Лекция 8. Тензорные методы

Метод Ньютона

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

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 \} \longrightarrow \text{град. метод}$$

$$\frac{1}{2}\langle \nabla^2 f(x^k)(x-x^k), x - x^k \rangle \} \to \text{метод Ньютона}$$

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

$\nabla^2 f(x^k) \preccurlyeq L \cdot I_d$, т.е.

$L I_d - \nabla^2 f(x^k)$ – неотр. опр.

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

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

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

$\{ \|\nabla^3 f(x)\|_2 \le M_2$

$$\|A\|_2 = \max_{\|x\|_2 \le 1} \|Ax\|_2$$

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

метод Ньютон: $x^{k+1} = x^k - \underbrace{[\nabla^2 f(x^k)]^{-1}}_{\text{обратная матрица}} \nabla f(x^k)$

$M_2 < \infty$, x^0 – точка старта достаточно близка к x_*

$$\lambda_{min}(\nabla^2 f(x)) \geq \mu > 0 \Rightarrow [\nabla^2 f(x^k)]^{-1} \preccurlyeq \frac{1}{\mu} I_d \Rightarrow$$

$$\Rightarrow \|[\nabla^2 f(x^k)]^{-1}\|_2 \le \frac{1}{\mu}$$

Цель: Доказать, что (*) $\boxed{\|\nabla f(x^{k+1})\|_2 \le \frac{M_2}{\mu^2} \|\nabla f(x^k)\|_2^2}$

$$\frac{M_2}{\mu^2}\|\nabla f(x^k)\|_2 < 1 \quad \left\{ \begin{array}{l} \frac{\mu}{2}\|x-x_*\|_2^2 \le f(x) - f(x_*) \le \\ \le \frac{1}{2\mu}\|\nabla f(x)\|_2^2 < \frac{\mu^3}{2M_2^2} \end{array} \right.$$

$$\Downarrow$$

$$\|\nabla f(x^{k+1})\|_2 < \|\nabla f(x^k)\|_2 \qquad X = \{x : \|x - x_*\|_2^2 \le \frac{\mu^2}{M_2^2}\}$$

Если $x^k \in x \in X = \{x : \frac{M_2}{\mu^2}\|\nabla f(x)\|_2 < 1\} \Rightarrow$

$\Rightarrow x^{k+1} \in X$

$$c_k = \|\nabla f(x^k)\|_2$$

$$c_{k+1} \le \text{const} \cdot c_k^\gamma \quad , \gamma = 2$$

$$c_N \le \varepsilon \quad \gamma > 1$$

$$\underbrace{N = O(\log\log(c_0/\varepsilon))}$$

А.С. Немировский, конец 70х

быстрее не возможно (даже при $d = 1$ и бесконеч. гладк.) и сильно вып.

Докажем (*)

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

$$0 = \nabla f(x^k) + [\nabla^2 f(x^k)](x^{k+1} - x^k)$$

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

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

$$= \| \overbrace{\nabla^2 f(\tilde{x}^k)(x^{k+1}-x^k)}^{\shortparallel} - \nabla^2 f(x^k)(x^{k+1}-x^k)\|_2 \leq$$

$\tilde{x}^k \in [x^k, x^{k+1}]$

$$\leq \|\nabla^2 f(\tilde{x}^k) - \nabla f(x^k)\|_2 \, \|x^{k+1}-x^k\|_2 \leq$$

$$\leq M_2 \|\tilde{x}^k - x^k\|_2 \, \|x^{k+1}-x^k\|_2 \leq M_2 \|x^{k+1}-x^k\|_2^2$$

(*)

$$\|\nabla f(x^{k+1})\|_2 \leq M_2 \|\underbrace{x^{k+1}-x^k}_{[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)}\|_2^2 = M_2 \|[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)\|_2^2 \leq$$

$$\leq \frac{M_2}{\mu^2} \|\nabla f(x^k)\|_2^2$$

В 2006 Нестеров-Поляк
кубически-регуляризов. метод. Ньютона

Nesterov
Implementable
2020

$$x^{k+1} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \Big\{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2!}\langle \nabla^2 f(x^k)(x-x^k), x-x^k\rangle + \frac{M_2}{3!}\|x-x^k\|_2^3 \Big\}$$

$$f(x^N) - f(x_*) \leq C \frac{M_2 R^3}{N}, \quad \mu = 0$$

локально сходится $\quad c_{k+1} \leq \text{const} \cdot c_k^{4/3} \quad \mu > 0$

2018
2 сентября

Ускоренный мета-алгоритм (УМ)

$$\min_{x\in\mathbb{R}^d} F(x) := f(x) + \overset{\text{выпукл.}}{g(x)}$$

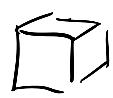
$p \geq 1$

$p = 2, 3$

$$\bar{f}^{\,p}_{\tilde{x}^k}(x) = f(\tilde{x}^k) + \langle \nabla f(\tilde{x}^k), x-\tilde{x}^k\rangle +$$

$$+ \frac{1}{2!}\langle \nabla^2 f(\tilde{x}^k)(x-\tilde{x}^k), x-\tilde{x}^k\rangle + \ldots +$$

$$+ \frac{1}{p!}\nabla^p f(\tilde{x}^k)\underbrace{(x-\tilde{x}^k)\ldots(x-\tilde{x}^k)}_{p}$$

 p-мерный тензор

$$\sum_{i_1+\ldots+i_d=p}\frac{\partial^p f(\tilde{x}^k)}{(\partial \tilde{x}^k_1)^{i_1}\ldots(\partial \tilde{x}^k_d)^{i_d}}[x-\tilde{x}^k]_1^{i_1}\ldots[x-\tilde{x}^k]_d^{i_d}$$

УМ

1: $A_0 = 0, \quad y^0 = x^0$

2: Определить (λ_{k+1}, y^{k+1}):

(Line search) $$\frac{1}{2} \leq \lambda_{k+1}\frac{M\|y^{k+1}-\tilde{x}^k\|_2^{p-1}}{p!} \leq \frac{p}{p+1}, \text{ где}$$

$$y^{k+1} = \arg\min_{y\in\mathbb{R}^d}\Big\{\overbrace{\bar{f}^{\,p}_{\tilde{x}^k}(y) + g(y) + \frac{M}{(p+1)!}\|y-\tilde{x}^k\|_2^{p+1}}^{\Omega^p(y)}\Big\},$$

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1} A_k}}{2}, \quad A_{k+1} = A_k + a_k$$

$$\tilde{x}^k = \frac{A_k}{A_{k+1}} y^k + \frac{a_{k+1}}{A_{k+1}} x^k$$

3: $x^{k+1} = x^k - a_{k+1} \nabla f(y^{k+1}) - a_{k+1} \nabla g(y^{k+1})$

Теорема. Пусть $\|D^p f(y) - D^p f(x)\|_2 \le M_p \|y-x\|_2$,

(для УМа) при $p \ge 3$ $H \ge (p+1) M_p$. Тогда

Nesterov 2018

$$F(y^N) - F(y_*) \le \frac{c_p H R^{p+1}}{N^{\frac{3p+1}{2}}}, \text{ где}$$

$$c_p = 2^{p-1} (p+1)^{\frac{3p+1}{2}} / p!, \quad R = \|x^0 - x_*\|_2.$$

$$\frac{3p+1}{2} \xrightarrow{p=1} 2 \sim \frac{1}{N^2}$$

$$\frac{3p+1}{2} \xrightarrow{p=2} \frac{7}{2} \sim \frac{1}{N^{7/2}}$$

$$\frac{3p+1}{2} \xrightarrow{p=3} 5 \sim \frac{1}{N^5}$$

Если $p=1$, то (Line Search) отсутствует.

гарантируем выпуклость задачи поиска y^{k+1}.

Nesterov'2018 при $p=2,3$ задача поиска УКН почти экв. по сложн. шагу методу Ньютона (если $g \equiv 0$).

Замечание 1. $\nabla^2 f(x) V \simeq$

$$\simeq \frac{\nabla f(x+\tau V) - \nabla f(x)}{\tau}$$

$\nabla^3 f(x) V_1 V_2 \rightarrow$ не нужно считать $\nabla^3 f(x)$

Слайдинг (градиент)

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

$\swarrow$ $\nabla f(x)$, L_f

$\searrow$ $\nabla g(x)$, L_g

дорогой $\rightarrow \# \nabla f$

дешёв. $\rightarrow \# \nabla g$

$\# \nabla f, \# \nabla g \longrightarrow N \sim \sqrt{\frac{(L_f + L_g) R^2}{\varepsilon}}$

$L_g \geqslant L_f$

$\# \nabla f \rightarrow N_f \sim \sqrt{\frac{L_f R^2}{\varepsilon}}$

$\# \nabla g \rightarrow N_g \sim \sqrt{\frac{L_g R^2}{\varepsilon}}$

Применяем УМ $p = 1$

$$\min_y \underbrace{\left\{ f(\tilde{x}^k) + \langle \nabla f(\tilde{x}^k), y - \tilde{x}^k \rangle + g(y) + \frac{H}{2}\|y - \tilde{x}^k\|_2^2 \right\}}$$

$\sqrt{\frac{HR^2}{\varepsilon}}$ - число итераций. $\to \# \nabla f$

$\sqrt{\frac{HR^2}{\varepsilon}}$ (внеш. итер.) $* \sqrt{\frac{L_g}{H}} \ln \frac{HR^2}{\varepsilon} \to \# g$

$H = L_f$

$$\sqrt{\frac{L_f R^2}{\varepsilon}} \to \# \nabla f$$

$$\sqrt{\frac{L_f R^2}{\varepsilon}} \sqrt{\frac{L_g}{L_f}} \ln \frac{L_f R^2}{\varepsilon} = \sqrt{\frac{L_g R^2}{\varepsilon}} \ln \frac{L_f R^2}{\varepsilon} \to \# \nabla g$$

<u>Замечание 2</u>.

y^{k+1} — идеальн. реш.

$$\|\nabla \Omega^p(\tilde{y}^{k+1})\|_2 \leq$$
$$\leq \frac{1}{4p(p+1)} \|\nabla F(\tilde{y}^{k+1})\|_2$$

$p = 1$

точка сопряж для внутр задачи

$$\|\tilde{y}^{k+1} - y^{k+1}\|_2 \le \underbrace{\frac{\mu}{3\mu + 2L_g}\|\tilde{x}^{k} - y^{k+1}\|_2}_{\text{вычисл.}}$$