Лекция 5. Метод условного градиента

X_* – реш. задачи.

$$\min_{x\in Q} f(x) \leftarrow \text{выпуклая}$$

$$R_p = \sup_{x,y\in Q} \|x-y\|_p$$

$$\|\nabla f(y) - \nabla f(x)\|_q \le L_p \|y-x\|_p, \quad \frac{1}{p}+\frac{1}{q}=1$$

$p \in [1,2]$

$$\gamma_k = \frac{2}{k+1}$$

$$\begin{cases} y^k = \operatorname*{argmin}_{y\in Q} \langle \nabla f(x^k), y\rangle \Rightarrow \langle \nabla f(x^k), y^k\rangle \le \\ \qquad \le \langle \nabla f(x^k), x_*\rangle \\ x^{k+1} = (1-\gamma_k)x^k + \gamma^k y^k \Rightarrow x^{k+1} - x^k = \\ \qquad = \gamma_k (y^k - x^k) \end{cases}$$

Теорема.

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

можно брать min по p

○ $\forall x,y\in Q$

$$0 \le f(y) - f(x) - \langle \nabla f(x), y-x\rangle \le \frac{L_p}{2}\|y-x\|_p^2$$

$y = x^{k+1}$, $x = x^k$

$$f(x^{k+1}) - f(x^k) \le \langle \nabla f(x^k), \underbrace{x^{k+1} - x^k}\rangle + \frac{L_p}{2}\|x^{k+1}-x^k\|_p^2 =$$

$$= \gamma_k \langle \nabla f(x^k), \underbrace{y^k - x^k}\rangle + \frac{L_p\gamma_k^2}{2}\underbrace{\|y^k-x^k\|_p^2}_{\le R_p^2} \le$$

$$\le \gamma_k \langle \nabla f(x^k), x_* - x^k\rangle + \frac{L_p\gamma_k^2 R_p^2}{2} \le$$

$$\le \gamma_k (f(x_*) - f(x_k)) + \gamma_k^2 \frac{L_pR_p^2}{2} \qquad (*)$$

$$\delta_k = \frac{f(x^k) - f(x_*)}{L_pR_p^2}$$

Из (*) $\Rightarrow \quad \delta_{k+1} \le (1-\gamma_k)\delta_k + \frac{\gamma_k^2}{2} = \left(1 - \frac{2}{k+1}\right)\delta_k + \frac{2}{(k+1)^2}$

$\delta_2 \le \frac{1}{2}$, дальше по индукции ○

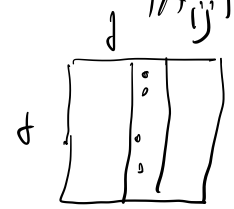
Пример.

$$\min_{\substack{x \in Q = S_d(1) \\ \| \\ \{x \in \mathbb{R}_+^d :\ \sum_{i=1}^d x_i = 1\}}} f(x) := \frac{1}{2}\langle x, Ax\rangle$$

$\downarrow$ симм. матр.

$A \succeq 0$

$|A_{ij}| \le M$



$= s$ ненул. эл-тов

$s \ll d$

① БГМ Нестерова

$d(x) = \sum_{i=1}^d x_i \ln x_i$

$V(y,x) = KL(y,x) = \sum_{i=1}^d y_i \ln(y_i/x_i)$

Дальше см. конец лекции 4.

$x^0 = \left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right) \quad f(x^N) - f(x_*) \le \frac{L_1 \tilde{R}_1^2}{N^2} = \varepsilon \quad N \sim \sqrt{\frac{L_1\tilde{R}_1^2}{\varepsilon}}$

$L_1 = M$

$R_1^2 = O(\ln d)$

стоимость итерации $O(s \cdot d) \to nnz(A) \leftarrow \nabla f(x) = Ax$

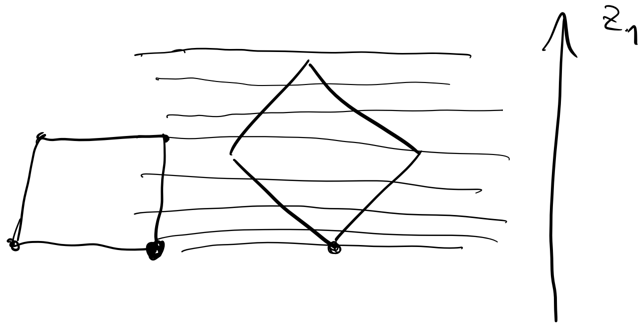
Общее число арифм. опер.

$O\left(sd\sqrt{\frac{M \ln d}{\varepsilon}}\right)_{\text{а.о.}}$ — БГМ

② Метод 1955 Франк-Вульф (условн. град.)

$\gamma_k = \frac{2}{k+1}$ $\qquad x^0 = (\underbrace{1, 0, \dots, 0}_{d})^T$

$$\begin{cases} y^k = \arg\min\limits_{y\in S_d(1)} \langle \nabla f(x^k), y\rangle = \langle Ax^k, y\rangle \\ x^{k+1} = (1-\gamma_k)x^k + \gamma^k y^k \end{cases}$$



$$y^k = (\underbrace{\overbrace{0,0,\dots,0,1}^{i_k},0,\dots,0}_{d})^T$$

$$i_k \in \operatorname{Argmin}_{i=\overline{1,d}} \frac{\partial f(x^k)}{\partial x_i}$$

$$f(x^N) - f(x_*) \le \frac{L_1 R_1^2}{N+2} = \varepsilon \qquad L_1 = M, \quad R_1^2 = 2$$

$$\Downarrow$$

$$Ax^{k+1} = (1-\gamma_k)\underbrace{Ax^k}_{\text{знаем}} + \gamma^k \underbrace{Ay^k}_{A^{i_k}} \qquad N \sim \frac{2M}{\varepsilon}$$

Упр. Покажите что стоимость итерации $O(s\log_2 d)$.

Первая итерация (препроцессинг) $O(d)$

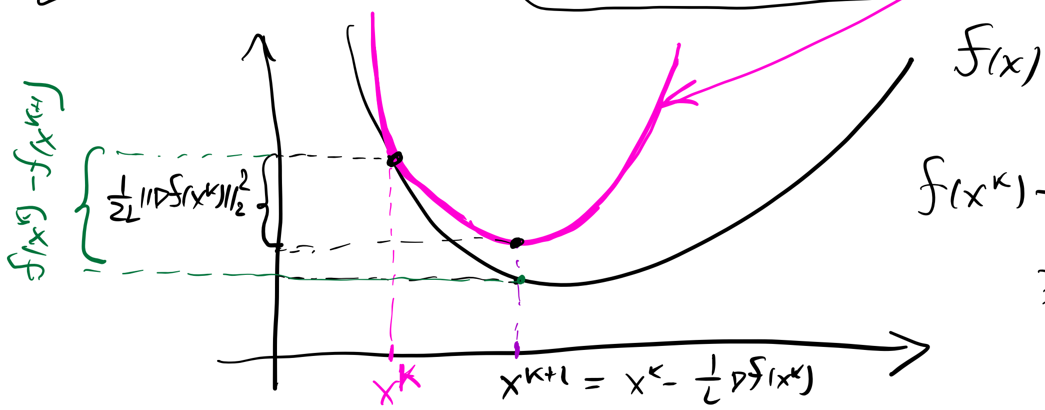
$$O(\underbrace{d}_{\text{препроц.}} + \underbrace{\frac{M}{\varepsilon}}_{\text{число итер.}} \underbrace{s\log_2 d}_{\text{стоим. итерации}}) \quad \text{а.о.}$$

н.о.:	БГМ	$\frac{d}{\sqrt{\varepsilon}}$	Условн. град.	$\frac{\log_2 d}{\varepsilon}$

Универсальный градиентный метод.

$$x^{k+1} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \{ \underbrace{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2}_{\bar{f}_{x^k}(x)} \} = x^k - \frac{1}{L} \nabla f(x^k)$$

$$\boxed{\bar{f}_{x^k}(x^{k+1}) \geq f(x^{k+1})}$$



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

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L \|y - x\|_2$$
$$\Downarrow$$
$$\lambda_{max}(\nabla^2 f(x)) \leq L$$

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2!} \langle \nabla^2 f(\underset{[x,y]}{\tilde{x}})(y - x), y - x \rangle \leq$$

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

$$\nabla^2 f(x) \preccurlyeq L I_d$$

Правило типа Армихо-Нестерова

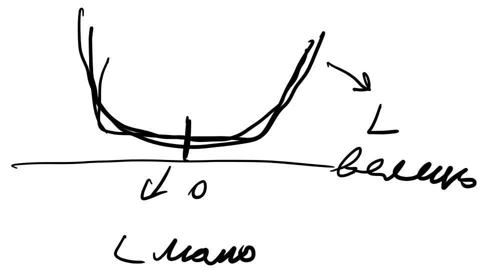
do $L_k := 2L_k$

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

until $\bar{f}_{x^k}(x^{k+1}) \geq f(x^{k+1})$

$\boxed{\text{в среднем } \leq 3 \text{ пересчета}}$

$$L_{k+1} = L_k/4$$

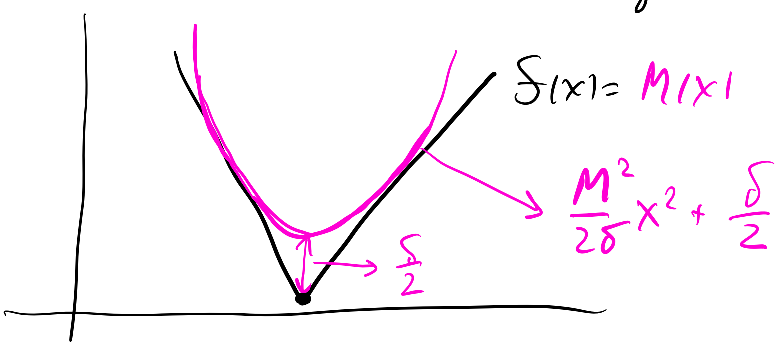


Пример $f(x) = x^{1000}$

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

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

Негладкий случай



$$|f(y) - f(x)| \leq M\|y-x\|_2$$

$\forall \delta > 0$

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

$\frac{M^2}{2\delta}$, $L(\delta) = \frac{M^2}{\delta}$

§2 порядок роста гладкости минимума

$$f(\bar{x}^N) - f(x_*) \le \underbrace{\frac{L(\delta)R^2}{N} + \frac{\delta}{2}}_{\varepsilon}$$

$$\frac{L(\delta)R^2}{N} = \frac{\varepsilon}{2} \qquad \frac{\varepsilon}{2} \qquad \delta = \varepsilon$$

$$N \sim \frac{2L(\varepsilon)R^2}{\varepsilon} = \frac{2M^2R^2}{\varepsilon^2}$$

$$x^{k+1} = x^k - \frac{1}{L}\nabla f(x^k) = x^k - \underset{\varepsilon/M^2}{\underset{\shortparallel}{h}}\nabla f(x^k)$$

$$\delta = \varepsilon, \quad L = \frac{M^2}{\delta} = \frac{M^2}{\varepsilon}$$

$$\bar{f}^{\delta}_{x^k}(x) = f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2}\|x - x^k\|_2^2 + \frac{\delta}{2}$$

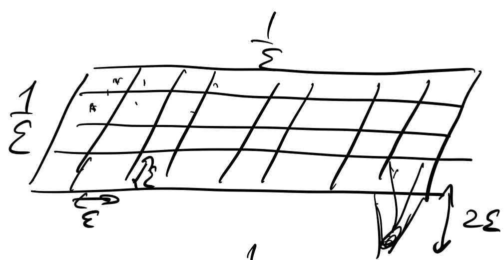
$$\delta = \varepsilon$$

Невыпуклая оптимизация

$$\min_x f(x)$$

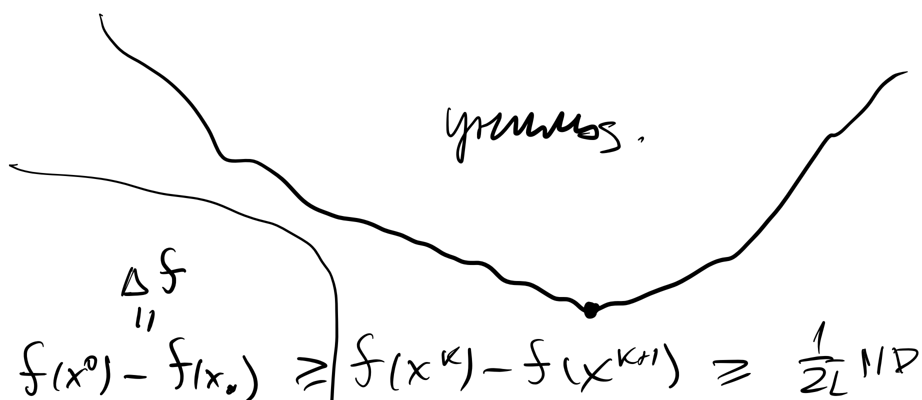
$$\underbrace{\|\nabla f(x^N)\|_2 \le \varepsilon}$$

шаг $\frac{1}{\varepsilon}$



$\sim \frac{1}{\varepsilon^2}$ $\quad d = 2$

$\sim \frac{1}{\varepsilon^d}$



$$\underset{f(x^0) - f(x_*)}{\underset{\shortparallel}{\Delta f}} \ge f(x^k) - f(x^{k+1}) \ge \frac{1}{2L}\|\nabla f(x^k)\|_2^2 \ge \frac{\varepsilon^2}{2L}$$

§1 пособия
МЦНМО

$$\frac{\varepsilon^2}{2L}N \simeq \Delta f$$

$$\boxed{N \simeq \frac{2L\Delta f}{\varepsilon^2}}$$

оптим. оценка

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