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

0. Коши [illegible]

$$Ax = b \iff Ax - b = 0$$

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2$$

$$\nabla = 2A^T(Ax - b) = 2(A^TAx - A^Tb)$$

Алгоритм 1 Градиентный спуск

Вход: размеры шагов $\{\gamma_k\}_{k=0} > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \dots, K-1$ **do**

2: Вычислить $\nabla f(x^k)$

3: $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$

4: **end for**

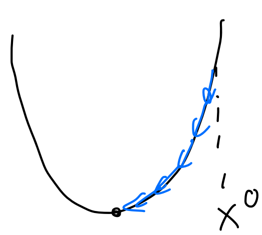
Выход: x^K

Физический смысл:

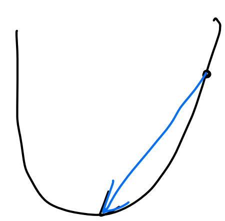
- против антиградиента – направление лок. убыв.
- выбирать шаг

$$\min_{x \in \mathbb{R}} x^2 \qquad x^0 = 1$$

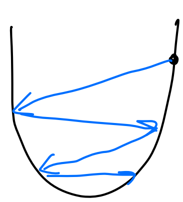
$$x^1 = x^0 - \gamma \cdot 2x^0$$



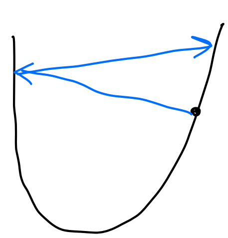
γ малое



γ оптимальное



γ большое



γ слишком большое

Теория сходимости

- $\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x-y\|_2$
- $\frac{\mu}{2}\|x-y\|_2^2 \le f(y) - f(x) - \langle \nabla f(x); y-x \rangle$

$$\|x^{k+1} - x^*\|_2^2 = \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2$$

подставляем ст. метода

$$= \|x^k - x^*\|_2^2 + \gamma_k^2 \|\nabla f(x^k)\|_2^2 - 2\gamma_k \langle \nabla f(x^k); x^k - x^* \rangle$$

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

$$= \|x^k - x^*\|_2^2 + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2 - 2\gamma_k \langle \nabla f(x^k); x^k - x^* \rangle$$

L - гладкость

$$\le \|x^k - x^*\|_2^2 + \gamma_k^2 L^2 \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k); x^k - x^* \rangle$$

! $\frac{\mu}{2}\|x-y\|_2^2 \le f(y) - f(x) - \langle \nabla f(x); y-x \rangle$

$$-\langle \nabla f(\underset{\substack{\uparrow \\ x^k}}{x}); x - \underset{\substack{\uparrow \\ x^*}}{y} \rangle \le f(y) - f(x) - \frac{\mu}{2}\|x-y\|_2^2$$

$$\le \|x^k - x^*\|_2^2 + \gamma_k^2 L^2 \|x^k - x^*\|_2^2 + 2\gamma_k \left(f(x^*) - f(x^k) - \frac{\mu}{2}\|x^k - x^*\|_2^2 \right)$$

$$= \left(1-\gamma_k\mu+\gamma_k^2L^2\right)\|x^k-x^*\|_2^2 + 2\gamma_k \underbrace{\left(f(x^*)-f(x^k)\right)}_{\leq 0}$$

$$\boxed{\|x^{k+1}-x^*\|_2^2 \leq \left(1-\gamma_k\mu+\gamma_k^2L^2\right)\|x^k-x^*\|_2^2}$$

$$\underbrace{\left(1-\gamma_k\mu+\gamma_k^2L^2\right)}_{<1}$$

$$\min_{\gamma_k} \Rightarrow \gamma_k^* = \frac{\mu}{2L^2}$$

$$\gamma_k \equiv \frac{\mu}{2L^2}$$

$$\|x^{k+1}-x^*\|_2^2 \leq \left(1-\frac{\mu^2}{2L^2}+\frac{\mu^2}{4L^2}\right)\|x^k-x^*\|_2^2$$

$$\leq \left(1-\frac{\mu^2}{4L^2}\right)\|x^k-x^*\|_2^2$$

Заменяем рекурсивно

$$\leq \left(1-\frac{\mu^2}{4L^2}\right)^2\|x^{k-1}-x^*\|_2^2$$

$$\cdots$$

$$\leq \left(1-\frac{\mu^2}{4L^2}\right)^{k+1}\|x^0-x^*\|_2^2$$

Скорости сходимости

- Сублинейная
  $$\|x^k-x^*\|_2 \leq \frac{C}{k^\alpha} \qquad C>0 \quad \alpha>0$$
- Линейная
  $$\|x^k-x^*\|_2 \leq Cq^k \qquad C>0 \quad q\in(0;1)$$

- Сверхлинейная

$$\|x^k - x^*\|_2 \le C q^{k^p} \qquad C>0 \quad q\in(0;1) \quad p>1$$

- Квадратичная

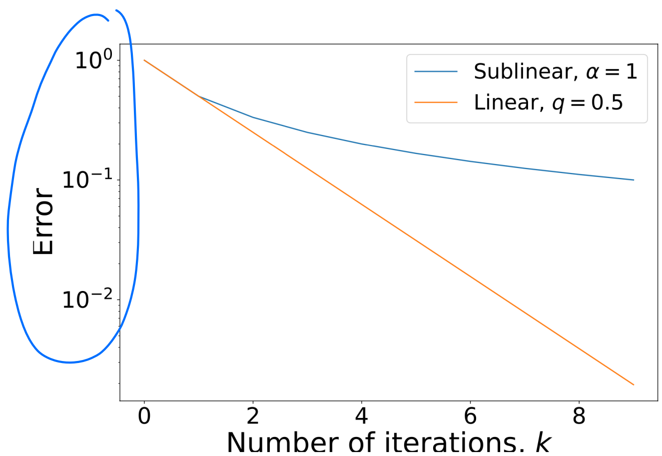
$$\|x^k - x^*\|_2 \le C q^{2^k} \qquad C>0 \quad q\in(0;1)$$

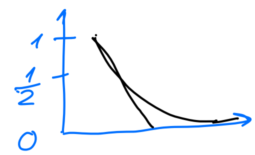
$$\|x^{k+1} - x^*\|_2 \le C \|x^k - x^*\|_2^2 \qquad C>0$$

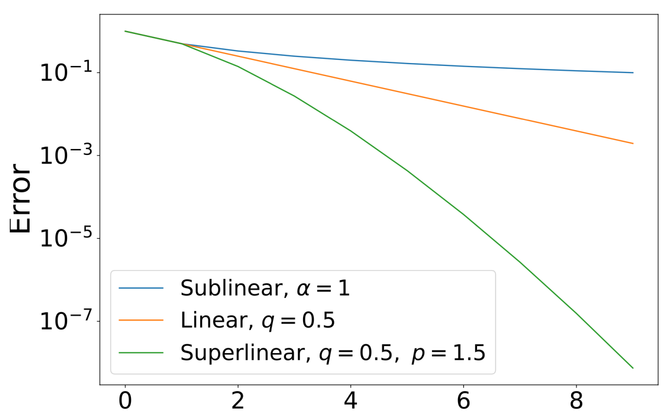
$C=1$ $\quad \|x^0 - x^*\|_2 = \frac{1}{2}$ $\quad \|x^1 - x^*\| = \frac{1}{4}$

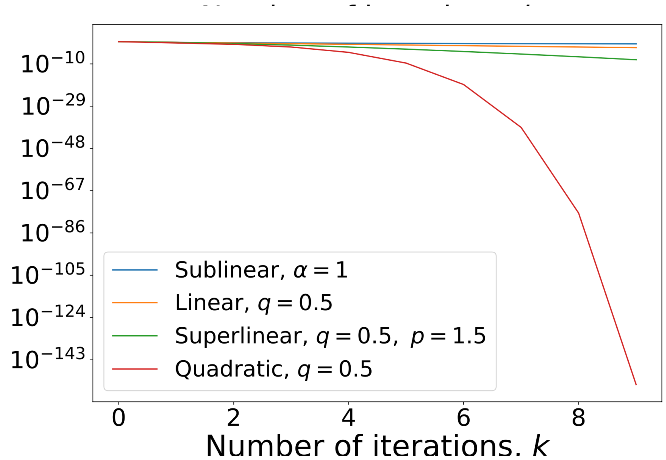
$$\frac{1}{2} \to \frac{1}{2^2} \to \frac{1}{(2^2)^2} \to \frac{1}{((2^2)^2)^2}$$

лог. шкала









$$\|x^k - x^*\|_2^2 \le \left(1 - \frac{\mu^2}{4L^2}\right)^k \|x^0 - x^*\|_2^2$$

Хотим $\|x^k - x^*\|_2^2 \le \varepsilon^2$

$$1 - x \le \exp(-x)$$

$$\le \exp\left(-\frac{\mu^2}{4L^2} \cdot k\right) \|x^0 - x^*\|_2^2 \le \varepsilon^2$$

$$\boxed{k \ge \frac{4L^2}{\mu^2} \log \frac{\|x^0 - x_*\|_2^2}{\varepsilon^2}}$$

плохой $\frac{L^2}{\mu^2}$ ← можно лучше

Теорема (свойства L - гладкой выпуклой функции)

Пусть дана L - гладкая *выпуклая* функция $f : \mathbb{R}^d \to \mathbb{R}$. Тогда для любых $x, y \in \mathbb{R}^d$ выполнено

оц. вып. функ. степень гладкости (пр. лек.)

$$0 \le f(y) - f(x) - \langle \nabla f(x), y - x \rangle \le \frac{L}{2}\|x - y\|_2^2$$

и

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le f(y).$$

Док-во: $1 \Rightarrow 2$

тек. точка

$$\Delta\ \varphi(y) = f(y) - \langle \nabla f(x); y \rangle$$

- φ - L_φ-гладкая? да, $L_\varphi = L$

$$\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\|_2 = \|\nabla f(y_1) - \cancel{\nabla f(x)} - \nabla f(y_2) + \cancel{\nabla f(x)}\|_2$$

$$= \|\nabla f(y_1) - \nabla f(y_2)\|_2 \le L\|x - y\|_2$$

- φ - выпуклой? да

$$0 \le \varphi(y_1) - \varphi(y_2) - \langle \nabla \varphi(y_2); y_1 - y_2 \rangle$$

?

$$= f(y_1) - \langle \cancel{\nabla f(x)}; y_1 \rangle - f(y_2) + \langle \cancel{\nabla f(x)}; y_2 \rangle$$

$$- \langle \nabla f(y_2) - \cancel{\nabla f(x)}; y_1 - y_2 \rangle$$

$$= \underbrace{f(y_1) - f(y_2) - \langle \nabla f(y_2); y_1 - y_2 \rangle}_{\text{справед. вып. } f}$$

для φ верно: $f \to \varphi \quad x \to y \quad y \to y - \frac{1}{L}\nabla\varphi(y)$

Теорема (свойства L - гладкой выпуклой функции)

Пусть дана L - гладкая *выпуклая* функция $f : \mathbb{R}^d \to \mathbb{R}$. Тогда для любых $x, y \in \mathbb{R}^d$ выполнено

$$0 \le f(y) - f(x) - \langle \nabla f(x), y - x \rangle \le \frac{L}{2}\|x - y\|_2^2$$

$y - \frac{1}{L}\nabla\varphi(y) \qquad y$

и

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le f(y).$$

$$\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) - \varphi(y) - \langle \nabla\varphi(y); \cancel{y} - \frac{1}{L}\nabla\varphi(y) - \cancel{y} \rangle \le \frac{L}{2}\|\cancel{y} - \frac{1}{L}\nabla\varphi(y) - \cancel{y}\|_2^2$$

$$\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) - \varphi(y) + \langle \nabla\varphi(y); \frac{1}{L}\nabla\varphi(y) \rangle \le \frac{L}{2} \cdot \frac{1}{L^2}\|\nabla\varphi(y)\|_2^2$$

$$\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \le \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

- $\nabla\varphi(x) = \nabla f(x) - \nabla f(x) = 0$

x – минимум φ

$$\varphi(x) \le \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \le \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

$$\varphi(x) \le \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

$$\varphi(y) = f(y) - \langle \nabla f(x); y \rangle$$

$$f(x) - \langle \nabla f(x); x \rangle \le f(y) - \langle \nabla f(x); y \rangle - \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|_2^2 \quad \blacksquare$$

$$\|\nabla f(y) - \nabla f(x)\|_2^2 \le 2L \left(f(y) - f(x) - \langle \nabla f(x); y - x \rangle \right)$$

Вернемся к док-ву град. спуска:

$$\|x^{k+1} - x^*\|_2^2 = \|x^k - x^*\|_2^2 + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2 - 2\gamma_k \langle \nabla f(x^k); x^k - x^* \rangle$$

$y = x^k \qquad x = x^* \qquad \nabla f(x^*) = 0$

$$\le \|x^k - x^*\|_2^2 + \gamma_k^2 \cdot 2L \left(f(x^k) - f(x^*) + \overset{0}{\cancel{\langle\,\rangle}} \right) - 2\gamma_k \langle \nabla f(x^k); x^k - x^* \rangle$$

$$\le \|x^k - x^*\|_2^2 + 2L\gamma_k^2 \left(f(x^k) - f(x^*) \right) + 2\gamma_k \left(f(x^*) - f(x^k) - \frac{\mu}{2} \|x^k - x^*\|_2^2 \right)$$

$$\|x^{k+1} - x^*\|_2^2 \le (1 - \gamma_k \mu) \|x^k - x^*\|_2^2 + \underbrace{2\gamma_k}_{\ge 0} \underbrace{(L\gamma_k - 1)}_{< 0} \underbrace{\left(f(x^k) - f(x^*) \right)}_{\ge 0}$$

$$\gamma_k \le \frac{1}{L}$$

$$\|x^{k+1} - x^*\|_2^2 \le (1 - \gamma_k \mu) \|x^k - x^*\|_2^2$$

$$\gamma_k \equiv \frac{1}{L}$$

$$\le \left(1 - \frac{\mu}{L}\right) \|x^k - x^*\|_2^2$$

$$\le \left(1-\frac{\mu}{L}\right)^{k+1} \|x^0 - x^*\|_2^2$$

кол-во итераций до точн. ε

$$k \ge \frac{L}{\mu} \log \frac{\|x^0 - x^*\|_2^2}{\varepsilon^2}$$

было $\frac{4L^2}{\mu^2}$

Теорема сходимость градиентного спуска для L-гладких и μ-сильно выпуклых функций

Пусть задача безусловной оптимизации (1) с L-гладкой, μ-сильно выпуклой целевой функцией f решается с помощью градиентного спуска. Тогда справедлива следующая оценка сходимости

$$\|x^K - x^*\|_2^2 \le \left(1 - \frac{\mu}{L}\right)^K \|x^0 - x^*\|_2^2.$$

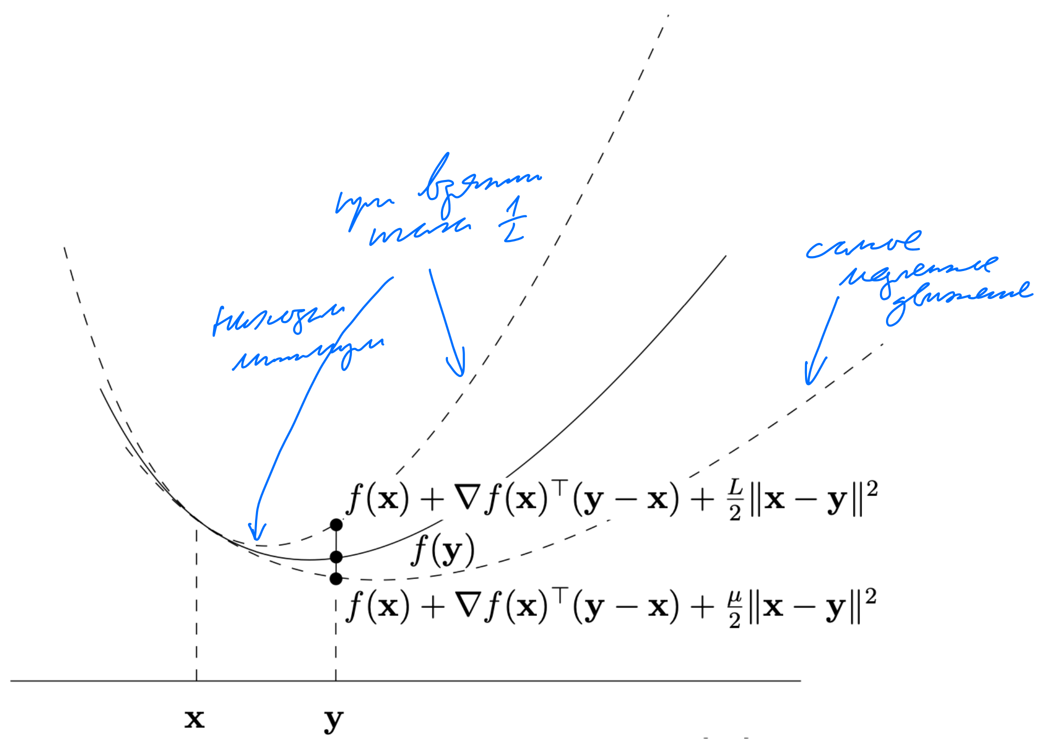
Более того, чтобы добиться точности ε по аргументу, необходимо

$$K = O\left(\frac{L}{\mu} \log \frac{\|x^0 - x^*\|_2}{\varepsilon}\right) = \tilde{O}\left(\frac{L}{\mu}\right) \text{ итераций.}$$

2 должна быть

$$\frac{L}{\mu} \to \infty \qquad \frac{\|x^0 - x^*\|}{\varepsilon} \to \infty$$

Физический смысл:



Подбор шага:

- Поляка – Шора:

$$\|x^{k+1}-x^*\|_2^2 = \|x^k-x^*\|_2^2 + \gamma_k^2 \|\nabla f(x^k) - \cancel{\nabla f(x^*)}\|_2^2 - 2\gamma_k \langle \nabla f(x^k) ; x^k - x^* \rangle$$

выпуклость $\mu = 0$

$$\le \|x^k-x^*\|_2^2 + \gamma_k^2 \|\nabla f(x^k)\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*)\right)$$

$\min\limits_{\gamma_k}$

$$\gamma_k^* = \frac{f(x^k) - f(x^*)}{\alpha_k \|\nabla f(x^k)\|_2^2}$$ ← шаг Поляка – Шора

не всегда знаем

$\alpha_k > 1$

- Наискорейший спуск

$$\gamma_k^* = \underbrace{\arg\min_{\gamma > 0} \left\{ f\left(x^k - \gamma \nabla f(x^k)\right) \right\}}_{\text{метод дихот. / золотое сечение}}$$