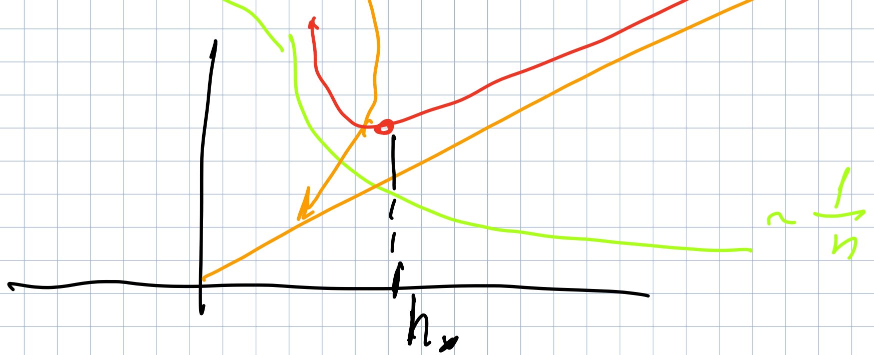



$$f\left(\frac{1}{N}\sum_{k=0}^{N-1} x^k\right) \leq \frac{1}{N}\sum_{k=0}^{N-1} f(x^k)$$

$$f(\bar{x}^N) - f(x_*) \le \underbrace{\frac{R^2}{2Nh} + \frac{M^2 h}{2}}_{\text{верно для любого } h>0}$$

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

$$F(h) = \underbrace{\frac{R^2}{2Nh}}_{} + \underbrace{\frac{M^2 h}{2}}_{} \to \min_h$$




$$F'(h) = 0$$

$$-\frac{R^2}{2Nh^2} + \frac{M^2}{2} = 0$$

$$h_* = \frac{R}{M\sqrt{N}}$$

$$F(h_*) = \frac{MR}{\sqrt{N}}$$

$$\min_{x \in Q} f(x)$$
$$\|\nabla f(x)\|_2 \le M$$
$$\|x^0 - x_*\|_2 \le R$$

$$f(\tilde{x}^N) - f(x_*) \le \frac{MR}{\sqrt{N}} = \varepsilon$$

$$\tilde{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k \qquad N = \frac{M^2R^2}{\varepsilon^2}$$

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

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

$$f(\tilde{x}^N) - f(x_*) \ge \frac{MR}{\sqrt{N+1}}$$

---

# Лекция 8 Ускоренный градиентный спуск

$$\min_{x \in \mathbb{R}^d} f(x) \quad \text{выпуклая}$$
$$\|x^0 - x_*\|_2 \le R$$
$$\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y-x\|_2 \quad (L)$$

$$f(x) = x^2 \qquad L_f = ?$$