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

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

SAGA

$y_i = \nabla f_i(x^0) \qquad i = 1..n$

случайно выбираем $i_k = 1..n$

$y^k_{i_k} = \nabla f_{i_k}(x^k) \qquad y^k_{i \neq i_k} = y^{k-1}_i$

$$g^k = \frac{1}{n} \sum_{i=1}^{n} y^k_i = \frac{1}{n} \sum y^{k-1}_i + \nabla f_{i_k}(x^k) - y^{k-1}_{i_k}$$

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

L - гладких (всех f_i) $\qquad \mu$ - сильно вып (f)

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

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

? g_k

$$\mathbb{E}\left[g^k \mid x^k\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} y^{k-1}_i + \nabla f_{i_k}(x^k) - y^{k-1}_{i_k} \,\Big|\, x^k\right]$$

$$= \frac{1}{n} \sum y^{k-1}_i + \mathbb{E}\left[\nabla f_{i_k}(x^k) - y^{k-1}_{i_k} \mid x^k\right]$$

$$= \cancel{\frac{1}{n} \sum y^{k-1}_i} + \frac{1}{n} \sum_{i=1}^{n} \nabla f_i(x^k) - \cancel{\frac{1}{n} \sum y^{k-1}_i}$$

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

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

$$= \mathbb{E}\left[\left\| \frac{1}{n}\sum_{i=1}^{n} y_i^{k-1} + \nabla f_{i_k}(x^k) - y_{i_k}^{k-1} - \nabla f(x^*) \right\|^2\right]$$

$$= \mathbb{E}\Big[\Big\| \nabla f_{i_k}(x^k) - \nabla f_{i_k}(x^*) + \nabla f_{i_k}(x^*) + \frac{1}{n}\sum y_i^{k-1} - y_{i_k}^{k-1} - \nabla f(x^*) \Big\|^2$$

$$\|a+b\|^2 = \|a\|^2 + 2\langle a; b\rangle + \|b\|^2 \le 2\|a\|^2 + 2\|b\|^2$$

$$\le \|a\|^2 + \|b\|^2$$

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

$$+ 2\mathbb{E}\left[\left\| y_{i_k}^{k-1} - \nabla f_{i_k}(x^*) - \left[\frac{1}{n}\sum y_i^{k-1} - \nabla f(x^*)\right] \right\|^2\right]$$

$$\mathbb{E}[\quad] = \mathbb{E}\left[\mathbb{E}\left[\quad | x^k\right]\right]$$

$$\le 2\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\nabla f_i(x^k) - \nabla f_i(x^*)\|^2\right]$$

$$+ 2\mathbb{E}\left[\mathbb{E}\left[\left\| y_{i_k}^{k-1} - \nabla f_{i_k}(x^*) - \mathbb{E}\left[\| y_{i_k}^{k-1} - \nabla f_{i_k}(x^*)\|^2 \,|\, x^k\right]\right\|^2 \,\Big|\, x^k\right]\right]$$

для 1-го исп. гладкость | для 2-го $\mathbb{D}x \le \mathbb{E}x^2$

$$\le 4L\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(f_i(x^k) - f_i(x^*) - \langle \nabla f_i(x^*); x^k - x^* \rangle\right)\right]$$

$$+ 2\mathbb{E}\left[\mathbb{E}\left[\| y_{i_k}^{k-1} - \nabla f_{i_k}(x^*)\|^2 \,|\, x^k\right]\right]$$

$$= 4L\,\mathbb{E}\left[f(x^k) - f(x^*) \right]$$

$$+ 2\,\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \| y_i^{k-1} - \nabla f_i(x^*) \|^2 \right] \quad (+)$$

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \| y_i^k - \nabla f_i(x^*) \|^2 \right] =$$

$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\| y_i^k - \nabla f_i(x^*) \| \;\middle|\; x^k \right] \right]$$

$y_i^k = y_i^{k-1}$ с вер. $\frac{n-1}{n}$

$y_i^k = \nabla f_i(x^k)$ с вер. $\frac{1}{n}$

$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \frac{n-1}{n} \| y_i^{k-1} - \nabla f_i(x^*) \|^2 + \frac{1}{n} \| \nabla f_i(x^k) - \nabla f_i(x^*) \|^2 \right]$$

$$= \left(1 - \frac{1}{n}\right) \mathbb{E}\left[\frac{1}{n} \sum \| y_i^{k-1} - \nabla f_i(x^*) \|^2 \right]$$

$$\underbrace{+ \mathbb{E}\left[\frac{1}{n^2} \sum \| \nabla f_i(x^k) - \nabla f_i(x^*) \|^2 \right]}_{\text{из гладкости}}$$

$$\leq \left(1 - \frac{1}{n}\right) \mathbb{E}\left[\frac{1}{n} \sum \| y_i^{k-1} - \nabla f_i(x^*) \|^2 \right]$$

$$\frac{2L}{n} \mathbb{E}\left[f(x^k) - f(x^*) \right] \quad (+)$$

Итог:

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

$$\left.\begin{aligned}&-2\gamma\,\mathbb{E}\left[f(x^k)-f(x^*)\right]\\ &+\gamma^2\cdot 4L\,\mathbb{E}\left[f(x^k)-f(x^*)\right]\end{aligned}\right] \quad \text{зачем } \gamma \le 0 \text{ (удоб.)}$$

$$+\gamma^2\cdot 2\,\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|y_i^{k-1}-\nabla f_i(x^*)\|^2\right]$$

$$\underline{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|y_i^{k}-\nabla f_i(x^*)\|^2\right]} \le$$

$$\left.\left(1-\frac{1}{n}\right)\mathbb{E}\left[\frac{1}{n}\sum\|y_i^{k-1}-\nabla f_i(x^*)\|^2\right]\right] \quad \text{лин. сход}$$

$$\left.+\frac{2L}{n}\,\mathbb{E}\left[f(x^k)-f(x^*)\right]\right] \quad \text{упрощ.}$$

$$\mathbb{E}\left[\|x^{k+1}-x^*\|^2 + M\cdot\frac{1}{n}\sum_{i=1}^{n}\|y_i^{k}-\nabla f_i(x^*)\|^2\right]$$

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

$$-2\gamma\,\mathbb{E}\left[f(x^k)-f(x^*)\right]$$

$$+\gamma^2\cdot 4L\,\mathbb{E}\left[f(x^k)-f(x^*)\right]$$

$$+\gamma^2\cdot 2\,\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|y_i^{k-1}-\nabla f_i(x^*)\|^2\right]$$

$$+ M \cdot \left(1-\frac{1}{n}\right) \mathbb{E}\left[\frac{1}{n}\sum \|y_i^{k-1} - \nabla f_i(x^*)\|^2\right]$$

$$+ M \cdot \frac{2L}{n} \mathbb{E}\left[f(x^k) - f(x^*)\right]$$

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

$$\left(1-\frac{1}{n}+\frac{\gamma^2}{M}\right) M \cdot \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|y_i^{k-1} - \nabla f_i(x^*)\|^2\right]$$

$$-\left(2\gamma + 4L\gamma^2 + M \cdot \frac{2L}{n}\right) \mathbb{E}\left[f(x^k) - f(x^*)\right]$$

1) M $\quad 1-\gamma\mu = 1-\frac{1}{n}+\frac{\gamma^2}{M}$

2) M $\quad 1-\frac{1}{n}+\frac{\gamma^2}{M} = 1-\frac{1}{2n}$ $\quad +$

$$M = 2n\gamma^2$$

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

$$\left(1-\frac{1}{2n}\right) M \cdot \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|y_i^{k-1} - \nabla f_i(x^*)\|^2\right]$$

$$-\left(2\gamma + 4L\gamma^2 + 4L\gamma^2\right) \mathbb{E}\left[f(x^k) - f(x^*)\right]$$

$$-2\gamma + 8L\gamma^2 \leq 0$$

$$\gamma \leq \frac{1}{2L}$$

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

$$\left(1 - \frac{1}{2n}\right) M \cdot \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|y_i^{k-1} - \nabla f_i(x^*)\|^2\right]$$

$$\leq \max\left(1-\mu\gamma;\ 1-\frac{1}{2n}\right)\mathbb{E}\left[\|x^k - x^*\|^2 + \frac{M}{n}\sum_{i=1}^{n}\|y_i^{k-1} - \nabla f_i(x^*)\|^2\right]$$

$$O\left(\frac{1}{\min\left(\mu\gamma;\ \frac{1}{2n}\right)}\right) =$$

$$= O\left(\max\left(\frac{1}{\mu\gamma};\ n\right)\right) \overset{\gamma}{=}$$

$$= O\left(\max\left(\frac{L}{\mu};\ n\right)\right) = \begin{cases} O\left(\frac{L}{\mu}\right) & \frac{L}{\mu} > n \\ O(n) & n \geq \frac{L}{\mu} \end{cases}$$

$$O\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right) \text{ — GD}$$

$$O\left(\left(\frac{L}{\mu}+n\right)\log\frac{1}{\varepsilon}\right) \text{ — SAGA}$$

$$R_k^2+\Psi_k^2 \le (1-\rho)^k\left(R_0^2+\Psi_0^2\right)$$

$$R_k^2 \le R_k^2+\Psi_k^2$$

$$R_k^2 \le (1-\rho)^k\left(R_0^2+\Psi_0^2\right)$$

Нижние оценки:

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

$$\boxed{2\sqrt{n}\sqrt{\frac{L}{\mu}} \le n+\frac{L}{\mu}}$$

Оптимальный метод: Katyusha

SVRG

(проблема SAGA, использ. y_i)

SVRG решает пробл. SAGA

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

ω - обновляется редко

SVRG: $O\left(\left(n + \frac{L}{\mu}\right) \log \frac{1}{\varepsilon}\right)$

храним 1 вектор доп. $\nabla f(\omega)$, но считаем 2 град. за итерацию

SARAH

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

SARAH часто на практике лучше - устойчивее раз в большое число итераций

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

если для SAGA и SVRG

$\mathbb{E}[g^k | x^k] = \nabla f(x^k)$

не для SARAH

$$\mathbb{E}[g^k | x^k] = \nabla f(x^k) - \nabla f(x^{k-1}) + g^{k-1} \neq \nabla f(x^k)$$

$\neq 0$

Loopless SVRG SARAH/PAGE

SVRG раз в T итераций

ω обновление на x^k

— нужно заново $\nabla f(w)$

1) τ можно выбирать детерм.

2) τ можно брать стох.

Loopless $\tau \sim Geom(p)$

$$w^{k+1} = \begin{cases} w^k & 1-p \\ x^k & p \end{cases}$$

---

1) $\max(1-\mu\gamma;\ 1-p) \leftarrow$ SVRG

$$p = \mu\gamma \approx \mu/L$$

2) $p = \frac{1}{n}$

раз (в среднем) в n итераций считаем полн. град

$\Theta(1)$ — стох град

+ 1 полный град $(\Theta(n))$

$= \Theta(n)$

$\Theta(\tau)$ смен узла $+$ $\underbrace{\Theta(n)}$

назад

$=$

n смен узла.

$= \Theta(n+\tau) \quad \tau \sim n$