Рандомизированные методы

$$f(x) := \frac{1}{m}\sum_{i=1}^{m} f_i(x) \to \min_{x\in\mathbb{R}^n}$$

M - Липш.

μ - сильно вып.

L - Липш. град.
μ - сильн. вып.

L - Липш. град.

// $f_k(x) = f(x, \xi_k)$

$\mathbb{E}_\xi f(x,\xi) \to \min_x$

M-С

$\boxed{m \sim \frac{M^2}{\mu\varepsilon}}$

$\underbrace{\frac{1}{m}\sum_{k=1}^{m} f(x,\xi^k)}_{f_k(x)}$

FGM
Быстрый град. метод

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

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

$$N = O\left(\sqrt{\frac{L}{\mu}}\ln\left(\frac{LR^2}{\varepsilon}\right)\right) \to \#\nabla f(x)$$

$$\#\nabla f_k(x) \to m\cdot N \to O\left(m\sqrt{\frac{L}{\mu}}\ln\frac{LR^2}{\varepsilon}\right) // \tilde{O} = O$$

с точн. до log.

SGD: $x^{k+1} = x^k - \frac{1}{\mu k}\nabla f_{\xi(k)}(x^k)$, $\xi(k)$ - с.в.

рандомизация (сумма)

$\mathbb{P}(\xi(k) = j) = \frac{1}{m}$

$\forall j = 1,\dots,m$

$\{\xi(k)\}_k$ i.i.d.

$$\mathbb{E} f(\bar{x}^N) - f(x_*) \le \varepsilon, \quad \bar{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k$$

$$N = O\left(\frac{M^2}{\mu\varepsilon}\right) \quad \text{vs} \quad O\left(m\sqrt{\frac{L}{\mu}}\ln\frac{LR^2}{\varepsilon}\right)$$

$\frac{M^2}{\mu\varepsilon}$

1) $\mathbb{E}[\nabla f_{\xi(k)}(x)] \equiv \nabla f(x)$!

2) $\mathbb{E}[\|\nabla f_{\xi(k)}(x)\|_2^2] \le M^2 \Leftarrow$ если $\|\nabla f_i(x)\|_2 \le M$

Variance Reduction

$$\nabla f_{\xi(k)}(x) \longrightarrow \underbrace{\nabla f_{\xi(k)}(x) - \nabla f_{\xi(k)}(x_*) + \overset{=0}{\nabla f(x_*)}}$$

$$\underbrace{\mathbb{E}_{\xi(k)} \nabla f_{\xi(k)}(x)}_{\nabla f(x)} - \underbrace{\mathbb{E}_{\xi(k)} \nabla f_{\xi(k)}(x_*)}_{-\cancel{\nabla f(x_*)} \to 0} + \underbrace{\cancel{\nabla f(x_*)}}_{=0}$$

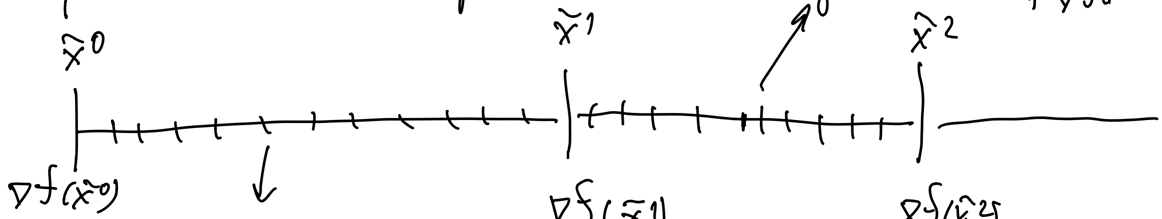
Выполнено условие 1) несмещенности

Если $x^k \to x_*$, то

$$\|\underbrace{\nabla f_{\xi(k)}(x^k) - \nabla f_{\xi(k)}(x_*)}_{\leq L^2\|x^k - x_*\|_2^2}\|_2^2 + \nabla f(x_*) \to 0$$

$$M^2 \sim L^2 \|x^k - x_*\|_2^2$$

Практичный вариант



$$x^{k+1} = x^k - h g_k$$

$$g_k = \left(\nabla f_{\xi(k)}(x^k) - \nabla f_{\xi(k)}(\tilde x^0) + \nabla f(\tilde x^0)\right)$$

\# $\nabla f(\tilde x) \to O\left(\ln\left(\frac{\Delta f}{\varepsilon}\right)\right)$ \# $\tilde x^j$

\# итераций, т.е. \# $\nabla f_{\xi(k)}(x^k) \to O\left(\sqrt{m\frac{L}{\mu}} \ln\frac{\Delta f}{\varepsilon}\right)$

\# $\nabla_\xi f_{\xi(k)}(x^k)$ $O\left(\left(m + \sqrt{m\frac{L}{\mu}}\right)\ln\left(\frac{\Delta f}{\varepsilon}\right)\right)$ vs $O\left(m\sqrt{\frac{L}{\mu}}\ln\frac{LR^2}{\varepsilon}\right)$

$m \simeq \frac{M^2}{\mu\varepsilon}$ $\searrow$ $\tilde{O}(m)$

Покомпонентные методы

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

μ-сильно вып.
L - липш. град.

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

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

$$N \simeq \frac{L}{\mu}\ln\frac{LR^2}{\varepsilon} \quad \cdot \# \nabla f(x^k)$$

Покомп. методы

$i(k)$ случ. выб. i.i.d.

$$\mathbb{P}(i(k) = j) = \frac{L_i}{\sum_{j=1}^n L_j}, \quad j = 1, \dots, n$$

$$x^{k+1}_{i(k)} = x^k_{i(k)} - \frac{1}{L_i}\underbrace{\nabla_i f(x^k)}$$

$$\left|\frac{\partial f}{\partial x_i}(x + \tau e_i) - \frac{\partial f}{\partial x_i}(x)\right| \le L_i \tau \qquad \frac{\partial f}{\partial x_i}(x^k)$$

$\tau \in R_+$

$$f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle, \quad A \succ 0$$

Nesterov'2012

$$L_i = A_{ii}$$

$$L = \lambda_{max}(A)$$

$$\begin{bmatrix} n & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \qquad L_1 = n, \; L_2 = 1, \; \dots, \; L_n = 1$$

$$L = n$$

$$\mathbb{E} f(x^N) - f(x_*) \le \varepsilon$$

$$N = n\frac{\bar{L}}{\mu}\ln\frac{LR^2}{\varepsilon}$$

$\frac{1}{n}\sum_{i=1}^n L_i$ V_i^2

GM

$$N \sim \frac{L}{\mu}\ln\frac{LR^2}{\varepsilon}$$

$$\#\nabla f(x) \simeq n^2$$

CGM

$$N \sim n\frac{\bar{L}}{\mu}\ln\frac{LR^2}{\varepsilon}$$

$$\nabla_i f(x) \simeq O(n)$$

$$\mu \ln\left(\sum_{i=1}^{n} \exp(x_i/\mu)\right) \longrightarrow \text{soft max}$$

$$\Big\downarrow \mu \to 0+$$

$$\max_{i=\overline{1,n}}\{x_i\}$$

$$\nabla f(x) = Ax$$

$$\nabla f(x+he_i) = A(x+he_i) = \overset{\text{известно}}{Ax} + h\underbrace{Ae_i}_{h\cdot A^{<i>}} =$$

$$g(Ax)$$

$$A^T \nabla g(Ax)$$

$$\mathbb{E}[\nabla_i f(x)] = \nabla f(x)$$

$$\mathbb{E}\left[\|\nabla_i f(x)\|_2^2\right] \simeq \frac{1}{n}\|\nabla f(x)\|_2^2 \qquad x \simeq x_*$$

$$\begin{array}{c} n \\ \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} n \end{array}$$

Метод Ньютона

$$f(x) \to \min_x \ - \ \mu\text{-сильно вып.} \qquad \nabla^2 f(x) \preccurlyeq L I_n, \quad \succcurlyeq \mu I_n$$

$$x^{k+1} = \arg\min_x \Big\{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2}\underbrace{\langle \nabla^2 f(x^k)(x-x^k), x - x^k \rangle}_{\leq L\|x-x^k\|_2^2} \Big\} = x^k - \frac{1}{L}\nabla f(x^k)$$

$$\dim x \lesssim 10^3 - 10^4$$

$$x^{k+1} = x^k - \underbrace{[\nabla^2 f(x^k)]^{-1}}_{O(n^{\log_2 7})} \nabla f(x^k)$$

$O(n^{\log_2 7})$ алг. Штрассена

$(\wedge n^3$

$O(n^{2.3\ldots})$ не практич.

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

$$= \|\nabla f(x^{k+1}) - \underbrace{(\nabla f(x^k) + \nabla^2 f(x^k)(x^{k+1} - x^k))}_{=0}\|_2 \le$$

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

$$\nabla^2 f(\bar{x})(y-x)$$

$$\|\nabla^2 f(y) - \nabla^2 f(x)\|_2 \le M_2 \|y-x\|_2$$

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2}$$

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

$$\le M_2 \|y-x\|_2^2$$

$$\|(\nabla^2 f(\hat{x}) - \nabla^2 f(x))(y-x)\|_2 \le \underbrace{\|\nabla^2 f(\hat{x}) - \nabla f(x)\|_2}_{\le M_2 \underbrace{\|\hat{x} - x\|_2}_{\le \|y-x\|_2}} \cdot \|y-x\|_2$$

$$\le M_2 \|x^{k+1} - x^k\|_2^2 =$$

$$= M_2 \|\underbrace{[\nabla^2 f(x^k)]^{-1}}_{\preceq \frac{1}{\mu} I_n} \nabla f(x^k)\|_2^2 \le \frac{M_2}{\mu^2} \|\nabla f(x^k)\|_2^2$$

$$\|\nabla f(x^{k+1})\|_2 \le \frac{M_2}{\mu^2} \|\nabla f(x^k)\|_2^2 \quad (*)$$

$$c_{k+1} \le \text{const} \cdot c_k^{\gamma}, \quad \gamma > 1$$

Если c_0 - дост. мало, то чтобы $c_N \le \varepsilon$ достаточно

$$N = O(\log(\log(c_0/\varepsilon)))$$

$$\frac{M_2}{\mu^2}\|\nabla f(x^k)\|_2 < 1 \quad (**)$$

$$\Downarrow$$

$$\frac{\mu}{2}\|x^k - x_*\|_2^2 \le f(x^k) - f(x_*) \le \underbrace{\frac{1}{2\mu}\|\nabla f(x^k)\|_2^2}_{\Uparrow (**)\ \frac{\mu^3}{2M_2^2}}$$

$$\|x^0 - x_*\|_2^2 \le \frac{\mu^2}{M_2^2}$$

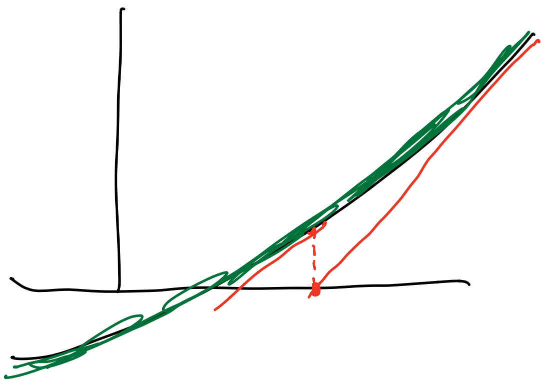
Nesterov-Polyak'06 (Griewank'81)

по сложн. почти экв. итер. м. Ньютона

$$x^{k+1} = \operatorname*{argmin}_x \Big\{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2}\langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle + \frac{M_2}{6}\|x - x^k\|_2^3 \Big\}$$

$$N \sim \left(\frac{M_2 R}{\mu}\right)^{\frac{2}{7}} \qquad f(x^N) - f(x_*) \lesssim \frac{M_2 R^3}{N^{7/2}} \qquad p = 2$$

$$N \sim \sqrt{\frac{L}{\mu}} \qquad \frac{1}{N^2} \qquad p = 1$$



$F(x) \to \min$

$F(x) = 0$

невырожд.

$\nabla F(x)$

$\frac{1}{2}\|F(x)\|_2^2 \to \min_x$

$\underbrace{\nabla f(x)} = \underbrace{F(x)} = 0$

Nesterov '06

$$x^{k+1} = x^k - \underbrace{[\nabla^2 f(x^k)]\nabla f(x^k)}_{[\nabla F(x^k)]^{-1} F(x^k)}$$

$\underline{F(x) = 0}$

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

$$\frac{1}{2}\|F(x^k) + \langle \nabla F(x^k), x - x^k \rangle\|_2^2 + \frac{\tilde{M}}{2}\|x - x^k\|_2^2 \to \min_x$$

Гаусс - Ньютон

$$\lambda_{\min}\left(\nabla F(x)(\nabla F(x))^T\right) \geq \mu$$