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

$$f(x) = \frac{1}{2} x^T A x - b x \qquad \nabla f(x) = \underbrace{A}_{d \times d} \underbrace{x}_{d} - b$$

операций $d \times d + d = O(d^2)$

$$\nabla f_{(i)}(x) = (Ax - b)_{(i)} =$$

компонента

$a_i x - b_i$ — i строка

$O(d)$ операций

$$f(x) = \frac{1}{2n} \sum_{i=1}^{n} (a_i x - b_i)^2$$

$$\nabla f_i(x) = \nabla \left[\frac{1}{2}(a_i x - b_i)^2\right] = (a_i x - b_i)\, a_i^T$$

Было слагаемых по i данным

$$\nabla f_i(x) = \begin{pmatrix} \text{весь} \\ \text{ненулевой} \end{pmatrix}$$

$$f(x) = \frac{1}{2n} \sum_{i=1}^{n} (a_i x - b_i)^2$$

$$\nabla f_{(j)}(x) = \frac{1}{2n} \sum_{i=1}^{n} (a_i x - b_i)\, a_{ij}$$

$$\nabla f_{(i)}(x) = \begin{pmatrix} 0 \\ \text{один ненулевой} \\ 0 \end{pmatrix}$$

SGD

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

i_k — случ. из 1 до n

Coord.-SGD

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

i_k — ~~сл~~ случ. от 1 до d

Док-во: μ-сильно вып. f, L-гладкая f

$$\mathbb{E}\|x^{k+1} - x^*\|^2 = (1-\mu\gamma)\,\mathbb{E}\|x^k - x^*\|^2 - \gamma\mathbb{E}\left(f(x^k) - f(x^*)\right) + \gamma^2 \|\nabla f_{(i_k)}(x^k)\|^2$$

$$\mathbb{E}\left[\nabla f_{(i_k)}(x^k) \,\middle|\, x^k\right] = \frac{1}{d}\begin{pmatrix} \nabla f_{(1)}(x^k) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \frac{1}{d}\begin{pmatrix} 0 \\ \nabla f_{(2)}(x^k) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \ldots$$

$$\ldots = \frac{1}{d}\sum_{i=1}^{d} \langle \nabla f(x^k); e_i \rangle e_i$$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{pmatrix}$$ i-место

$$= \frac{1}{d}\cdot \nabla f(x^k) \neq \nabla f(x^k)$$

Модификация метода:

$$\nabla f_{(i_k)}(x^k) \to d \cdot \nabla f_{(i_k)}(x^k)$$

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

Несмещ. → в худшем случае влечет, что реальный шаг $\times\, d$

$$\mathbb{E}\left[d \cdot \nabla f_{(i_k)}(x^k)\right] = \nabla f(x^k)$$ → анализ работает

$$\mathbb{E}\|x^{k+1}-x^*\|^2 = (1-\mu\gamma)\,\mathbb{E}\|x^k-x^*\|^2 - \gamma\mathbb{E}(f(x^k)-f(x^*)) + \gamma^2 d^2\,\mathbb{E}\|\nabla f_{(i_k)}(x^k)\|^2$$

$$\mathbb{E}\left[\|\nabla f_{(i_k)}(x^k)\|^2 \mid x^k\right] =$$

$$= \frac{1}{d}\|\nabla f_{(1)}(x_k)\|^2 + \dots = \frac{1}{d}\sum_{i=1}^{d}\|\nabla f_{(i)}(x_k)\|^2 = \|\nabla f(x^k)\|^2$$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 = (1-\mu\gamma)\,\mathbb{E}\|x^k-x^*\|^2 - \gamma\mathbb{E}(f(x^k)-f(x^*)) + \frac{\gamma^2 d^{\cancel{2}}}{\cancel{d}}\cdot\mathbb{E}\|\nabla f(x^k)\|^2$$

гладкость

$$\mathbb{E}\|x^{k+1}-x^*\|^2 = (1-\mu\gamma)\,\mathbb{E}\|x^k-x^*\|^2 - \gamma\mathbb{E}(f(x^k)-f(x^*)) + \gamma^2 d\,\mathbb{E}(2L(f(x^k)-f(x^*)))$$

$$-\gamma + 2Ld\gamma^2 \le 0 \qquad -1+\gamma 4Ld \le 0$$

$$\gamma \le \frac{1}{4Ld}$$

$$\mathbb{E}\left[\|x^{k+1}-x^*\|^2\right] \le (1-\mu\gamma)\,\mathbb{E}\left[\|x^k-x^*\|^2\right]$$

$\frac{1}{\gamma\mu}\log\frac{1}{\varepsilon}$ итераций

$\gamma = \frac{1}{4Ld}$ $\qquad O\left(\frac{Ld}{\mu}\log\frac{1}{\varepsilon}\right)$ итераций до точн. ε

SAGA

$$y^k_{i_k} = \nabla f_{i_k}(x^k) \qquad y^k_{j\ne i_k} = y^{k-1}_j$$

$$x^{k+1} = x^k - \gamma\cdot\frac{1}{n}\sum_{i=1}^n y^k_i$$

SEGA

$$h^k_i \in \mathbb{R} \qquad h^k_{(i_k)} = \langle \nabla f(x^k); e_{i_k}\rangle$$

$$h^k_{j\ne i_k} = h^{k-1}_j$$

$$x^{k+1} = x^k - \gamma\cdot d\cdot\begin{pmatrix} h^k_1 \\ \vdots \\ h^k_d \end{pmatrix}$$

Док-во SEGA

μ-сильно вып f, L-глад. f

$$h^k = \begin{pmatrix} h^k_1 \\ \vdots \\ h^k_d \end{pmatrix} \qquad h^{k+1} = h^k - \begin{pmatrix} 0 \\ \vdots \\ 0 \\ h^k_{(i_k)} - \langle \nabla f(x^k); e_{i_k}\rangle \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= h^k - \langle h^k; e_{i_k}\rangle e_{i_k} + \langle \nabla f(x^k); e_{i_k}\rangle e_{i_k}$$

$$= h^k + \langle \nabla f(x^k) - h^k; e_{i_k}\rangle e_{i_k}$$

$$x^{k+1} = x^k - \gamma g^k = x^k - \gamma\left(h^k + d\langle \nabla f(x^k) - h^k; e_{i_k}\rangle e_{i_k}\right)$$

$$\mathbb{E}\left[g^k \mid x^k \right] =$$

$$= \mathbb{E}\left[g^k \mid x^k \right] =$$

$$= \mathbb{E}\left[h^k + d \langle \nabla f(x^k) - h^k ; e_{i_k} \rangle e_{i_k} \mid x^k \right] =$$

$$= \left(h^k + d \frac{1}{d} \underbrace{\sum_{i=1}^{d} \langle \nabla f(x^k) - h^k ; e_i \rangle e_i}_{= \nabla f(x^k) - h^k} \right)$$

$$= h^k + \nabla f(x^k) - h^k$$

$$= \nabla f(x^k)$$

$$\mathbb{E} \| x^{k+1} - x^* \|^2 \leq (1 - \mu\gamma) \mathbb{E}\left[\| x^k - x^* \|^2 \right] - \gamma \left(f(x^k) - f(x^*) \right) + \gamma^2 \mathbb{E}\left[\| g^{k+1} - \nabla f(x^*) \|^2 \right]$$

$$\mathbb{E}\left[\| h^{k+1} - \nabla f(x^*) \|^2 \mid x^k \right] =$$

$$= \mathbb{E}\left[\| h^k - \nabla f(x^*) + \langle \nabla f(x^k) - h^k ; e_{i_k} \rangle e_{i_k} \|^2 \mid x^k \right]$$

$$= \mathbb{E}\left[\| h^k - \nabla f(x^*) + \langle \nabla f(x^k) - \nabla f(x^*) + \nabla f(x^*) - h^k ; e_{i_k} \rangle e_{i_k} \|^2 \right]$$

$$= \mathbb{E}\Big[\| h^k - \nabla f(x^*) - \langle h^k - \nabla f(x^*) ; e_{i_k} \rangle e_{i_k}$$
$$+ \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k} \rangle e_{i_k} \|^2 \,\Big|\, x^k \Big]$$

$$= \mathbb{E}\left[\| h^k - \nabla f(x^*) - \langle h^k - \nabla f(x^*) ; e_{i_k} \rangle e_{i_k} \|^2 \right]$$
$$+ \mathbb{E}\left[\| \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k} \rangle e_{i_k} \|^2 \right]$$
$$+ 2\mathbb{E}\Big[\langle h^k - \nabla f(x^*) - \langle h^k - \nabla f(x^*) ; e_{i_k} \rangle \underline{e_{i_k}} ;$$
$$\underline{e_{i_k}} \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k} \rangle \rangle \Big]$$

$$\|$$

$$2\mathbb{E}\left[\langle h^k - \nabla f(x^*) ; e_{i_k} \rangle \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k} \rangle \right]$$
$$- 2\mathbb{E}\left[\langle h^k - \nabla f(x^*) e_{i_k} \rangle \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k} \rangle \right]$$

$$\| \; 0$$

$$\mathbb{E}\left[\| h^{k+1} - \nabla f(x^*) \|^2 \,\middle|\, x^k \right]$$

$$= \mathbb{E}\left[\|h^k - \nabla f(x^*) - \langle h^k - \nabla f(x^*); e_{i_k}\rangle e_{i_k}\|^2 \,|\, x^k\right]$$

$$+ \mathbb{E}\left[\| \langle \nabla f(x^k) - \nabla f(x^*); e_{i_k}\rangle e_{i_k}\|^2 \,|\, x^k\right]$$

$$\| \quad \frac{1}{d}\|\nabla f(x^k) - \nabla f(x^*)\|^2$$

$$\| \quad \frac{1}{d}\|\nabla f(x^k)\|^2$$

$$= \frac{1}{d}\|\nabla f(x^k)\|^2 + \|h^k - \nabla f(x^*)\|^2$$

$$+ \mathbb{E}\left[\|\langle h^k - \nabla f(x^*); e_{i_k}\rangle e_{i_k}\|^2 \,|\, x^k\right]$$

$$- 2\mathbb{E}\left[\langle h^k - \nabla f(x^*); e_{i_k}\rangle \langle h^k - \nabla f(x^*); e_{i_k}\rangle\right]$$

$$= \frac{1}{d}\|\nabla f(x^k)\|^2 + \|h^k - \nabla f(x^*)\|^2$$

$$+ \frac{1}{d}\|\nabla f(x^*) - h^k\|^2$$

$$- 2\mathbb{E}\left[\|\langle h^k - \nabla f(x^*); e_{i_k}\rangle e_{i_k}\|^2 \,|\, x^k\right]$$

$$= \frac{1}{d}\|\nabla f(x^k)\|^2 + \left(1 - \frac{1}{d}\right)\|h^k - \nabla f(x^*)\|^2$$

$$\boxed{\mathbb{E}\left[\|h^{k+1} - \nabla f(x^*)\|^2\right] = \left(1 - \frac{1}{d}\right)\|h^k - \nabla f(x^*)\|^2 + \frac{1}{d}\|\nabla f(x^k)\|^2}$$

$$\mathbb{E}\left[\|g^k\|^2 \mid x^k\right] =$$

$$= \mathbb{E}\left[\left\| h^k + d\langle \nabla f(x^k) - h^k ; e_{i_k}\rangle e_{i_k} - \nabla f(x^*)\right\|^2 \mid x^k\right]$$

$$= \mathbb{E}\left[\left\| h^k - \nabla f(x^*) + d\langle \nabla f(x^k) - h^k ; e_{i_k}\rangle e_{i_k}\right\|^2\right]$$

$$= \mathbb{E}\Big[\Big\| h^k - \nabla f(x^*) - d\langle h^k - \nabla f(x^*) ; e_{i_k}\rangle e_{i_k}$$

$$+ d\langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k}\rangle e_{i_k}\Big\|^2\Big]$$

$$\le 2\,\mathbb{E}\left[\left\| \underbrace{h^k - \nabla f(x^*)}_{\mathbb{E}} - \underbrace{d\langle h^k - \nabla f(x^*) ; e_{i_k}\rangle e_{i_k}}\rangle\right\|^2\right]$$

$$+ 2d^2\,\mathbb{E}\left[\left\| \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k}\rangle e_{i_k}\right\|^2\right]$$

$$\le 2d^2\,\mathbb{E}\left[\left\| \langle h^k - \nabla f(x^*) ; e_{i_k}\rangle e_{i_k}\rangle\right\|^2\right]$$

$$+ 2d^2\,\mathbb{E}\left[\left\| \langle \nabla f(x^k) - \nabla f(x^*) ; e_{i_k}\rangle ; e_{i_k}\right\|^2\right]$$

$$= \frac{2d^2}{d}\left\| h^k - \nabla f(x^*)\right\|^2$$

$$+ \frac{2d^2}{d}\left\| \nabla f(x^k) - \nabla f(x^*)\right\|^2$$

$$\mathbb{E}\|g^k\|^2 \le 2d\cdot\mathbb{E}\|\nabla f(x^k)\|^2 + 2d\cdot\mathbb{E}\|h^k-\nabla f^*\|^2$$

$$\begin{aligned}\mathbb{E}\|x^{k+1}-x^*\|^2 \le{} & (1-\mu\gamma)\,\mathbb{E}\|x^k-x^*\|^2 \\ & -\gamma\,\mathbb{E}\big(f(x^k)-f(x^*)\big) \\ & +4d\gamma^2L\,\mathbb{E}\big(f(x^k)-f(x^*)\big) \\ & +2d\gamma^2\,\mathbb{E}\|h^k-\nabla f^*\|^2\end{aligned}$$

$$\begin{aligned}\mathbb{E}\|x^{k+1}-x^*\|^2 + \mathbb{E}\left[M\gamma^2\|h^{k+1}-\nabla f(x^*)\|^2\right] \le{} & (1-\mu\gamma)\,\mathbb{E}\|x^k-x^*\|^2 \\ & +\left(4d\gamma^2L+\frac{2LM\gamma^2}{d}-\gamma\right)\big(f(x^k)-f(x^*)\big) \\ & +\underbrace{\left(1-\frac{1}{d}+\frac{2d}{M}\right)}_{\left(1-\frac{1}{2d}\right)}\gamma^2\cdot\mathbb{E}\left[\|h^k-\nabla f(x^*)\|^2\right]\end{aligned}$$

$M = 4d^2$

$$4d\gamma^2L + \frac{2LM\gamma^2}{d} - \gamma = 4d\gamma^2L + 8dL\gamma - \gamma$$

$$= 12dL\gamma^2 - \gamma \le 0$$

$$\gamma = \frac{1}{24dL}$$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 + \mathbb{E}\left[M\gamma^2\|h^{k+1}-\nabla f(x^*)\|^2\right]$$

$$\leq (1-\mu\gamma)\mathbb{E}\|x^k-x^*\|^2$$

$$+ \left(1-\frac{1}{2d}\right)\mathbb{E}\left[M\gamma^2\|h^k-\nabla f(x^*)\|^2\right.$$

$$\leq \max\left(1-\mu\gamma;\ 1-\frac{1}{2d}\right)\left(\ .\ \ .\ \ .\ \right)$$

$$O\left(\frac{1}{\mu\gamma}+d\right) = O\left(\frac{dL}{\mu}\log\frac{1}{\varepsilon}\right)$$ итерации

нижняя $\Omega\left(d\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}\right)$