Задача безусловной оптимизации:

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

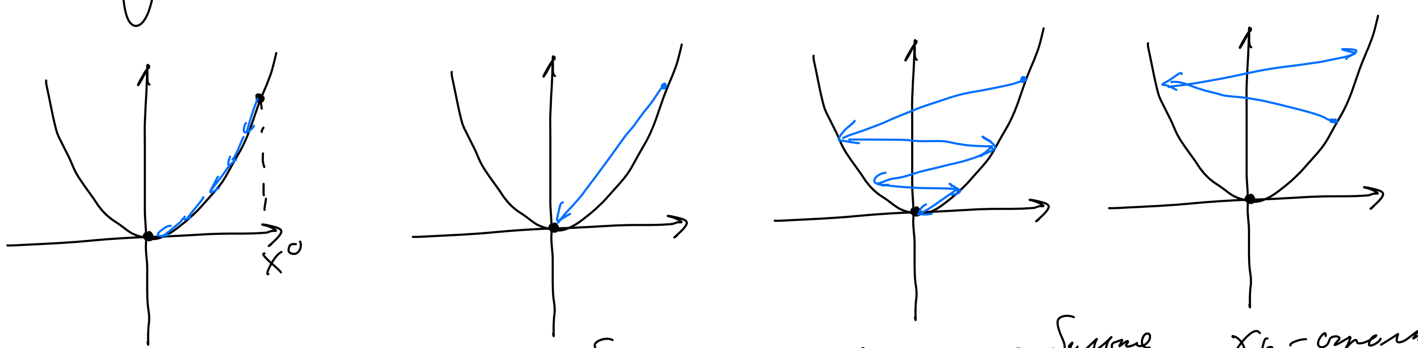
Метод град. спуска:

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

NB: x^k_i k ← номер итер., i ← номер коорд.

Физика: ∇f — направл. роста $\Rightarrow$ $-\nabla f$ — направл. уб-я

γ_k : $\min\limits_{x\in\mathbb{R}} \frac{1}{2}x^2$ $x^0 = 1$ $x^{k+1} = x^k - \gamma_k x^k$



γ_k — маленькие
+ — сойдется
— — медленно

γ_k — больше
+ — быстрая сходимость

γ_k — еще больше
+ — сход.
— — медленно

γ_k — огромный
— — расход.

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

- f — L-гладкая
- f — μ-сильно выпуклой

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

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

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

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

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

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

сильная вып.

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

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

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

оптимально: $1 - \gamma\mu + \gamma^2 L^2 \to \min_{\gamma > 0}$ $\quad \gamma = \frac{\mu}{2L^2}$

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

рекурсия

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

$$\le \ldots$$

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

<u>Сходимость</u> (плохая!) град. спуска для L-гладких μ-сильно выпуклых задач $\gamma = \frac{\mu}{2L^2}$

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

линейная сходимость

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

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

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

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

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

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

$$\boxed{k \ge 1+\frac{4L^2}{\mu^2}\log\frac{\|x^0-x^*\|^2}{\varepsilon}}$$

оценка на число итераций

оценка на оракул. сложность

Будем улучшать!

Св-во L-гладкой выпуклой функции:

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

NB $\frac{\mu}{2}\|x-y\|^2$ с.в.

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

Новое: $\forall x,y\in\mathbb{R}^d$

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

Док-во: из старого сделаем новое

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

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

- φ − L_φ-гладкая? по опред.

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

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

L − гладк. f

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

- φ − выпукла? как сумма выпуклых (можно по определению) $\boxed{\text{выпуклая}}$

- $\nabla \varphi(y^*) = 0 \Rightarrow \nabla f(y^*) - \nabla f(x) = 0$ $\boxed{y^* = x}$ (возможно, не ед.)

С помощью свойств выше:

$$\varphi(x) = \varphi^*$$

x − оптимум

$\min\limits_y \varphi(y)$

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

L − гладкость φ

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

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

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

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

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

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

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

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

Док-во сход. град. спуска: $\gamma_k \equiv \gamma$

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

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

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

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

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

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

0

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

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

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

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

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

Выше: $\gamma \sim \frac{\mu}{L^2}$

Здесь $\gamma \sim \frac{1}{L}$ лучше! $\mu \ll L$

Подставим $\gamma \equiv \frac{1}{L}$

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

рекурсия

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

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

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

Было $k \sim \frac{L^2}{\mu^2}\log\frac{1}{\varepsilon}$ хуже!

$$\boxed{k \ge 1 + \frac{L}{\mu}\log\frac{\|x^0-x^*\|^2}{\varepsilon}}$$

← Не улучш. для GD

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

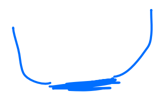
O скрывает все числен. факторы

Сходимость GD для других классов задач

- L-гладкие, выпуклые ф.

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

$f(x^k)-f^*\le\varepsilon$ (сходимость по аргументу нет)



сублинейная сход.

- L-гладкие, невыпуклые

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

$\|\nabla f(x^k)\|\le\varepsilon$ (только к стац. точке)

NB f^* — может не сущ. нужно предп. $f^*>-\infty$