Лекция 7. Невыпуклая оптимизация

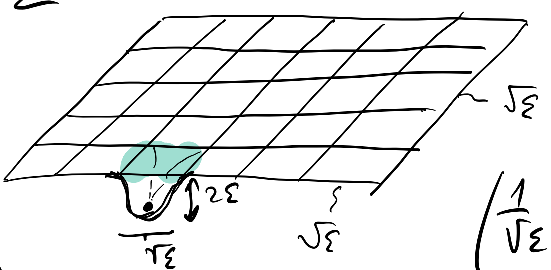
$$\min_{x \in \mathbb{R}^d} f(x)$$

$$f(x_\varepsilon) - f(x_*) \le \varepsilon$$

$$\sim \left(\frac{1}{\varepsilon}\right)^{\frac{d-1}{2}}$$

$d = 2$

$\frac{1}{\varepsilon^2}, \frac{1}{\varepsilon}, \frac{1}{\sqrt{\varepsilon}}$ → вып. оптим.



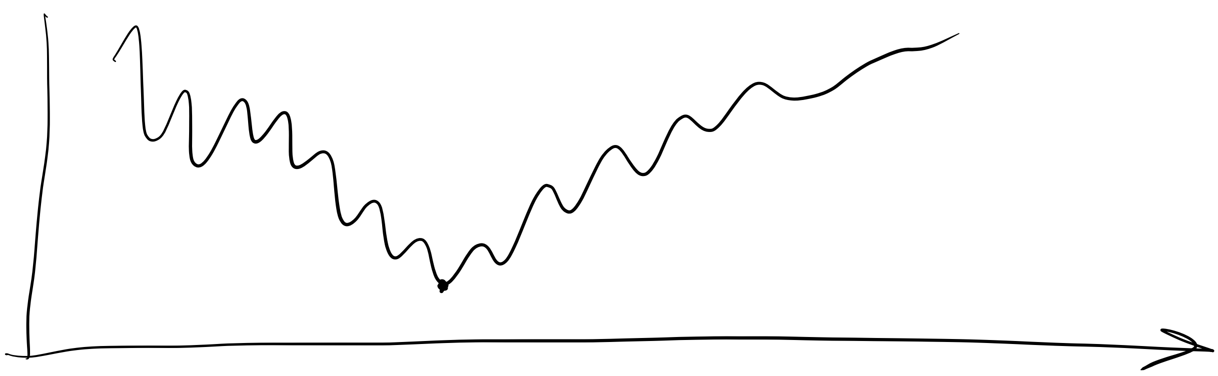
$$\left(\frac{1}{\sqrt{\varepsilon}}\right)^{2 = d}$$

$$\sim \left(\frac{1}{\sqrt{\varepsilon}}\right)^{d}$$

$d = 100$

$\varepsilon = 0.01 = 10^{-2}$

Сложность $\sim 10^{100}$

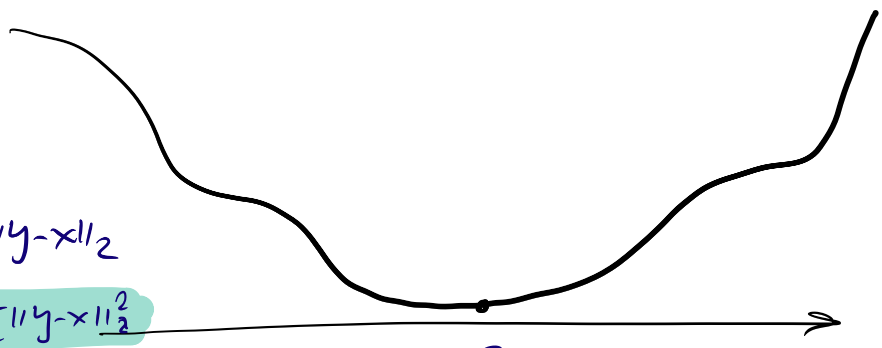


$$x^{k+1} = x^k - h \underset{=0}{\nabla f(x^k)}, \quad h = \frac{1}{L}$$

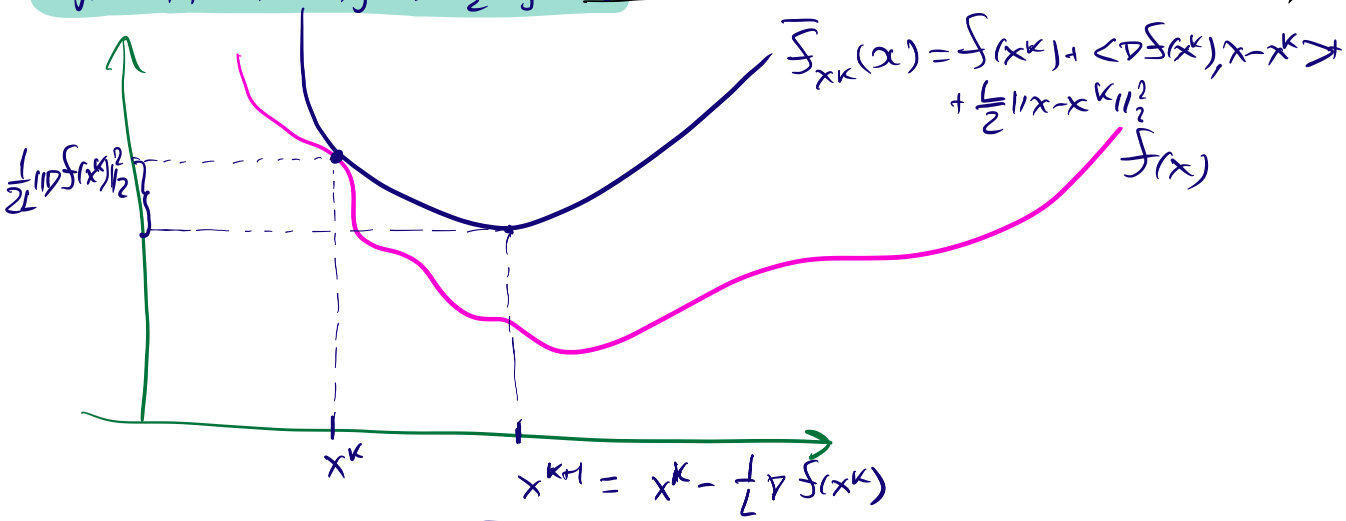
Можно застрять в седловой точке!

$$\|\nabla f(x^N)\|_2 \le \tilde{\varepsilon} \qquad (0)$$

унимодальная ф-ция



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



$$f(x^k) - f(x^{k+1}) \ge \frac{1}{2L}\|\nabla f(x^k)\|_2^2$$
$$\bar{f}_{x^k}(x^k) - \bar{f}_{x^k}(x^{k+1}) = \frac{1}{2L}\|\nabla f(x^k)\|_2^2$$

$$f(x^k) - f(x^{k+1}) \ge \bar{f}_{x^k}(x^k) - \bar{f}_{x^k}(x^{k+1}) = \frac{1}{2L}\|\nabla f(x^k)\|_2^2$$

$$f(x^k) - f(x^{k+1}) \ge \frac{\tilde{\varepsilon}^2}{2L}$$ до вып. кр. остановки (0)

$$f(x^0) - f(x_*) \overset{def}{=} \Delta f$$

N – оценка сверху на число итер. до вып (0)

За N итераций $\quad \frac{\tilde{\varepsilon}^2}{2L} N \simeq \Delta f$

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

$\nabla f(x)$ мин. град.

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

$\|\nabla f(x^N)\|_2 \le \tilde{\varepsilon}$ невып.

если гесс. гладк.

$\sim \frac{1}{\varepsilon^{8/5}}$

если использ. р-е произв

$\sim \frac{1}{\varepsilon^{\frac{p+1}{p}}}$

вып. (ускор.)
$\|\nabla f(x^N)\|_2 \le \tilde{\varepsilon}$
$N \sim \frac{1}{\sqrt{\varepsilon}}$

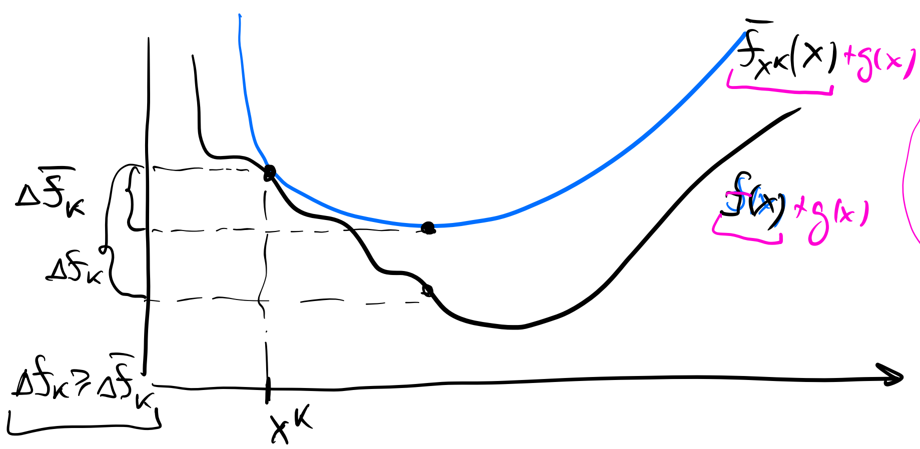
в независим. от степени гладк.

Композитная оптимизация

$$\min_{x \in \mathbb{R}^d} f(x) + \underbrace{g(x)}_{\text{простая ф-ция}}$$

$\lambda \|x\|_1$

$\frac{\lambda}{2} \|x\|_2^2$



$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 + g(x) \} \quad (\square)$$

Пример. LASSO

$Ax = b$

$\|x\|_0 \to \min$
$\Downarrow$ число ненул. эл-тов

$\|x\|_1 \to \min$
$Ax = b$
$\Downarrow$ $\|Ax - b\|_2^2 = 0$

В качестве $g(x)$ можно брать индик. ф-цию мн-ва Q

$\min\limits_{x \in Q} f(x)$

$\Updownarrow$

$\min\limits_{x \in \mathbb{R}^d} f(x) + g(x)$

$$\frac{1}{2}\|Ax - b\|_2^2 + \underbrace{\lambda \|x\|_1}_{g(x)} \to \min_{x \in \mathbb{R}^d}$$

$(\square)$ – реш. аналит., т.е. за $O(d)$

Ускор. мет-д (КЭМ)

$y^0 = z^0; \quad A_0 = \alpha_0 = 0$

$$\alpha_{k+1} = \frac{1}{2L} + \sqrt{\frac{1}{4L^2} + \alpha_k^2}$$

$$A_{k+1} = A_k + \alpha_{k+1}$$

$$x^{k+1} = \frac{\alpha_{k+1} z^k + A_k y^k}{A_{k+1}}$$

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

$x \in Q$

$V(x)$

В общем случае $V(x, z^k)$

$$y^{k+1} = \frac{\alpha_{k+1} z^{k+1} + A_k y^k}{A_{k+1}}$$

если не использ. ускор. $f(y^N) - f(x_*) \sim \frac{1}{\sqrt{N}}$

$$f(y^N) - f(x_*) \le \frac{4LR^2}{N^2}, \quad R^2 = \|y^0 - x_*\|_2^2$$

Пример. (Tomogravity model)

$$\sum_{k=1}^{d} x_i \ln x_i \to \min$$
$$Ax = b,\ x \in S_d(1)$$

$$F(x) = \boxed{\frac{1}{2}\underbrace{\|Ax-b\|_2^2}_{f} + \lambda \underbrace{\sum_{i=1}^{d} x_i \ln x_i}_{g(x)} \to \min_{x \in \underbrace{S_d(1)}_{Q}}}$$

гладк. выпукл.

не липш.

1-сильно выпукл. в 1-норме на $S_d(1)$

$$\nabla^2 g(x) = \begin{bmatrix} 1/x_1 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{x_d} \end{bmatrix}$$

$$\frac{\partial g(x)}{\partial x_i} = \ln x_i + 1$$

$\downarrow\ x_i \to 0+$

$$\min_{\|h\|_1 \le 1} \langle h, \nabla^2 g(x) h\rangle = 1 \qquad \sum_{i=1}^{d} x_i = 1$$

$h = (1/d, \dots, 1/d)$ $\qquad \frac{1}{x_i} \geq d$

$$\frac{\partial g}{\partial x_i} \uparrow -$$

$$N = \tilde{O}\left(\sqrt{\frac{L_f}{\mu_g \to \lambda}}\right)$$

$$L_f = \max_{\|h\|_1 \le 1} \langle hA^TA, h\rangle =$$

$$= \max_{i=1,\dots,d} \|A^{\langle i\rangle}\|_2^2$$

$\downarrow$ i-столбец

конст.

если прав. выбрать ✓

$$\frac{\lambda}{2}\|y^N - x_*\|_1^2 \le F(y^N) - F(x_*) \le \frac{4L_f V(x_*, y^0)}{N^2} \le$$

$$\le \frac{4L_f\, C \cdot \ln d \cdot \|x_* - y^0\|_1^2}{N^2}$$

$$\|y^N - x_*\|_1^2 = \frac{1}{2}\|y^0 - x_*\|_1^2$$

$$N \simeq \sqrt{\frac{L_f}{\lambda}}\, C \ln d$$

$y^0 := y^N$

↓

идея рестарта

Число рестартов

$\sim \log_2 \frac{1}{\varepsilon}$

Условие Поляка – Лоясиевича

Михаил Белкин

→ over param.

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

$$f(x) - f(x_*) \le \frac{1}{2\mu}\|\nabla f(x)\|_2^2 \quad (PL)$$

$f(x)$ – μ-сильно вып.

следствие

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

$g: \mathbb{R}^d \to \mathbb{R}^m$, $m \le d$

$$\frac{\partial g}{\partial x} = \left\| \frac{\partial g_i(x)}{\partial x_j} \right\|_{\substack{i = 1, m \\ j = 1, \dots, d}}$$

Цель: $g(x) = 0$ – совместн.

$$f(x) = \frac{1}{2}\|g(x)\|_2^2$$

$$f(x_*) = 0$$

Предполож. $\lambda_{min}\left(\frac{\partial g}{\partial x}\left(\frac{\partial g}{\partial x}\right)^T\right) \geq \mu$ (*)

$$\nabla f(x) = \nabla \tfrac{1}{2}\|g(x)\|_2^2 = \left(\frac{\partial g}{\partial x}\right)^T g(x)$$

$$\|\nabla f(x)\|_2^2 = \nabla f(x)^T \nabla f(x) = g(x)^T \frac{\partial g}{\partial x}\left(\frac{\partial g}{\partial x}\right)^T g(x)$$

Из (*) $\|\nabla f(x)\|_2^2 \geq \mu \|g(x)\|_2^2 \geq 2\mu f(x)$

$$f(x) - \underbrace{f(x_*)}_{0} \leq \frac{1}{2\mu}\|\nabla f(x)\|_2^2$$

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

$$f(x) - f(x_*) \leq \frac{1}{2\mu}\|\nabla f(x)\|_2^2$$

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

$$\|\nabla f(x^N)\|_2^2 \leq \varepsilon^2 = \frac{2L\Delta f}{N}$$

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

$$\|\nabla f(x^N)\|_2^2 \leq \frac{2L}{N}\underbrace{(f(x^0) - f(x_*))}_{\text{оцениваем по (PL)}} \leq$$

$$\leq \frac{2L}{N}\cdot\frac{1}{2\mu}\|\nabla f(x^0)\|_2^2$$

Если $N = 2\frac{L}{\mu}$, то

$$\|\nabla f(x^N)\|_2^2 \le \frac{1}{2}\|\nabla f(x^0)\|_2^2$$

$r \sim \log_2 \varepsilon^{-1}$

$$\|\nabla f(x^{N,r})\|_2^2 \le \varepsilon$$

$$\# \nabla f(x^k) \to 2\frac{L}{\mu}\log_2 \varepsilon^{-1}$$

<u>Пример.</u> (Нестеров-Скоков)

$x^0 = (-1, 1, \ldots, 1)$

$x_* = (1, 1, \ldots, 1)$

$g(x) = 0$:

$$\begin{cases} x_1 = 1 \\ x_2 = 2x_1^2 - 1 \\ x_3 = 2x_2^2 - 1 \\ \ldots \\ x_d = 2x_{d-1}^2 - 1 \end{cases}$$

$x_1 = x_2 = \ldots = x_d = 1$

$\|\nabla f(x^N)\|_2 \simeq 10^{-9}$, $\boxed{\begin{matrix} f(x^N) - f(x_*) = \\ \simeq 0{,}98 \\ \hline f(x^0) - f(x_*) = \\ = 1 \end{matrix}}$

$\frac{1}{2}\|g(x)\|_2^2 \to \min$

нуль по-другому

$$f(x) = \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{d-1}(x_{i+1} - 2x_i^2 + 1)^2 =$$

$$= \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{d-1}(x_{i+1} - \underbrace{P_2(x_i)}_{\text{мн-н Чебышёв}})$$

$$P_n(P_m(x)) = P_{mn}(x)$$

Фикс. $x_1 = c$, $x_{i+1}(c) = P_{2^i}(x_1) = P_{2^i}(c)$