Лекция 9. Приложения ЧМа

Покомпонентный метод.
Nesterov' 2010

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

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

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

$f(x)$ — μ-сильно выпукл. : $\forall x,y \rightarrow f(y) \ge f(x) + \langle \nabla f(x), y-x\rangle + \frac{\mu}{2}\|y-x\|_2^2$

$$\mu I_d \preceq \nabla^2 f(x) \preceq L I_d$$

$$f(x^N) - f(x_*) \le \varepsilon, \quad N = O\left(\frac{L}{\mu} \ln \frac{\Delta f}{\varepsilon}\right)$$

$$\forall i = 1, \dots, d \rightarrow \left| \frac{\partial f}{\partial x_i}(x + h e_i) - \frac{\partial f}{\partial x_i}(x) \right| \le L_i h, \quad h \ge 0.$$

$$e_i = (\underbrace{0, 0, \dots, 0, 1}_{i}, 0, \dots, 0)$$

$$\max_{x \in \mathbb{R}^d} \frac{\partial^2 f}{\partial x_i^2}(x)$$

Пример. $f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle$

$$A = \begin{bmatrix} L_1 & & * \\ & \ddots & \\ * & & L_d \end{bmatrix}$$

$$p_i = \frac{L_i}{\sum_{j=1}^{d} L_j}, \quad \mathbb{P}(i = i') = p_{i'}, \quad i' = 1, \dots, d$$

RCD

шаг k:

1: Случ. выб. i_k согласно распр. $\{p_i\}_{i=1}^d$

2: $x_{i_k}^{k+1} = x_{i_k}^k - \frac{1}{L_{i_k}} \frac{\partial f}{\partial x_{i_k}}(x^k)$

$x_j^{k+1} = x_j^k \quad , \quad j \neq i_k$

Напоминание

Если $f(x)$ – μ-сильно вып, то

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

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

$$f(x_*) = \min_y f(y) \geq \min_y \{ f(x) + \langle \nabla f(x), y-x \rangle + \frac{\mu}{2}\|y-x\|_2^2 \} =$$

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

$\nabla f(x) + \mu(\hat{y} - x) = 0$

$y = x - \frac{1}{\mu}\nabla f(x)$

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

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

$(1) + (2) \Rightarrow$

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

$$\overset{(1)}{\leq} -\frac{\mu}{L}(f(x^k) - f(x_*))$$

$$f(x^{k+1}) - f(x_*) \leq \left(1 - \frac{\mu}{L}\right)(f(x^k) - f(x_*))$$

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

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

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

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

$$f(x^N) - f(x_*) \le \left(1 - \frac{\mu}{L}\right)^N \left(f(x^0) - f(x_*)\right)$$

$$f(x^N) - f(x_*) \le \varepsilon, \quad N = O\left(\frac{L}{\mu}\ln\frac{\Delta f}{\varepsilon}\right)$$

∎

(2') $$f(x^k) - \mathbb{E}_{i_k} f(x^{k+1}) = \sum_{i=1}^{d} p_i \left(f(x^k) - f(x^{k+1})\right) \ge$$

расчит. алгоритмом RCD

$$\ge \sum_{i=1}^{d} \frac{p_i}{2L_i}\left|\frac{\partial f}{\partial x_i}(x^k)\right|^2 = \frac{1}{2\left(\sum_{j=1}^{d} L_j\right)}\|\nabla f(x^k)\|_2^2$$

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

Для RCD: $$N = O\left(\frac{\sum_{j=1}^{d} L_j}{\mu}\ln\frac{\Delta f}{\varepsilon}\right)$$

Для GM: $$N = O\left(\frac{L}{\mu}\ln\frac{\Delta f}{\varepsilon}\right)$$

град. спуск

$$L_j \le L$$

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

$$A = \left[A^{\langle 1\rangle} \quad A^{\langle d\rangle}\right]$$

$$\nabla f(x) = \underbrace{Ax} - b$$

softmax$(y) =$

$$= \ln\left(\sum_{i=1}^{m} \exp(y_i)\right)$$

$$y_i = \langle A_i, x\rangle$$

$$\frac{\partial f}{\partial x_i} = A^{\langle i\rangle} x_i - b_i$$

soft-max

$$f(x) = F(Ax)$$

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

вообще говоря меняются все комп.

для GM Ax^k , Ax^{k+1}

знание этого

не упрощает вычисление этого

для RCD

$$Ax^k$$

$$Ax^{k+1} = A\left(x^k - \frac{1}{L_i}\begin{pmatrix} 0 \\ \vdots \\ \partial f/\partial x_i(x^k) \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) =$$

знаю

$$= \underbrace{Ax^k}_{\text{посчитали}} - \underbrace{\frac{1}{L_i} A^{\langle i\rangle} \frac{\partial f}{\partial x_i}(x^k)}_{O(d)}$$

можно пересчитать за $O(d)$

$$\nabla f(x) = A^T \nabla F(Ax)$$

$$\frac{\partial f}{\partial x_i}(x) = \langle A^{\langle i\rangle}, \overbrace{\nabla F(Ax)}^{O(d) \text{ по предполож.}}\rangle$$

$O(d)$ пересч.

Вывод:

	стоимость итерации	N (число итер.)
RCD	T/d	$\sim\tilde{O}\left(\frac{\sum_{i=1}^{d} L_j}{\mu}\right)$
GM	T	$\tilde{O}\left(\frac{L}{\mu}\right)$

$$L = \lambda_{max}(A) \geq \lambda_{max}(\mathbb{1}_d \mathbb{1}^T) = d \qquad L_j \leq L$$

$$A = A^T \longrightarrow 1 \leq A_{ij} \leq 2$$
$$A \succ 0$$

$$\underset{\substack{\parallel \\ A_{jj}}}{L_j} \leq 2 \quad , \quad L \geq d$$

$$\frac{1}{d}\sum_{j=1}^{d} L_j \leq 2 \quad , \quad L \geq d$$

Задача: Предложить ускоренный RCD.
(Nesterov'10)

<u>Идея</u>: использовать УМ

$$\min_{x \in \mathbb{R}^d} f(x) \leftarrow \mu\text{-сильно вып}$$

самая затратная часть УМ

$$\min_{x \in \mathbb{R}^d} \left\{ f(x) + \frac{H}{2}\|x - x^k\|_2^2 \right\} \qquad (*)$$

(*) надо решать $\sqrt{\frac{H}{\mu}} \ln \frac{\Delta f}{\varepsilon}$ раз.
(рестартованный УМ)

решаем серую задачу $\rightarrow$ RCD $\frac{\sum_{j=1}^{d}(L_j+H)}{\mu+H} \ln \frac{\Delta f'}{\varepsilon'(\varepsilon)}$

$\varepsilon'(\varepsilon) \sim \varepsilon^2$

Общее число итераций

$$\sqrt{\frac{H}{\mu}} \cdot \frac{\sum_{j=1}^{d}(L_j+H)}{\mu+H} \rightarrow \min_H$$

$\mu \ll H$

$$\Updownarrow$$

$$\sqrt{\frac{H}{\mu}} \cdot \frac{\sum_{j=1}^{d}(L_j+H)}{H} \rightarrow \min_H$$

$$H = \frac{1}{d}\sum_{j=1}^{d} L_j = \bar{L}$$

Общее число итераций

$$\sqrt{\frac{\bar{L}}{\mu}} \cdot \frac{2d\bar{L}}{\bar{L}} \sim d\sqrt{\frac{\bar{L}}{\mu}}$$

Итого $N = O\left(d\sqrt{\frac{\bar{L}}{\mu}} \ln^2 \frac{\Delta f}{\varepsilon}\right)$

Для ARCD

$$N = \tilde{O}\left(d\sqrt{\frac{\bar{L}}{\mu}}\right)$$

Для AGM

$$N = \tilde{O}\left(\sqrt{\frac{L}{\mu}}\right)$$

$L \geq \bar{L}$

$\overline{L} \leq 2 \quad , \quad L \geq d$

$$\frac{\partial f(x)}{\partial x_i} \simeq \frac{f(x+\tau e_i) - f(x)}{\tau}$$