

$$\min_{x \in \mathbb{R}^d} \left[f(x) := \frac{1}{n} \sum_{i=1}^n f_i(x) \right]$$

SGD:

$$x^{k+1} = x^k - \gamma \nabla f_{i_k}(x^k)$$

↑
случ. выбор и субградиент

Проблема: сходимость к оптимальности

$$x = x^* - \gamma \nabla f_{i_k}(x^*)$$

↑
градиент не равен нулю

≠ 0

$\nabla f(x^*) = 0$

$\nabla f_{i_k}(x^*) \neq 0$

$$x^{k+1} = x^k - \gamma \nabla f_{i_k}(x^k)$$

↑
заменим на $g^k: g^k \rightarrow 0$
 $x^k \rightarrow x^*$

$\nabla f_{i_k}(x^*) \neq 0$

$x^k \rightarrow x^*$

SAGA

g_i^k — нерегуляризованный наклон

$$\nabla f_5(x^k), \nabla f_{21}(x^{k+1})$$

$$x^k \leftarrow x^{k+1}$$

выбираю 2, 3... выбираю любое 1 BSGD

$$g_i^k = \begin{cases} \nabla f_{i_k}(x^k), & i = i_k \\ g_i^{k-1}, & \text{иначе} \end{cases}$$

$$1) \quad X^{k+1} = X^k - \gamma \cdot \frac{1}{n} \sum_{i=1}^n g_i^k$$

$$g_i^k = \nabla f_i(x^{k-\tau})$$

$$x^k \rightarrow x^* \quad g_i^k \rightarrow \nabla f_i(x^*)$$

$$\frac{1}{n} \sum g_i^k \rightarrow \frac{1}{n} \sum \nabla f_i(x^*) = \nabla f(x^*) \rightarrow 0$$

$$K = O\left(\left[\frac{L}{\mu} + n\right] \log \frac{1}{\varepsilon}\right) \text{ итераций}$$

$$K = O\left(\frac{L}{\mu} \log \frac{1}{\varepsilon}\right)$$

← константа $O(1)$

↑
звездочка (константа $O(n)$)

⊕ скалярное, так $\gamma \in \mathbb{D}$ (но итераций)

⊕ итераций константа $O(1)$

⊕ скалярное k -мощный процессор

⊖ $O(nd)$ минимум

⊖ так γ SAGA не более $\frac{1}{L}$ (также, так $\gamma \in \mathbb{D}$)

SVRG

$$x^{k+1} = x^k - \gamma g^k$$

$$g^k = \nabla f_{i_k}(x^k) - \nabla f_{i_k}(\omega^k) + \nabla f(\omega^k)$$

ω^k - выборка погр

$$\omega^k = \begin{cases} x^k \\ \omega^{k-1} \end{cases} \text{ / раз } b \text{ и неперезагружен}$$

Теорема: $x^k \rightarrow x^*$, но $\omega^k \rightarrow x^*$

$$g^k = \nabla f_{i_k}(x^k) - \nabla f_{i_k}(\omega^k) + \nabla f(\omega^k)$$

$$\underbrace{\nabla f_{i_k}(x^*) - \nabla f_{i_k}(x^*)}_{\rightarrow 0} + \underbrace{\nabla f(x^*)}_{\rightarrow 0}$$

$$\underbrace{\phantom{\nabla f_{i_k}(x^*) - \nabla f_{i_k}(x^*) + \nabla f(x^*)}}_{g^k \rightarrow 0}$$

Сложность:

$$k = O\left(\left[\frac{L}{\mu} + n\right] \log \frac{1}{\varepsilon}\right) \text{ итераций}$$

⊕ много SAGA

⊕ шаг $O(d)$

⊖ нужно считать все новые градиенты