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

$$\gamma_k = \gamma$$

• Если хотим $\|x^k - x^*\|_2^2 \le \varepsilon$, тогда

плохой результат! $\longrightarrow$ $k = O\left(\frac{L^2}{\mu^2} \log \frac{\|x^0 - x^*\|_2^2}{\varepsilon}\right)$

O-нотация = прятать численные константы $(2, 6, \dots)$

$\left(\frac{L}{\mu} \to \infty \quad \varepsilon \to 0\right)$

Уже доказали: если f явл. L-гладкой и выпуклой

$$\text{a) } 0 \overset{\text{опр. вып}}{\le} f(y) - f(x) - \langle \nabla f(x); y - x \rangle \overset{\text{гладкость}}{\le} \frac{L}{2}\|y - x\|_2^2$$

Новое:

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

Докажем, что a) $\Rightarrow$ б).

Док-во:

$$\varphi(y) = f(y) - \langle \nabla f(x); y \rangle \quad \text{для } \forall x \in \mathbb{R}^d$$

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

• φ L_φ-гладкий? по определению

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

$$= \|\nabla f(y_1) - \nabla f(y_2)\|_2$$

L-гладкость f

$$\le L\|y_1 - y_2\|_2 \Rightarrow \boxed{L_\varphi = L}$$

• φ выпукла ? по определению

? ← если да, то выпукла

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

(+) по выпуклости f

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

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

$$= \underline{f(y_1) - f(y_2) - \langle \nabla f(y_2); y_1 - y_2 \rangle} \qquad \boxed{\text{выпукла}}$$

• $\nabla\varphi(y^*) = 0 \Rightarrow \nabla f(y^*) - \nabla f(x) = 0$ $\quad \boxed{y^* = x}$

(возможно, не ед. опт.)

С помощью фактов выше:

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

x глоб. оптимум

φ L-гладкая

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

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

$$x = y$$

$$\le \varphi(y) + \langle \nabla\varphi(y); \underbrace{y - x}_{-\frac{1}{L}\nabla\varphi(y)} \rangle + \frac{L}{2}\|\underbrace{y - x}_{-\frac{1}{L}\nabla\varphi(y)}\|^2$$

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

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

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

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

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

∎

Перепишем: L-гладкость + выпуклость

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

Еще было μ-сильная выпуклость

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

Док-во сходимости град. спуска: $\gamma_k = \gamma$

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

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

μ-сильная выпуклость f

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

$\nabla f(x^*) = 0$ $\quad \min\limits_{x \in \mathbb{R}^d} f(x)$

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

L-гладкость в новом виде

$$\le \|x^k - x^*\|_2^2 - 2\gamma\left(f(x^k) - f(x^*) + \frac{\mu}{2}\|x^k - x^*\|_2^2\right) + \gamma^2 \cdot 2L\left(f(x^k) - f(x^*) - \langle \underbrace{\cancel{\nabla f(x^*)}}_{0}; x^k - x^* \rangle\right)$$

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

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

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

Такое же как в прош. раз:

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

- Тогда $\gamma \leq \frac{\mu}{2L^2}$
- Сейчас $\boxed{\gamma \leq \frac{1}{L}}$ лучше!

$\mu \ll L$

Подставим $\gamma = \frac{1}{L}$

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

запускаем рекурсию

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

$$\ldots$$

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

$1 - x \leq \exp(-x) \quad x \in (0;1)$

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

Хотим $\|x^{k+1} - x^*\|_2^2 \leq \varepsilon$, тогда

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

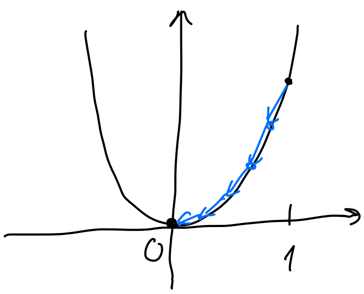
Было $\frac{L^2}{\mu^2}$ значительно хуже → $\boxed{k+1 = \frac{L}{\mu}\log\frac{\|x^0-x^*\|_2^2}{\varepsilon}}$ ← не улучшаемо для GD (нижняя оценка)

Про шаги и функцию

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

пример $\min\limits_{x\in\mathbb{R}^d} \frac{1}{2}x^2 \qquad x^0 = 1$

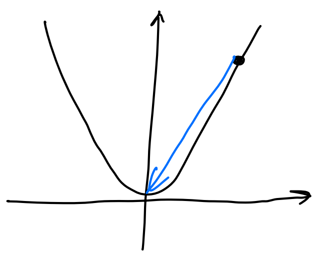
- с маленьким шагом $\gamma \in (0; \frac{1}{L})$



сходится
не очень шибко быстрее

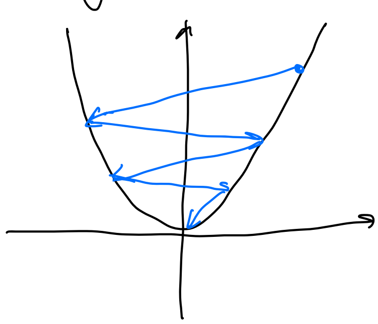
- $\gamma_{opt} = \frac{1}{L} \quad L = 1$

$x^1 = x^0 - \gamma_{opt}x^0 = 0$



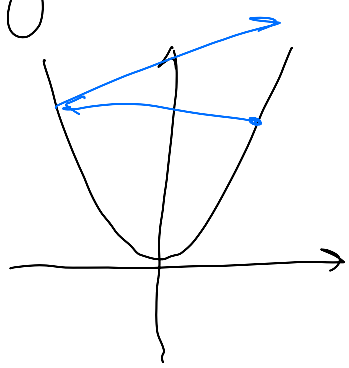
сходится за 1 итерацию

- с большим шагом $\gamma \in (\frac{1}{L}; \frac{2}{L})$

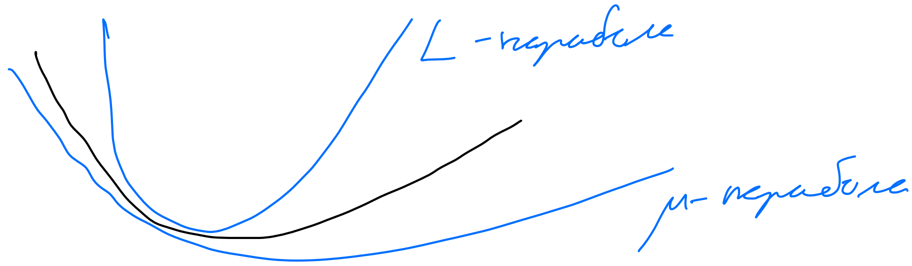


сходится
не шибко быстрее

- с огромными шагами $\gamma \geq \frac{2}{L}$



расходится



подбираем шаг, исходя из макс. резкости изменения

$$f \to \frac{L}{2}\|y-x\|_2^2 \qquad \gamma \sim \frac{1}{L}$$

шагаем в худшем случае, исходя из медленных св-в

$$f \to \frac{\mu}{2}\|y-x\|_2^2 \qquad (1-\gamma\mu)$$

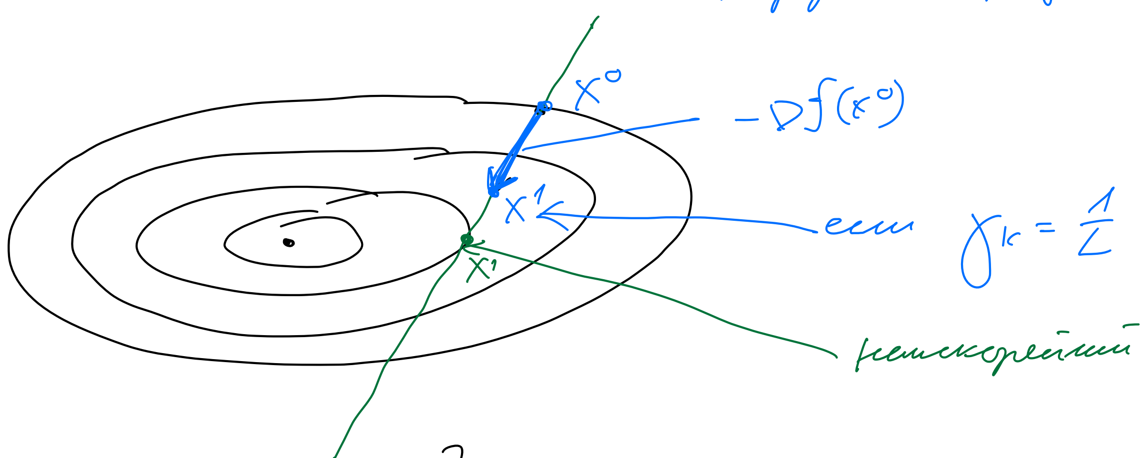
Итого: $\left(1-\frac{\mu}{L}\right)$

Стратегии подбора шага:

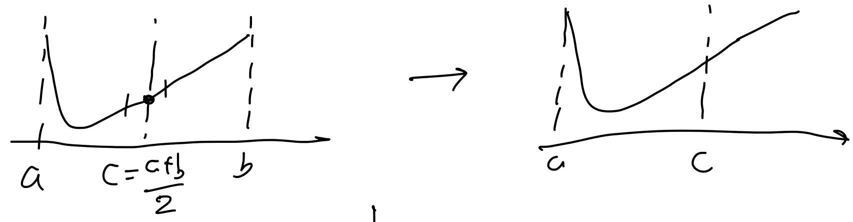
- const $\gamma = \frac{1}{L}$
- наискорейший спуск:

$$\gamma_k = \operatorname*{argmin}_{\gamma\in\mathbb{R}} \left(f(x^k - \gamma\nabla f(x^k))\right)$$

нелинейным образом



как решить $\operatorname*{argmin}_{\gamma}$?
— явная формула
— выпукла по γ, можно бинпоиск золотого сечения



$\log_2 \frac{b-a}{\varepsilon}$ итераций

- $\gamma_k = \frac{1}{k+1}$; $\frac{1}{\sqrt{k+1}}$ по номеру итерации
- Поляка – Шора (из док-ва с $\mu = 0$)

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

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

проблема

- знаем $f(x^*)$ (например 0)
- иначе оценить $f(x^*)$

$\alpha > 1$

- Адаптивный подход: L_k может меняться

$$\Downarrow$$

$$\gamma_k = \frac{1}{L_k}$$ (подбор L_k по св-ву гладкости)

- Armijo, Wolfe, Goldstein } Д/З

Сходимость GD для других классов задач

- L-гладкие, выпуклые функции

$$k = O\left(\frac{L\|x^0 - x^*\|_2^2}{\varepsilon}\right) \quad f(x^k) - f^* \leq \varepsilon$$

(сходимость по функционалу)

- L-гладкие, невыпуклые функции

$$k = O\left(\frac{L(f(x^0) - f^*)}{\varepsilon^2}\right) \quad \|\nabla f(x^k)\|_2 \leq \varepsilon$$

(только к стац. точке)

f^* в обоих случаях может не сущ.