$$\min_{x \in \mathbb{R}^d} f(x) = \mathbb{E}_D\left[f(x, D) \right]$$

$\xi_i \in D$

1) $\approx \frac{1}{n} \sum_{i=1}^{n} f(x, \xi_i)$

2) $f(x, \xi) \qquad \nabla f(x, \xi)$

$$\mathbb{E}\left[\|\nabla f(x,\xi) - \nabla f(x)\|^2 \right] < \sigma^2$$

$$\mathbb{E}\, \nabla f(x, \xi) = \nabla f(x)$$

Предположение: $\underline{\mathbb{E}\left[\|\overset{=0}{\nabla f(x^*)} - \nabla f(x^*, \xi)\|^2 \right]} \leq \sigma_*^2$

$+$ L-гладкость $\forall \xi$, $f(x, \xi)$

$+$ μ-сильная вып f

SGD: $x^{k+1} = x^k - \gamma \nabla f(x^k, \xi^k)$

$$\mathbb{E}\left[\|x^{k+1} - x^*\|^2 \right] \leq \mathbb{E}\left[\|x^k - x^*\|^2 \right] - 2\gamma\, \mathbb{E}\langle \underbrace{\mathbb{E}[\nabla f(x^k, \xi^k) | x^k]}_{\nabla f(x^k)}, x^k - x^* \rangle + \gamma^2 \mathbb{E}\left[\|\nabla f(x^k, \xi^k)\|^2 \right]$$

сильная вып:

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

КБШ

$(a+b)^2 \leq 2a^2 + 2b^2$

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

$$+2\gamma^2 \underbrace{\mathbb{E}\left[\|\nabla f(x^*,\xi^k)\|^2\right]}_{\le \sigma_*^2}$$

L - гладкость f_ξ

$$\le \mathbb{E}\left[(1-\mu\gamma)\|x^k-x^*\|^2\right]$$
$$-2\gamma\mathbb{E}\left(f(x^k)-f(x^*)\right)$$
$$+2\gamma^2\mathbb{E}\left[2L\left(f(x^k,\xi^k)-f(x^*,\xi^k) - \langle \nabla f(x^*,\xi^k); x^k-x^*\rangle\right)\right]$$
$$+2\gamma^2\sigma_*^2$$

$$\mathbb{E}\left[f(x^k,\xi^k)\,|\,x^k\right] = f(x^k)$$
$$\mathbb{E}\left[f(x^*,\xi^k)\right] = f(x^*)$$
$$\mathbb{E}\left[\nabla f(x^*,\xi^k)\,|\,x^k\right] = \nabla f(x^*) = 0$$

$$\le \mathbb{E}\left[(1-\mu\gamma)\|x^k-x^*\|^2\right]$$
$$-2\gamma\mathbb{E}\left(f(x^k)-f(x^*)\right) + 4L\gamma^2\left(f(x^k)-f(x^*)\right)$$
$$+2\gamma^2\sigma_*^2$$

$$-2+8L\gamma=0 \qquad \gamma \le \frac{1}{4L}$$

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

Предположение: $\|\nabla f(x^*) - \nabla f(x^*,\xi)\|^2 \le \sigma_*^2$

$+$ L-гладкость f, $f(x,\xi)$

$+$ μ-сильная вып. f, $f(x,\xi)$

$$
\begin{aligned}
\mathbb{E}\left[\|x^{k+1}-x^*\|^2\right] \le\ & \mathbb{E}\left[\|x^k - x^*\|^2\right] \\
& -2\gamma\, \mathbb{E}\langle \nabla f(x^k,\xi^k) ; x^k - x^* \rangle \\
& +2\gamma^2 \mathbb{E}\left[\|\nabla f(x^k,\xi^k) - \nabla f(x^*,\xi)\|^2\right] \\
& +2\gamma^2 \sigma_*^2 \\
\le\ & \mathbb{E}\left[\|x^k - x^*\|^2\right] \\
& -2\gamma\left(\frac{\mu}{2}\|x^k - x^*\|^2 + f(x^k,\xi^k) - f(x^*,\xi^k)\right) \\
& +4L\gamma^2 \mathbb{E}\Big[\big(f(x^k,\xi^k) - f(x^*,\xi^k) \\
& \qquad - \langle \nabla f(x^*,\xi^k) ; x^k - x^* \rangle\big)\Big] \\
& +2\gamma^2 \sigma_*^2 \\
=\ & (1-\mu\gamma)\,\mathbb{E}\left[\|x^k - x^*\|^2\right] \\
& -(2\gamma - 4L\gamma^2)\,\mathbb{E}\left[f(x^k,\xi^k) - f(x^*,\xi^k)\right] \\
& +4L\gamma^2 \mathbb{E}\Big[\underbrace{-\langle \nabla f(x^*,\xi^k), x^k - x^* \rangle}_{\text{КБШ}}\Big] \\
& +2\gamma^2\sigma^2
\end{aligned}
$$

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

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

$$+4\gamma^2 L\,\mathbb{E}\Big[\underbrace{\|\nabla f(x^*,\xi^k)\|\,\|x^k - x^*\|}_{ab \leq c^2 a^2 + \frac{1}{c^2} b^2}\Big]$$

$$+2\gamma^2 \sigma_*^2$$

$$\frac{4\gamma^2 L}{c^2} = \frac{\mu\gamma}{2} \qquad c^2 = \frac{8\gamma^2 L}{\mu\gamma} = \frac{8\gamma L}{\mu}$$

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

$$+ \frac{\mu\gamma}{2}\mathbb{E}\|x^k - x^*\|^2$$

$$+ 4\gamma^2 L \cdot \frac{8\gamma L}{\mu} \underbrace{\mathbb{E}\|\nabla f(x^*,\xi^k)\|^2}_{\leq \sigma_*^2}$$

$$+ 2\gamma^2\sigma_*^2$$

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

$$+ 2\gamma^2\sigma_*^2\Big(1 + \overbrace{\frac{16L^2\gamma}{\mu}}^{\gamma \leq \frac{1}{4L}}\Big)$$

$$1 + \frac{4L}{\mu} \leq \frac{5L}{\mu}$$

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

$$+ 2\gamma^2 \cdot \frac{5L}{\mu}\sigma_*^2$$

лин. $\left(\frac{L}{\mu}\right)$ + суб. $\frac{\sigma_*^2}{\varepsilon\mu^2} \cdot \frac{L}{\mu}$

$$+ \text{сув} \quad \frac{\sigma_*^2}{\varepsilon\mu^2} + \frac{\sigma_*^2 \frac{L}{\mu}}{\sqrt{\varepsilon}\,\mu^2}$$

$$\gamma = \min\left(\frac{1}{4L}; \frac{1}{\mu T}\right)$$

EG

$$x^{k+1/2} = x^k - \gamma F(x^k)$$
$$x^{k+1} = x^k - \gamma F(x^{k+1/2})$$

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - 2\gamma \langle F(x^{k+1/2}); x^{k+1/2} - x^* \rangle - \|x^k - x^{k+1/2}\|^2 + \gamma^2 \|F(x^{k+1/2}) - F(x^k)\|^2$$

1) незав. случ $\xi^k, \xi^{k+1/2}$

$$x^{k+1/2} = x^k - \gamma F(x^k, \xi^k) \qquad \mathbb{E}[\|F(x,\xi) - F(x)\|] \le \sigma^2$$
$$x^{k+1} = x^k - \gamma F(x^{k+1/2}, \xi^{k+1/2})$$

$$\mathbb{E}\|x^{k+1} - x^*\|^2 \le \mathbb{E}\|x^k - x^*\|^2 - 2\gamma \mathbb{E} \langle F(x^{k+1/2}, \xi^{k+1/2}), x^{k+1/2} - x^* \rangle - \mathbb{E}\|x^k - x^{k+1/2}\|^2 + \gamma^2 \mathbb{E}\|F(x^{k+1/2}, \xi^{k+1/2}) - F(x^k, \xi^k)\|^2$$

$$= \mathbb{E}\|x^k - x^*\|^2 - 2\gamma \mathbb{E} \langle F(x^{k+1/2}), x^{k+1/2} - x^* \rangle - \mathbb{E}\|x^k - x^{k+1/2}\|^2 + \gamma^2 \mathbb{E}\|F(x^{k+1/2}, \xi^{k+1/2}) - F(x^k, \xi^k)\|^2$$

μ - сильная мон $\quad F(x^*) = 0$

$$\le (1 - 2\mu\gamma)\, \mathbb{E}\|x^k - x^*\|^2 - \mathbb{E}\|x^k - x^{k+1/2}\|^2 + \gamma^2 \mathbb{E}\|F(x^{k+1/2}, \xi^{k+1/2}) - F(x^k, \xi^k)\|^2$$

$$\|F(x^{k+1/2}) - F(x^k) + F(x^{k+1/2},\xi^{k+1/2}) - F(x^{k+1/2}) + F(x^k) - F(x^k,\xi^k)\|^2$$

$$\le 2\|F(x^{k+1/2}) - F(x^k)\|^2$$

$$+ 2\|F(x^{k+1/2},\xi^{k+1/2}) - F(x^{k+1/2}) + F(x^k) - F(x^k,\xi^k)\|^2$$

$$\mathbb{E}\ \underbrace{\|F(x^{k+1/2},\xi^{k+1/2}) - F(x^{k+1/2})\|^2}_{\le \sigma^2} + \underbrace{\|F(x^k) - F(x^k,\xi^k)\|^2}_{\le \sigma^2}$$

$$+ 2\langle F(x^{k+1/2},\xi^{k+1/2}) - F(x^{k+1/2}),\ F(x^k) - F(x^k,\xi^k)\rangle$$

$$\mathbb{E}[\ \mid x^{k+1/2}] \quad 0$$

$$\mathbb{E}\|x^{k+1} - x^*\|^2 \le (1 - 2\mu\gamma)\,\mathbb{E}\|x^k - x^*\|^2 - \mathbb{E}\|x^k - x^{k+1/2}\|^2 + 2\gamma^2\mathbb{E}\|F(x^{k+1/2}) - F(x^k)\|^2 + 4\gamma^2\sigma^2$$

$$\gamma \le \frac{1}{\sqrt{2}L}$$

$$\le (1 - 2\mu\gamma)\,\mathbb{E}\|x^k - x^*\|^2 + 4\gamma^2\sigma^2$$

2) зав. стох.

$$x^{k+1/2} = x^k - \gamma F(x^k, \xi^k)$$
$$x^{k+1} = x^k - \gamma F(x^{k+1/2}, \xi^k)$$

$$\|F(x,\xi) - F(y,\xi)\|^2 \le L^2\|x-y\|^2$$

μ - сильно монот. $F(\cdot,\xi)$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 \le \mathbb{E}\|x^k - x^*\|^2 - 2\gamma\mathbb{E}\langle F(x^{k+1/2},\xi^k), x^{k+1/2}-x^*\rangle$$
$$-\mathbb{E}\|x^k - x^{k+1/2}\|^2 + \gamma^2\underbrace{\mathbb{E}\|F(x^{k+1/2},\xi^k) - F(x^k,\xi^k)\|^2}_{\le L^2\|x^{k+1/2}-x^k\|^2}$$

$$-\mathbb{E}\langle F(x^{k+1/2},\xi^k); x^{k+1/2}-x^*\rangle$$

$F(x^*)=0$

$F(x^*,\xi^k)\ne 0$

$$= -\mathbb{E}\langle F(x^{k+1/2},\xi^k)-F(x^*,\xi^k); x^{k+1/2}-x^*\rangle$$
$$-\mathbb{E}\langle F(x^*,\xi^k); x^{k+1/2}-x^*\rangle$$

сильн $F(\cdot,\xi)$

$$\le -\mu\|x^{k+1/2}-x^*\|^2$$
$$-\mathbb{E}\langle F(x^*,\xi^k); x^{k+1/2}-x^*\rangle$$

$$= -\mu\|x^{k+1/2}-x^*\|^2 +$$
$$-\underbrace{\mathbb{E}\langle F(x^*,\xi^k); x^k-x^*\rangle}_{\to 0}$$
$$-\underbrace{\mathbb{E}\langle F(x^*,\xi^k); x^{k+1/2}-x^k\rangle}_{КБШ}$$

? $\mathbb{E}[\,\cdot\,|x^k]$

$\mathbb{E}[F(x^*,\xi^k)|x^k] = F(x^*)=0$

$$\le -\mu\mathbb{E}\|x^{k+1/2}-x^*\|^2$$
$$+\frac{1}{4\gamma}\|x^{k+1/2}-x^k\|^2 + \gamma\|F(x^*,\xi^k)\|^2$$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 \le \mathbb{E}\|x^k - x^*\|^2 - 2\mu\gamma\, \mathbb{E}\left[\|x^{k+1/2}-x^*\|^2\right]$$
$$-\mathbb{E}\|x^k - x^{k+1/2}\|^2 + \gamma^2 L^2 \mathbb{E}\|x^{k+1/2} - x^k\|^2$$
$$+ \frac{1}{2}\mathbb{E}\|x^{k+1/2}-x^k\|^2 + \underbrace{2\gamma^2 \|F(x^*,\xi^k)\|^2}_{\sigma_*^2}$$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 \le \mathbb{E}\|x^k-x^*\|^2 - 2\mu\gamma\mathbb{E}\|x^{k+1/2}-x^*\|^2$$
$$-\left(\frac{1}{2}-\gamma^2 L^2\right)\|x^{k+1/2}-x^k\|^2 + 2\gamma^2\sigma_*^2$$

$$-\|x^{k+1/2}-x^k+x^k-x^*\|^2 \le -\frac{1}{2}\|x^k-x^*\|^2 + \|x^{k+1/2}-x^k\|^2$$

$$\mathbb{E}\|x^{k+1}-x^*\|^2 \le (1-\mu\gamma)\mathbb{E}\|x^k-x^*\|^2 + 2\gamma^2\sigma_*^2$$
$$-\left(\frac{1}{2}-\gamma^2 L^2 + 2\gamma\mu\right)\|x^{k+1/2}-x^k\|^2$$

$$\gamma \le \frac{1}{3L}$$

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

$\|x^k - x^*\|^2$ $\quad g(x^*, y) \le g(x^*, y^*) \le g(x, y^*)$ $\forall x$

$f(x^k) - f(x^*)$

$\min_x \max_y g(x, y)$

$\min_{x \in [-1;1]} \max_{y \in [-1;1]} xy$ $\quad x^* = 0$

$x_0 = 1$ $\quad y^* = 0$

$y_0 = -1$

1) $g(x^k, y^k) - g(x^*, y^*)$

$g(x^0, y^0) - g(x^*, y^*) = -1$

2) $g(x^k, y^*) - g(x^*, y^k)$

$g(x^k, y^*) - g(x^*, y^*) + g(x^*, y^*) - g(x^*, y^k) \ge 0$

3) $\max_{y \in Y} g(\bar{x}^k, y) - \min_{x \in X} g(x, \bar{y}^k)$

$= \max_{(x,y) \in X \times Y} (g(x^k, y) - g(x, y^k))$

$= \max_{(x,y) \in X \times Y} (\underbrace{g(x^k, y) - g(x^k, y^k)}_{\text{вогнутая}} + \underbrace{g(x^k, y^k) - g(x, y^k)}_{\text{выпуклая}})$

$\le \max_{(x,y)} \left(\langle \nabla_y g(x^k, y^k); y - y^k \rangle + \langle \nabla_x g(x^k, y^k); x^k - x \rangle \right)$

$F = \begin{pmatrix} \nabla x \\ -\nabla y \end{pmatrix} \quad = \max_{\substack{(x,y) \\ \| \\ z}} \langle F(z^k); z^k - z \rangle$

$$\frac{1}{k}\sum_{k=0}^{k-1} \langle F(z^k) ; z^k - z \rangle$$

$$\langle F(z^k) - F(z) ; z^k - z \rangle \geq 0$$

$$\geq \frac{1}{k}\sum \langle F(z) ; z^k - z \rangle = \langle F(z) ; \bar{z}^k - z \rangle$$