Гладкость (L-гладкость / L-липшицевость градиента)

Определение L-гладкой функции

Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что данная функция имеет L-Липшицев градиент (говорить, что она является L-гладкой), если для любых $x, y \in \mathbb{R}^d$ выполнено

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

Теорема (свойство L - гладкой функции)

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

$\frac{\mu}{2}\|x-y\|_2^2$ — сильная

для вып.

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

для выпуклой

Док-во:

Формула Ньютона-Лейбница:

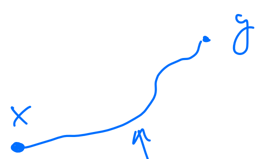
$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + \tau(y-x)); y - x\rangle d\tau$$

Общий вид

$r(t)$

$t \in [a, b]$

$$\int_a^b \langle \nabla f(r(t)); r'(t)\rangle dt$$



параметризация

$r(t) = x + t(y-x)$

$t \in [0, 1]$

В нашем случае

$t \in [0; 1]$

$$\int_0^1 \langle \nabla f(\overbrace{x + t(y-x)}^{r(t)}); y - x\rangle dt$$

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

Скал. прод. влево:

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

$$|\Sigma| \le \Sigma |\ |$$

$$\le \int_0^1 |\langle \overbrace{\nabla f(x+\tau(y-x)) - \nabla f(x)}; y-x\rangle| \, d\tau$$

КБШ $\langle a;b\rangle \le \|a\|_2 \cdot \|b\|_2$

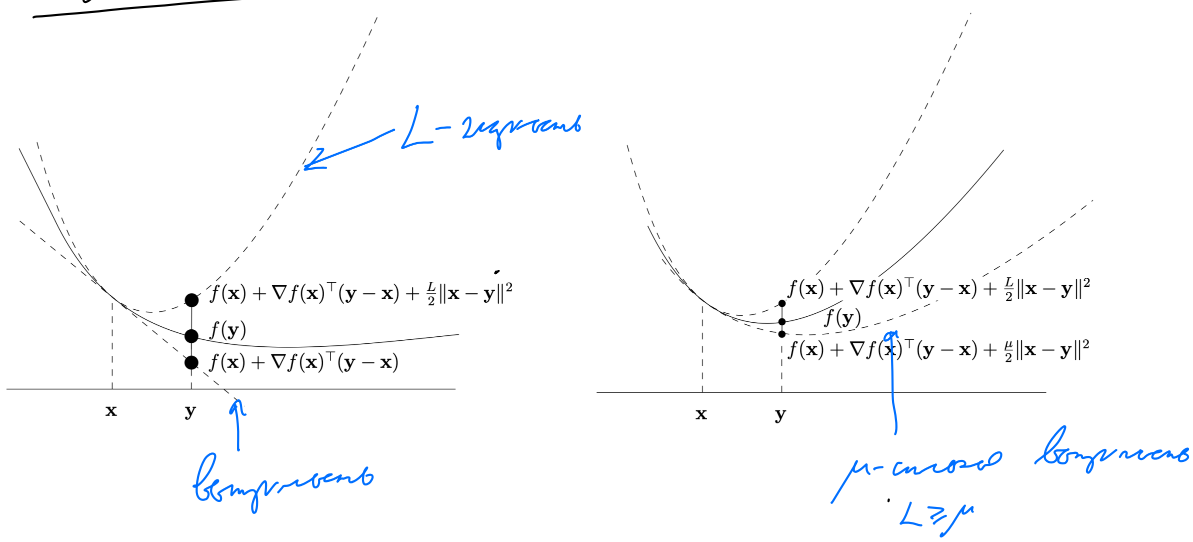
$$\le \int_0^1 \|\nabla f(x+\tau(y-x)) - \nabla f(x)\|_2 \cdot \|y-x\|_2 \, d\tau$$

L- гладкость

$$\le \int_0^1 L\|\cancel{x}+\tau(y-x)-\cancel{x}\|_2 \cdot \|y-x\|_2 \, d\tau$$

$$= L\|y-x\|_2^2 \int_0^1 \tau \, d\tau = \frac{L}{2}\|y-x\|_2^2 \quad \blacksquare$$

Физический смысл L-гладкости



Градиентный спуск

$$\min_{x\in\mathbb{R}^d} f(x) \qquad (1)$$

безусловная (в начале)

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** $k = 0, 1, \ldots, K-1$ **do**

2: Вычислить $\nabla f(x^k)$

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

4: **end for**

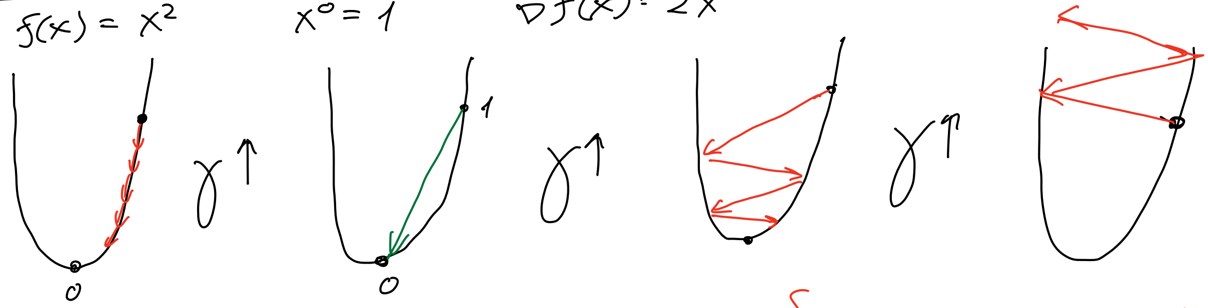
Выход: x^K

Идея: антиград указывает на локальные св-ва убывания

шаг / learning rate $\in \mathbb{R}_+$ (положит.)

Влияние шага

$f(x) = x^2 \qquad x^0 = 1 \qquad \nabla f(x) = 2x$



шаг маленький — шаг оптимален — шаг большой — шаг огромный

$$x^1 = x^0 - \frac{1}{1000} \cdot 2x_0$$

Док-во сходимости: $\gamma_k \equiv \gamma$

1) f — L-гладкая $\quad \|\nabla f(x) - \nabla f(y)\|_2 \le L\|x-y\|_2$

2) f — μ-сильно выпуклая

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

$\|x^{k+1} - x^*\|_2^2 \leftarrow$ как близок к решению на тек. шаге

x^* ↑ решение

$$\underbrace{\|x^{k+1} - x^*\|_2^2 <}_{\text{хотим!}} \underbrace{\|x^k - x^*\|_2^2 \to 0}_{\text{хотим!}}$$

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

$$= \|\underbrace{x^k - x^*}_{a} \underbrace{- \gamma\nabla f(x^k)}_{b}\|_2^2$$

$\|a+b\|_2^2 = \|a\|_2^2 + 2\langle a;b\rangle + \|b\|_2^2$

$\|a+b\|_2^2 = \langle a+b; a+b\rangle$

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

$\nabla f(x^*) = 0 \qquad \|\nabla f(x^k) - \nabla f(x^*)\|_2^2 = \|\nabla f(x^k)\|_2^2$

L - гладкость

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

сильная выпуклость

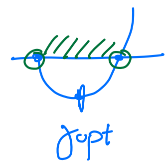
$$-\langle \nabla f(x^k); x^k - x^* \rangle \le -\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right)$$

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

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

$$\le (\underbrace{1 - \gamma\mu + \gamma^2 L^2}_{< 1})\|x^k - x^*\|_2^2$$

$\gamma_{opt} = \frac{\mu}{2L^2}$ (вершину параболы)



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

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

Запустим рекурсию

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

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

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

Скорости сходимости:

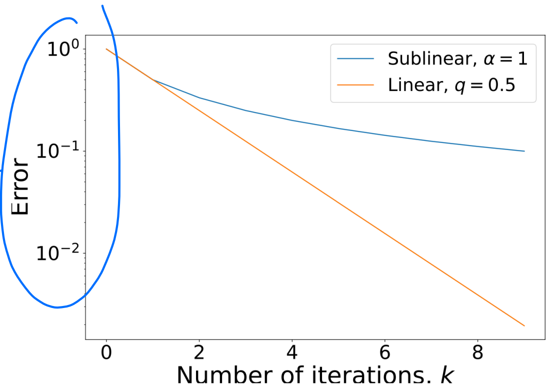
1) сублинейная $\quad C > 0 \quad \alpha > 0$

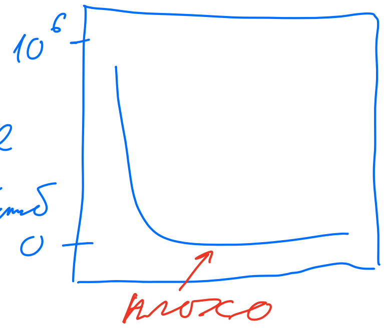
$$\|x^k - x^*\|_2^2 \le \frac{C}{k^\alpha}$$

2) линейная сходимость $\quad \rho \in (0;1) \quad C > 0$

$$\|x^k - x^*\|_2^2 \le C\rho^k$$

лог масштаб





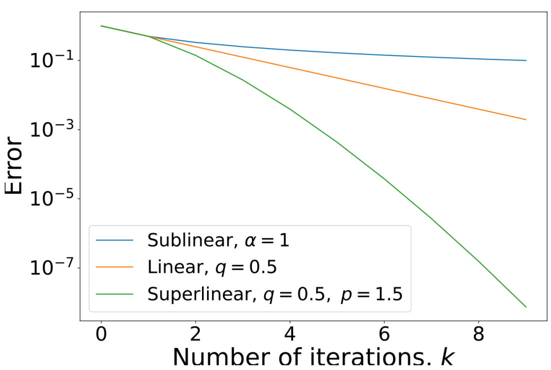
3) Сверхлинейная

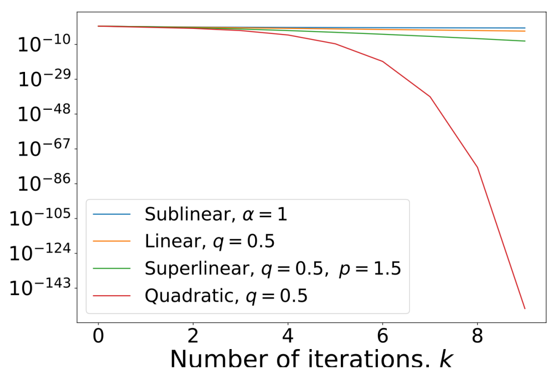
$$\|x^k - x^*\|_2^2 \le C\, q^{k^p} \qquad C>0 \quad q\in(0;1) \quad p>1$$

4) Квадратичная

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

$$\|x^k - x^*\| \le C\, \|x^{k-1} - x^*\|_2^2$$





Вернемся к нашему случаю:

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

$K = ?$ если хочу точность ε

$1 - x \le \exp(-x)$, $x \in (0;1)$

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

логарифмируем:

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

$$K = O\left(\frac{L^2}{\mu^2} \log \frac{\|x^0 - x^*\|_2^2}{\varepsilon}\right)$$

$\frac{L}{\mu} \to \infty$

$\varepsilon \to 0$

без числ. факторов

$$= \tilde{O}\left(\frac{L^2}{\mu^2}\right)$$

без лог. факторов

Плохая оценка!

Теорема (свойства L - гладкой выпуклой функции)

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

обх. выше

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

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: из выпуклости

второе св-во: $\triangleleft$ $\varphi(y) = f(y) - \langle \nabla f(x); y\rangle \quad \forall x$

что мож сказать про φ?

1) φ выпуклая как сумма выпуклых

2) φ — L_φ - гладкая

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

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

$$L_\varphi = L$$

3) $\nabla\varphi(y^*) = 0 \qquad \nabla f(y^*) - \nabla f(x) = 0 \qquad y^* = x$ минимум

Первое св-во для φ:

$$\varphi(y_1) - \varphi(y_2) - \langle \nabla\varphi(y_2); y_1 - y_2\rangle \le \frac{L_\varphi}{2}\|y_1 - y_2\|_2^2$$

$$y_1 = y - \frac{1}{L}\nabla\varphi(y) \qquad y_2 = y$$

$$\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) - \varphi(y) + \langle \nabla\varphi(y); \frac{1}{L}\nabla\varphi(y)\rangle$$

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

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

$y^* = x$

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

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

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

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

Обратно к град спуску:

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

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

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

по новому св-ву

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

μ - сильно выпуклая

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

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

$$2\gamma^2 L - 2\gamma = 2\gamma(\gamma L - 1) < 0$$

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

$$\boxed{\gamma_{opt} = \frac{1}{L}}$$

Разные

$\frac{1}{L} \cdot \frac{\mu}{L}$

маленькое

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

Результат

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

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

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

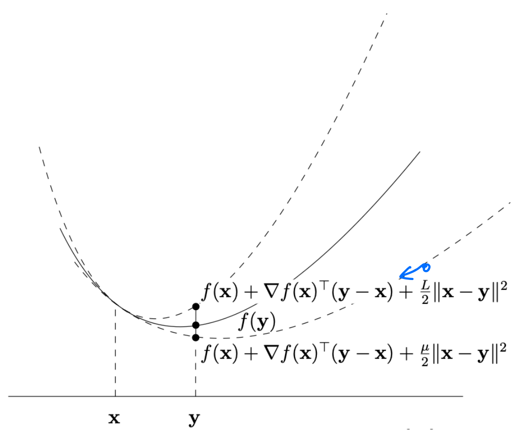
Более того, чтобы добиться точности ε по аргументу, необходимо

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

неулучшаемо!

хуже чем $\frac{L^2}{\mu^2}$!

Физический смысл:



1) шаг берется исходя из верхней параболы $\gamma \sim \frac{1}{L}$

2) ∇ в худшем случае отрезает св. во нижней параболы

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

Другие оценки

докажем

	μ-сильно выпуклая	выпуклая	невыпуклая
L-гладкая	$O\left(\frac{L}{\mu}\log\frac{\|x^0-x^*\|_2}{\varepsilon}\right)$	$O\left(\frac{L\|x^0-x^*\|_2^2}{\varepsilon}\right)$	$O\left(\frac{L(f(x^0)-f^*)}{\varepsilon^2}\right)$
M-липшицева	$O\left(\frac{M^2}{\mu^2\varepsilon}\right)$	$O\left(\frac{M^2\|x^0-x^*\|_2^2}{\varepsilon^2}\right)$	2 лекция

$\|x^k - x^*\|_2^2 \sim \varepsilon$

$f(x^k) - f(x^*) \sim \varepsilon$

$\|\nabla f(x^k)\|_2 \sim \varepsilon$