1964 г:

тяжелого шарика (HB)

$$x^{k+1} = x^k - \eta \nabla f(x^k) + \underbrace{\gamma (x^k - x^{k-1})}_{\text{по инерции}}$$

1981 г Nesterov's acceleration

$$y^k = x^k + \gamma (x^k - x^{k-1})$$
$$x^{k+1} = y^k - \eta \nabla f(y^k)$$

$$x_g^k = \theta x_f^k + (1-\theta) x^k$$
$$x_f^{k+1} = x_g^k - \eta \nabla f(x_g^k)$$
$$x^{k+1} = \gamma (x_f^{k+1} - x_f^k) + x_f^k$$

$$x_g^k \to \nabla f(x_g^k)$$
$$x^k \to \|x^k - x^*\|^2$$
$$x_f^k \to f(x_f^k)$$

$$\gamma = \sqrt{\frac{L}{\mu}}$$
$$\eta = \frac{1}{L}$$
$$\theta = \frac{1}{1 + \frac{1}{\gamma}} = \frac{1}{1 + \sqrt{\frac{\mu}{L}}}$$

$$\|x^{k+1} - x^*\|^2 \overset{(3)}{=} \|\gamma x_f^{k+1} + (1-\gamma) x_f^k - x^*\|^2$$
$$\overset{(2)}{=} \|\gamma x_g^k - \eta\gamma \nabla f(x_g^k) + (1-\gamma) x_f^k - x^*\|^2$$
$$= \|\gamma x_g^k + (1-\gamma) x_f^k - x^*\|^2 + \eta^2 \gamma^2 \|\nabla f(x_g^k)\|^2$$
$$- 2\eta\gamma \langle \nabla f(x_g^k); \gamma x_g^k + (1-\gamma) x_f^k - x^* \rangle$$

$\ominus$

$$\gamma x_g^k + (1-\gamma) x_f^k = \gamma x_g^k + \frac{1-\gamma}{\theta} \theta x_f^k$$

$$\overset{(1)}{=} \gamma x_g^k + \frac{1-\gamma}{\theta}\left(x_g^k - (1-\theta)x^k\right)$$

$$= \left(\gamma + \frac{1-\gamma}{\theta}\right) x_g^k - \frac{(1-\theta)(1-\gamma)}{\theta} x^k$$

$$\frac{1-\theta-\gamma+\theta\gamma}{\theta} = \frac{1-\gamma+\theta\gamma}{\theta} - 1$$

$$= x^k - \left(\frac{1-\gamma+\theta\gamma}{\theta}\right) x^k + \left(\frac{1-\gamma+\gamma\theta}{\theta}\right) x_g^k$$

$$= x^k - \left(\frac{1-\gamma+\theta\gamma}{\theta}\right)(x^k - x_g^k)$$

$$\gamma + \frac{1-\gamma}{\theta} = \gamma + (1-\gamma)\left(1+\frac{1}{\gamma}\right) =$$

$$= \cancel{\gamma} + \cancel{1} - \cancel{\gamma} + \frac{1}{\gamma} - \cancel{1}$$

$$= \frac{1}{\gamma} \simeq \sqrt{\frac{\mu}{L}}$$

$$= x^k - \frac{1}{\gamma}(x^k - x_g^k)$$

$$\underbrace{\| \gamma x_g^k + (1-\gamma) x_f^k - x^* \|^2}_{} = \| x^k - x^* - \frac{1}{\gamma}(x^k - x_g^k) \|^2$$

$$= \| x^k - x^* \|^2 + \frac{1}{\gamma^2} \| x^k - x_g^k \|^2 + \frac{1}{\gamma} \cdot 2 \underbrace{\langle x^k - x^*; x_g^k - x^k \rangle}_{}$$

$$\underset{\| x_g^k - x^* \|^2}{\overset{\|}{\| x_g^k - x^k + x^k - x^* \|^2}} = \| x_g^k - x^k \|^2 + \| x^k - x^* \|^2 + \underbrace{2 \langle x_g^k - x^k; x^k - x^* \rangle}_{}$$

$$\| \gamma x_g^k + (1-\gamma) x_f^k - x^* \|^2 = \| x^k - x^* \|^2 + \frac{1}{\gamma^2} \| x^k - x_g^k \|^2 + \frac{1}{\gamma}\left(\| x_g^k - x^* \|^2 - \| x^k - x^* \|^2 - \| x_g^k - x^k \|^2 \right)$$

$$= \underset{1-\sqrt{\frac{\mu}{L}}}{\left(1 - \frac{1}{\gamma}\right)} \| x^k - x^* \|^2 + \left(\frac{1}{\gamma^2} - \frac{1}{\gamma}\right) \| x_g^k - x^k \|^2 + \frac{1}{\gamma} \| x_g^k - x^* \|^2$$

$=$

$$\|x^{k+1}-x^*\|^2 = \left(1-\frac{1}{\gamma}\right)\|x^k-x^*\|^2 + \left(\frac{1}{\gamma^2}-\frac{1}{\gamma}\right)\|x_g^k-x^k\|^2 + \frac{1}{\gamma}\|x_g^k-x^*\|^2$$

$$+\eta^2\gamma^2\|\nabla f(x_g^k)\|^2$$

$$-2\eta\gamma\langle \nabla f(x_g^k);\ \underbrace{\gamma x_g^k + (1-\gamma)x_f^k - x^*}_{x_g^k - x^* + (1-\gamma)(x_f^k - x_g^k)}\rangle$$

$$(\gamma-1)(x_g^k - x_f^k)$$

$+$

L-гладкость

$$f(x_f^{k+1}) \le f(x_g^k) + \langle \nabla f(x_g^k);\ x_f^{k+1}-x_g^k\rangle + \frac{L}{2}\|x_f^{k+1}-x_g^k\|^2$$

$$\overset{(2)}{=} f(x_g^k) - \eta\|\nabla f(x_g^k)\|^2 + \frac{L}{2}\eta^2\|\nabla f(x_g^k)\|^2$$

$$= f(x_g^k) - \eta\left(1-\frac{L\eta}{2}\right)\|\nabla f(x_g^k)\|^2$$

сильная выпуклость

$$f(x_g^k) \le f(u) - \langle \nabla f(x_g^k);\ u-x_g^k\rangle - \frac{\mu}{2}\|u-x_g^k\|^2$$

$$f(x_f^{k+1}) \le f(u) - \langle \nabla f(x_g^k);\ u-x_g^k\rangle - \frac{\mu}{2}\|u-x_g^k\|^2$$

$$-\eta\left(1-\frac{L\eta}{2}\right)\|\nabla f(x_g^k)\|^2$$

$u = x_f^k$ $\quad$ $u = x^*$

$$f(x_f^{k+1}) \le f(x^*) - \langle \nabla f(x_g^k);\ x^*-x_g^k\rangle - \frac{\mu}{2}\|x^*-x_g^k\|^2$$

$+\ 2\eta\gamma$

$$-\eta\left(1-\frac{L\eta}{2}\right)\|\nabla f(x_g^k)\|^2$$

$$f(x_f^{k+1}) \le f(x_f^k) - \langle \nabla f(x_g^k);\ x_f^k-x_g^k\rangle - \frac{\mu}{2}\|x_f^k-x_g^k\|^2$$

$+\ 2\eta\gamma(\gamma-1)$

$$-\eta\left(1-\frac{L\eta}{2}\right)\|\nabla f(x_g^k)\|^2$$

складываем $+$, убиваем скалярное произв.

$$\|x^{k+1}-x^*\|^2 + 2\eta\gamma^2 f(x_f^{k+1}) \le$$

$$\le \left(1-\frac{1}{\gamma}\right)\|x^k-x^*\|^2 + \frac{1}{\gamma}\|x_g^k - x^*\|^2$$

$$-\frac{1}{\gamma}\left(1-\frac{1}{\gamma}\right)\|x_g^k - x^k\|^2 + \eta^2\gamma^2\|\nabla f(x_g^k)\|^2$$

$$+2\eta\gamma f(x^*) - \eta\gamma\mu\|x_g^k - x^*\|^2$$

$$+2\eta\gamma(1-\gamma) f(x_f^k) + \times$$ с μ косяк

$$+2\eta\gamma^2\left(\frac{L\eta^2}{2} - \eta\right)\|\nabla f(x_g^k)\|^2$$

$\eta = \frac{1}{L}$

$\gamma = \sqrt{\frac{L}{\mu}} \ge 1$

$$\underbrace{2\eta\gamma^2\left(\frac{L\eta^2}{2} - \eta + \frac{\eta}{2}\right)}_{0}\|\nabla f(x_g^k)\|^2$$

$$\underbrace{-\frac{1}{\gamma}\underbrace{\left(1-\frac{1}{\gamma}\right)}_{\ge 0}}_{\le 0}$$

$$\underbrace{\left(-\eta\mu\gamma + \frac{1}{\gamma}\right)}_{0}$$

$$R_{k+1}^2 + 2\eta\gamma^2 f(x_f^{k+1}) \le \left(1-\frac{1}{\gamma}\right)R_k^2 + 2\eta\gamma^2\left(1-\frac{1}{\gamma}\right)f(x_f^k) + 2\eta\gamma f(x^*)$$

$$R_{k+1}^2 = \|x^{k+1}-x^*\|^2$$

$$\underbrace{R_{k+1}^2 + 2\eta\gamma^2\left(f(x_f^{k+1}) - f(x^*)\right)}_{\varphi_{k+1}} \le$$

$$\underbrace{\left(1-\frac{1}{\gamma}\right)}_{\left(1-\sqrt{\frac{\mu}{L}}\right)}\left[\underbrace{R_k^2 + 2\eta\gamma^2\left(f(x_f^k) - f(x^*)\right)}_{\varphi_k}\right]$$

$$\varphi_k \le \left(1-\sqrt{\frac{\mu}{L}}\right)^k \varphi_0$$

$$R_k^2 \le \varphi_k$$

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

Nesterov

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

выпуклый: $\theta \to \theta_k$

$\gamma \to \gamma_k$

Нижние оценки:

1) x_0 - начальная $\quad M_0 = \{x_0\}$

M_k - мн-во всех достиж. на данной итер. k

2) считать $\nabla f(x_k)$ $\quad x_k \in M_k$

3) $M_{k+1} = \text{span}\{x', \nabla f(x'')\}$

$x', x'' \in M_k$

$$f(x) = \frac{L-\mu}{8} x^T A x + \frac{\mu}{2} x^T x - \frac{L-\mu}{4} e_1^T x$$

$$A = \begin{pmatrix} 2 & -1 & & 0 \\ -1 & 2 & -1 & \\ 0 & -1 & 2 & -1 \\ 0 & & \ddots & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad 0 \preceq A \preceq 4I$$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

μ-сильно-выпуклая

$$\frac{1}{2} x^T \left(\frac{L-\mu}{4} A + \mu \right) x$$

$$\lambda_{max} \left(\underbrace{\frac{L-\mu}{4} A}_{L-\mu} + \mu \right) \leq L$$

K – итераций $\quad d = \infty$

$d = 2K$

Идея размерности задачи $d > K$

Идея: $x^0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$$\nabla f(x) = \frac{L-\mu}{4} A x + \mu x - \frac{L-\mu}{4} e_1$$

1) $\nabla f(x^0) \in \text{span}(e_1)$

за 1 итерацию только 1 коорд. может быть ненулевой

2) $\nabla f(x^1) \in \text{span}(e_1, e_2)$

$x^1 \in M_1$

за 2 итерации максим 2 первых коорд могут быть ненулевыми

k) по инд.

за k итерации максимум k первых

контрпример Нестерова

$$\nabla f(x^*) = 0 \qquad Ax^* + \frac{4\mu}{L-\mu} x^* - e_1 = 0$$

$$2x^{(1)} - x^{(2)} + \frac{4\mu}{L-\mu} x^{(1)} = 1$$

$$\frac{2(L+\mu)}{L-\mu} \cdot x^{(1)} - x^{(2)} = 1$$

$$-x^{(k-1)} + \frac{2(L+\mu)}{L-\mu} x^{(k)} - x^{(k+1)} = 0$$

$$x^{(k)} = q^k \qquad q = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$$

$$\|x^0 - x^*\|^2 = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}$$

$$\|x^k - x^*\|^2 \geq \sum_{i=k+1}^{\infty} q^{2i} = q^{2k} \sum_{i=1}^{\infty} q^{2i}$$

$$= q^{2k} \frac{q^2}{1-q^2} =$$

$$= q^{2k} \|x^0 - x^*\|^2$$

$$\|x^k - x^*\|^2 \geq q^{2k} \|x^0 - x^*\|^2$$

$$\|$$

$$\left(\left(1 - \frac{2\sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}\right)^2\right)^k$$

$$\sim$$

$$1 - \frac{2\sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}$$

$$\sim$$

$$\|x^k - x^*\|^2 \geq \left(1 - \sqrt{\frac{\mu}{L}}\right)^k \|x^0 - x^*\|^2$$

$$A \varphi_0 \left(1 - \sqrt{\frac{\mu}{L}}\right)^k \geq \varphi_k \geq B \left(1 - \sqrt{\frac{\mu}{L}}\right)^k \varphi_0$$

$A \geq B$ [illegible]