Задача безусловной оптимизации:

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

Метод град. спуска:

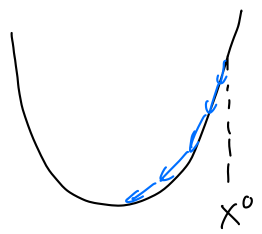
NB О. Коши $Ax=b \Rightarrow \min_{x\in\mathbb{R}^d}\|Ax-b\|^2$

$$\boxed{x^{k+1} = x^k - \underset{\uparrow \text{ шаг } > 0}{\gamma_k} \nabla f(x^k)}$$

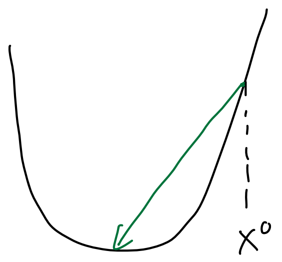
NB x_i^k ← номер итер (k), ← номер коорд (i)

Физика: ∇f – направл. роста $\Rightarrow$ $-\nabla f$ – направл. убыв-я

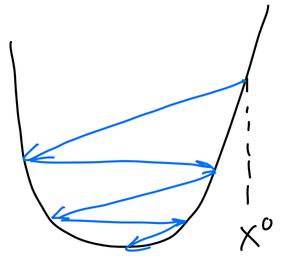
подбор $\gamma_k \equiv \gamma$: $\min_{x\in\mathbb{R}} \frac{1}{2}x^2$ $x^0=1$ $x^{k+1} = x^k - \gamma x^k = (1-\gamma)x^k$



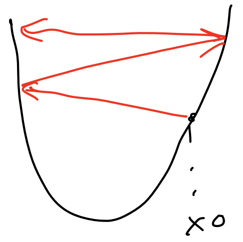
γ маленький сходимся, но медленно



γ идеальный (оптим.) сходимся за 1 итер.



γ больше сходимся, но снова медленно



γ огромный расходимся

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

- f – L-гладкая
- f – μ-сильно выпуклая

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

$$\|x^{k+1} - \underset{\uparrow}{x^*}\|^2 \overset{\text{итер.}}{=} \|\underline{x^k} - \underline{\underline{\gamma\nabla f(x^k)}} \underline{- x^*}\|^2$$

ед. для μ-сильно в. ф.

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

$$\|\nabla f(x^k)\|^2 = \|\nabla f(x^k) - \underset{\substack{\| \\ 0}}{\nabla f(x^*)}\|^2 \underset{\substack{\uparrow \\ \text{L-гладкость}}}{\le} L^2 \|x^k - x^*\|^2$$

$$-\langle \nabla f(x^k); x^k - x^*\rangle = \langle \nabla f(x^k); x^* - x^k\rangle \underset{\substack{\uparrow \\ \mu\text{-сильн. вып.}}}{\le} f(x^*) - f(x^k) - \frac{\mu}{2}\|x^k - x^*\|^2$$

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

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

$$\|x^{k+1} - x^*\|^2 \le (\underbrace{1 - \gamma\mu + \gamma^2 L^2}_{\in (0;1)})\|x^k - x^*\|^2$$

оптимально: $1 - \gamma\mu + \gamma^2 L^2 \to \min\limits_{\gamma > 0}$ $\boxed{\gamma = \frac{\mu}{2L^2}}$

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

рекурсия

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

$$\le \ldots$$

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

Сходимость (плохая!) град. спуска для L-гладких, μ-сильно вып. ф. $\gamma = \frac{\mu}{2L^2}$

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

линейная сходимость!

Следствие хотим $\|x^{k+1} - x^*\|^2 \leq \varepsilon$

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

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

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

$$\exp\left(\frac{\mu^2}{4L^2} \cdot (k+1)\right) \geq \frac{\|x^0 - x^*\|^2}{\varepsilon} \qquad \Big|\log$$

$$(k+1) \cdot \frac{\mu^2}{4L^2} \geq \log \frac{\|x^0 - x^*\|^2}{\varepsilon}$$

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

плохая!

надо улучшить

оценка на число итер

оценка на оракульн. сложн.

Св-ва L-гладкой выпуклой функции:

Уже доказано: $\forall\, x, y \in \mathbb{R}^d$

$$\frac{\mu}{2}\|x-y\|^2 + 0 \underset{\text{выпуклость}}{\leq} f(y) - f(x) - \langle \nabla f(x); y - x \rangle \overset{\text{гладкость}}{\leq} \frac{L}{2}\|y - x\|^2$$

новое: $\forall x, y \in \mathbb{R}^d$

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

Док-во: из старого $\Rightarrow$ новое

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

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

- $\varphi - L_\varphi$ - гладкая? не опред.

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

$$= \|\nabla f(y_1) - \nabla f(y_2)\| \overset{L\text{-гладкость } f}{\le} L \|y_1 - y_2\|$$

$$\boxed{L_\varphi = L}$$

- φ - выпуклая? как сумма 2х выпуклых (не опред.)

$\boxed{\text{выпуклая}}$

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

$\boxed{y^* = x}$ (возможно, не ед.)

С помощью выше

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

$$\varphi(x) = \min_y \varphi(y)$$

L-гладкость φ

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

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

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

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

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

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

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

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

■

Док-во сходимости град. спуска: $(\gamma_k \equiv \gamma)$

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

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

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 \underbrace{- 2\gamma \langle \nabla f(x^k); x^k - x^* \rangle}_{\mu\text{-сильн. вып.}} + \gamma^2 \|\nabla f(x^k)\|^2$$

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

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

$$\|\nabla f(x^k)\|^2 = \|\nabla f(x^k) - \underset{0}{\underset{\|}{\nabla f(x^*)}}\|^2$$

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

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

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

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

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

<u>Сходимость</u> (оптимальн.) град. спуска для L-гладких, μ-сильно вып. ф. $\gamma \leq \frac{1}{L}$

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

NB $\gamma \in (0; \frac{2}{L})$

Выше:

$$\gamma \sim \frac{\mu}{L^2}$$

Здесь:

$$\gamma \sim \frac{1}{L} \qquad \mu \ll L$$

Подставим $\gamma \equiv \frac{1}{L}$

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

рекуррентно

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

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

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

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

← Было $k \sim \frac{L^2}{\mu^2}$

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

O „съедает" „все" численные факторы

не улучш. для град. спуска

(не можно лучше с пом. др. метода!)

$\varepsilon \to 0 \qquad \frac{L}{\mu} \to \infty$

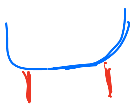
Сходимость град. спуска для других классов задач:

- L-гладкая, выпуклая

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

сублин. сх.

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



(сход. по аргум. нет)

(можно лучше др. методом)

- L-гладкая, невыпуклая

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

$\|\nabla f(x^k)\| \leq \varepsilon$

(только к стац. точке)

град. спуск оптимален

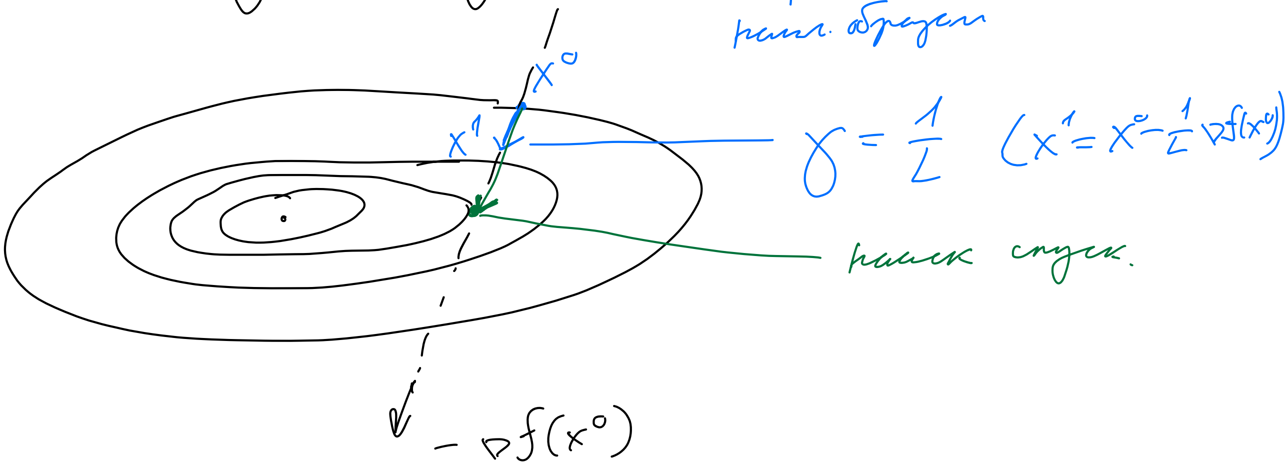
NB f^* может не сущ., нужно пред. $f^* > -\infty$

Способы подбора шага:

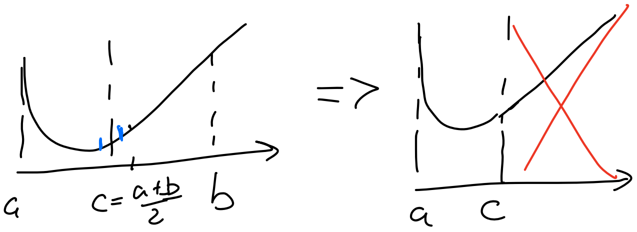
- const, пример $\gamma = \frac{1}{L}$
- наискорейший спуск

$$\gamma_k = \arg\min_{\gamma \in \mathbb{R}} \left(f(x^k - \gamma \nabla f(x^k)) \right)$$

наил. образом



как решить $\arg\min_{\gamma}$?
- явная формула
- выпукла по γ, если f выпукла
можно МЗС или бинпоиск.



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

- $\gamma_k = \frac{1}{k+1}$; $\frac{1}{\sqrt{k+1}}$ по номеру итер.
- Поляка - Шора

из док-ва сходимости:

$\mu = 0$

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

$$\underbrace{+\ \gamma_k^2 \|\nabla f(x^k)\|^2}_{\min \text{ по } \gamma_k}$$

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

— проблема

↓

- знать $f(x^*)$
- можно оценить
- добавить параметр

$\alpha > 1$

- Адаптивный подбор: L_k можно менять

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

- Armijo
  Wolfe
  Goldstein } Д/З