$$\min_{x \in Q} f(x)$$

1) $Q = \mathbb{R}^d$

2) Q – простое мн-во

Оракулы:

1) нулевого: $f(x)$
2) первого: $\nabla f(x)$
3) второго: $\nabla^2 f(x)$

p-порядка $\nabla^p f(x)$

Критерии сходимости:

1) $\|x^k - x^*\|_2^2 \le \varepsilon$ x^* – решение $\|x^k - x^{k-1}\|_2^2$
2) $f(x^k) - f^* \le \varepsilon$ $f(x^k) - f(x^{k-1})$

   $f^* = \min f(x)$
3) $\|\nabla f(x^k)\|_2^2 \le \varepsilon$

Скорости сходимости:

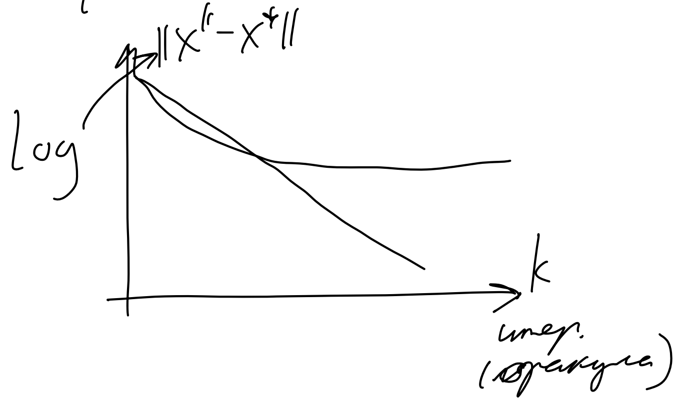
1) Сублинейная

$$\|x^k - x^*\| \le \frac{C}{k^\alpha} \qquad \alpha > 0 \quad C > 0$$

$$\frac{1}{k} \qquad \frac{1}{k^2}$$

2) Линейная (геом.)

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



3) Сверхлинейная

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

4) Квадратичная

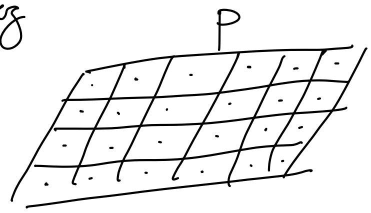
$$\|x^k - x^*\| \le C q^{2^k} \qquad q \in (0;1) \quad C > 0$$

Пример для липш. ф. метод оптим. дает плохие гарантии

$$|f(x)-f(y)| \le M\|x-y\|_\infty = M \max_i |x_i - y_i|$$

$$\min_{x\in B_\infty(1)} f(x)$$

1) Метод

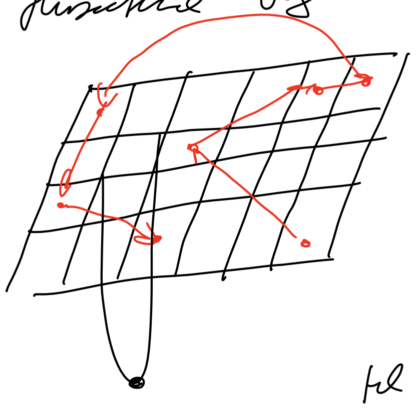


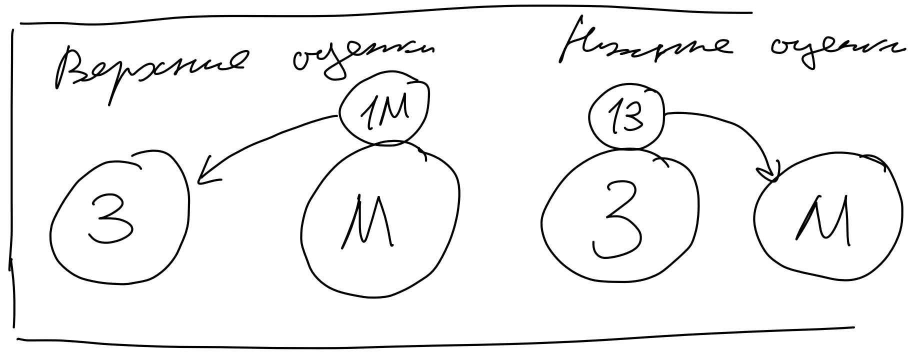
p^d

$f(x)-f^* \le \varepsilon$

$\left(\frac{M}{2\varepsilon}\right)^d$ - точек

2) Нижние оценки





не менее $\left(\frac{M}{2\varepsilon}\right)^d$ точек посмотреть

- Выпуклость

  $f(x)$ выпукла на $\mathbb{R}^d$, если $\forall x,y \in \mathbb{R}^d$

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

μ- сильная выпуклость

- липшиц. ф.
- Липшицевость градиента $\forall x,y \in \mathbb{R}^d$

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

Теорема Если $f(x)$ имеет L-липшицев градиент, то $\forall\, x, y \in \mathbb{R}^d$

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

Док-во:

$$f(y) - f(x) \overset{\text{ф. Н-Л}}{=} \int_0^1 \langle \nabla f(x + \tau \underbrace{(y-x)}_{\text{вектор}}); y - x \rangle \, d\tau$$

$$\langle \nabla f(\underbrace{\vec{r}(\tau)}_{x + \tau(y-x)}); \underbrace{d\vec{r}(\tau)}_{(y-x)\, d\tau} \rangle$$

x, y

τ от $[0;1]$

$x + \tau(y - x)$

$$= \underbrace{\langle \nabla f(x); y - x \rangle}_{+} + \int_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x); y - x \rangle \, d\tau$$

$$\left| f(y) - f(x) - \langle \nabla f(x); y - x \rangle \right|$$

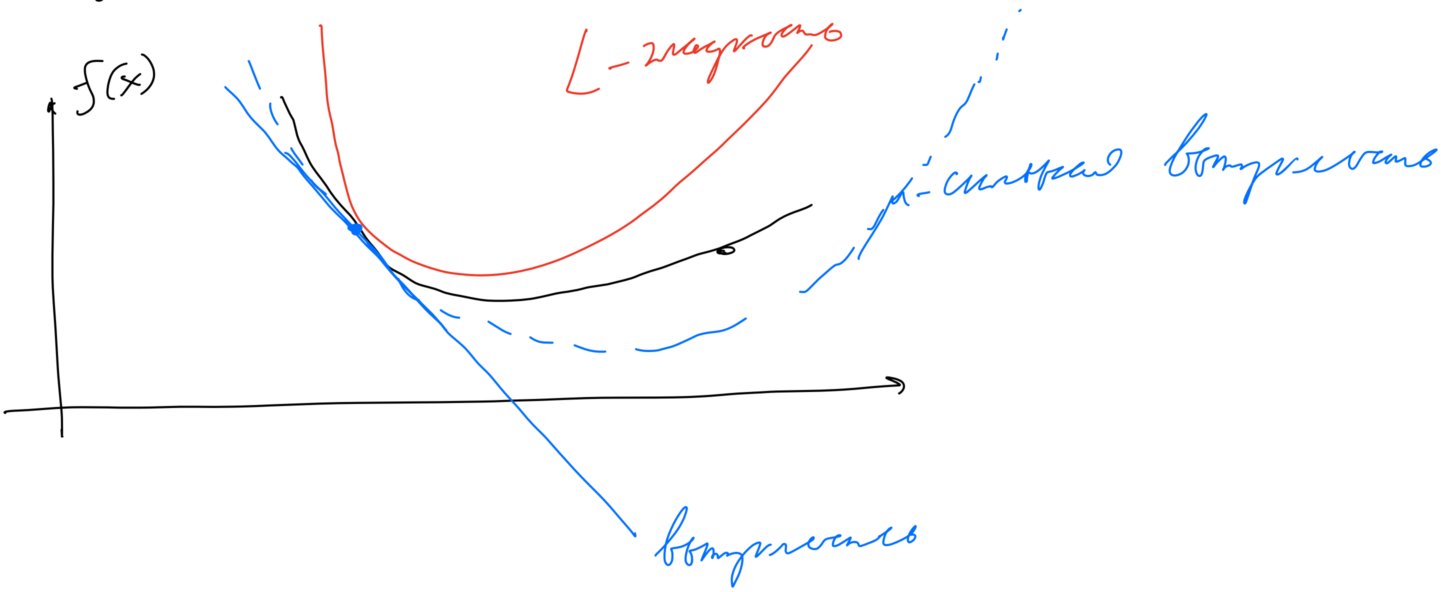
$$= \left| \int_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x); y - x \rangle \, d\tau \right|$$

$$\le \int_0^1 \underbrace{\left| \langle \nabla f(x + \tau(y-x)) - \nabla f(x); y - x \rangle \right|}_{\text{КБШ}} d\tau$$

$$\le \int_0^1 \|\nabla f(x + \tau(y-x)) - \nabla f(x)\| \cdot \|y - x\| \, d\tau$$

$$\underset{L\text{-гладкость}}{\le} \int_0^1 L \tau \|x - y\|^2 d\tau = L\|x-y\|^2 \int_0^1 \tau \, d\tau = \frac{L\|x-y\|_2^2}{2}$$

∎



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

L-гладк.

μ-сильно вып.

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

$$x^k - \gamma \nabla f(x^k)$$

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

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

$$\le \|x^k - x^*\|_2^2 - 2\gamma \langle \nabla f(x^k); x^k - x^* \rangle + \gamma^2 \|\nabla f(x^k) - \underbrace{\nabla f(x^*)}_{=0}\|_2^2$$

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

L-гладк.

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

μ-convex bound

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

$$= (1 - \gamma\mu + \gamma^2 L^2)\|x^k - x^*\|_2^2 - 2\gamma(f(x^k) - f(x^*))$$

$$\leq (1 - \gamma\mu + \gamma^2 L^2)\|x^k - x^*\|_2^2$$

min γ

$$-\mu + 2\gamma L^2 = 0 \qquad \gamma = \frac{\mu}{2L^2}$$

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

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

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

$$\dots$$

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

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

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

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

плохой результат можно лучше

Ещё одно св-во для L-гладких и выпуклых ф.

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

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

а) $\Rightarrow$ б). — доказать

Док-во: зафикс. x

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

выпуклая L-гладкая $y^* = x$

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

- $\nabla\varphi(y^*) = 0$ $\quad \nabla f(y^*) - \nabla f(x) = 0$

$$\varphi(x) = \varphi(y^*) \le \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \le$$

L-гладкости φ

см. а). $\tilde{y} = y - \frac{1}{L}\nabla\varphi(y)$ $\quad \tilde{x} = y$

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

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

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

подставить φ:

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

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

$$\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \le 2L\left(f(x^k) - f(x^*) - \underbrace{\underbrace{\langle \nabla f(x^*); x^k - x^*\rangle}_{= 0}}_{0}\right)$$

Док-во сходимости град. спуска:

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

μ-сильн. + L-гладкость

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

$$\|x^{k+1} - x^*\|_2^2 = (1-\gamma\mu)\|x^k - x^*\|_2^2 \underbrace{- 2\gamma(1-\gamma L)\underbrace{(f(x^k) - f(x^*))}_{\ge 0}}_{?\ \le 0}$$

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

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

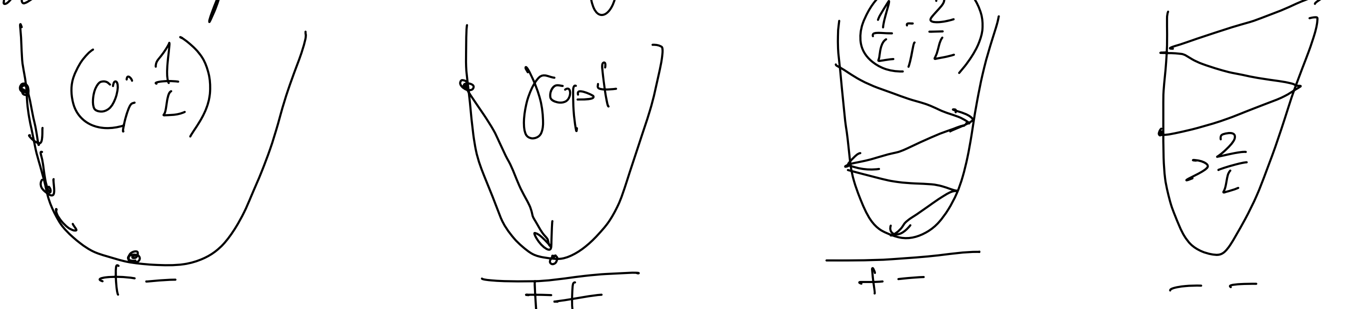
$\gamma = \frac{1}{L}$ хуже $\frac{\mu}{L^2}$

$$\left(1 - \frac{\mu}{L}\right)$$

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

правильная оценка

- шаг ограничен $\gamma \in \left(0; \frac{2}{L}\right)$ $\gamma_{opt} = \frac{1}{L}$



Стратегии подбора шага:

1) const $\frac{1}{L}$

2) наискорейший: $\gamma_k = \operatorname{argmin}_{\gamma \in \mathbb{R}} \left(\underbrace{f(x^k - \gamma \nabla f(x^k))}_{\text{выпукла}} \right)$

бинпоиск, золотое

3) $\frac{1}{k+1}$; $\frac{1}{\sqrt{k+1}}$ (больше про стохастику)

4) адаптивный подбор L_k $\quad \gamma_k = \frac{1}{L_k}$

5) Armijo
Wolfe
Goldstein

6) Поляка - Шора $\quad \gamma_k = \frac{f(x^k) - \cancel{f(x^*)}}{\alpha \, \|\nabla f(x^k)\|_2^2}$

? $= 0$

подобрать

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

min по γ_k