$$\boxed{\begin{array}{l}\min\limits_{x\in Q} f(x)\\ \|\nabla f(x)\|_2 \le M\\ \|x^0 - x_*\|_2 \le R\end{array}}$$

$$f(\tilde{x}^N) - f(x_*) \le \frac{MR}{\sqrt{N}} = \varepsilon$$

$$\tilde{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k$$

$$N = \frac{M^2R^2}{\varepsilon^2}$$

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

$$h = \frac{R}{M\sqrt{N}} = \frac{\varepsilon}{M^2}$$

$$f(\tilde{x}^N) - f(x_*) \ge \frac{MR}{\sqrt{N+1}}$$

Лекция 8

Ускоренный градиентный спуск

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

$$\|x^0 - x_*\|_2 \le R$$

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

$$f(x) = x^2 \qquad L_f = ?$$

$$\nabla f(x) = 2x$$

$$\| 2y - 2x \|_2 \le L \, \|y - x\|_2$$

$$L = 2$$

$$L = \max_{x \in \mathbb{R}^d} \lambda_{max} (\nabla^2 f(x))$$

$$\nabla^2 f(x) = \left(\frac{\partial f}{\partial x_i \partial x_j} \right)_{i,j=1,d}$$

$$\left\{ \begin{array}{l} \nabla^2 f(x) \preceq L I \\ \langle y, \nabla^2 f(x) y \rangle \le \\ \le L \langle y, y \rangle \end{array} \right.$$

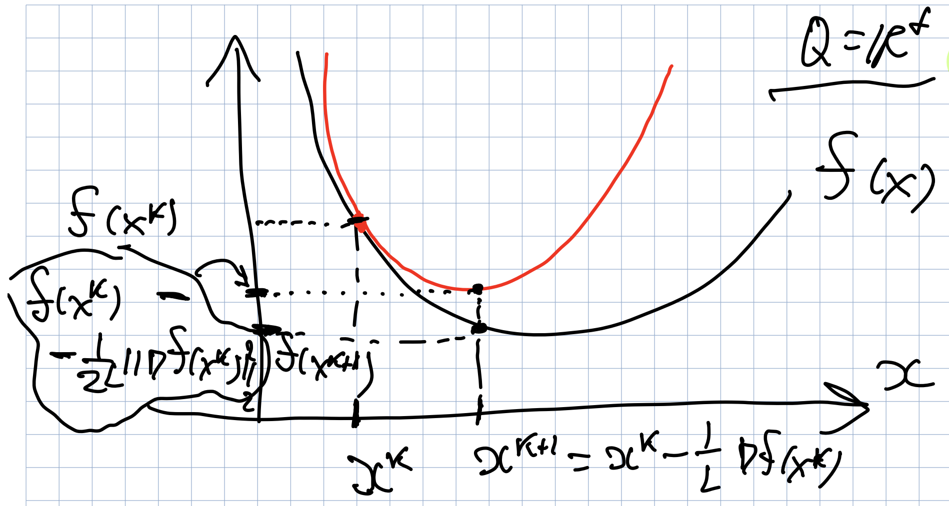
$$x^{k+1} = x^k - h \nabla f(x^k) \quad (0)$$

$$\underbrace{\frac{1}{N} \sum_{k=0}^{N-1} f(x^k)}_{\ge \frac{1}{N} f(\bar{x}^N)} - f(x_*) \le \frac{1}{2hN} \overset{\|x^0 - x_*\|_2^2}{R^2} - \cancel{\frac{1}{2h} \|x^N - x_*\|_2^2} + \frac{hM^2}{2}$$

$$\|\nabla f(x^k)\|_2 \le M$$

$$(0) \quad x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^d} \left\{ \underline{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle} + \underline{\frac{1}{2h} \|x - x^k\|_2^2} \right\}$$

$$h \le \frac{1}{L} \qquad \mathcal{U}_\delta(L) \Rightarrow f(y) \le f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|_2^2$$



$Q = \mathbb{R}^d$ $x^{k+1} = x^k - \frac{1}{L}\nabla f(x^k)$

град. спуск.

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{2L}\|\nabla f(x^k)\|_2^2$$

$$\wedge|$$

$$f(x^k) - f(x_*)$$

$$(\star) \quad \|\nabla f(x^k)\|_2^2 \leq 2L\left(f(x^k) - f(x_*)\right)$$

$$\frac{1}{N}\sum_{k=0}^{N-1} f(x^k) - f(x_*) \leq \frac{R^2}{2hN} + \frac{h}{2}\underbrace{\frac{1}{N}\sum_{k=0}^{N-1}\|\nabla f(x^k)\|_2^2}_{\leq M^2} \leq$$

грубо !!!

$$\overset{\star}{\leq} \frac{R^2}{2hN} + hL\left(\frac{1}{N}\sum_{k=0}^{N-1} f(x^k) - f(x_*)\right)$$

$hL \to 1/2$

~~$h = 1/L$~~ $h = 1/2L$

$$\frac{1}{2}\left(\frac{1}{N}\sum_{k=0}^{N-1} f(x^k) - f(x_*)\right) \leq \frac{LR^2}{N}$$

V) Н-во Иенсена

$$\frac{1}{2}\left(f(\bar{x}^N) - f(x_*)\right) \le \frac{LR^2}{N}$$

$$f(\bar{x}^N) - f(x_*) \le \frac{2LR^2}{N}$$

Напомним, что до этого было вот так

Это лучше

$$f(\bar{x}^N) - f(x_*) \le \frac{MR}{\sqrt{N}}$$

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

Не оптимальный метод для класса гладких задач

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

А.С. Немировский, 1979 г.

$$f(\tilde{x}^N) - f(x_*) \ge \frac{LR^2}{32N^2}$$

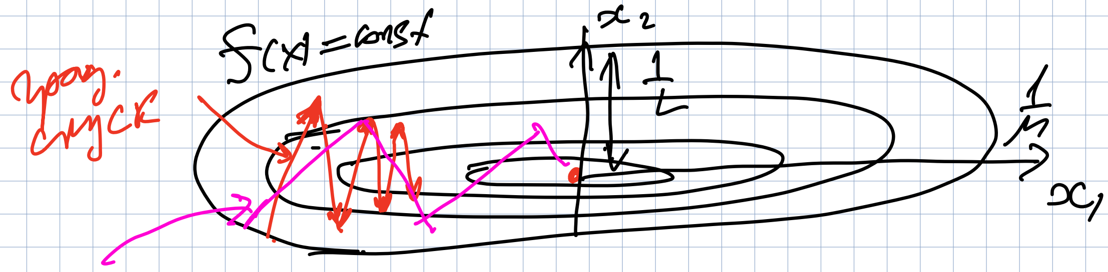
1983 г. Ю.Е. Нестеров

$$\int x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{k-1}{k+2}(x^k - x^{k-1})\right) + \frac{k-1}{k+2}(x^k - x^{k-1})$$

L

$$x^1 = x^0 - \frac{1}{L}\nabla f(x^0)$$

$$f(x^N) - f(x_*) \le \frac{4LR^2}{N^2}$$



$$f(x) = \frac{\mu}{2}x_1^2 + \frac{L}{2}x_2^2$$

$$\mu \ll L$$

$$\mu I \preceq \nabla^2 f(x) \preceq LI$$

конст. сильн. выпукл.

конст. Липш. град.

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

$$f(x^N) - f(x_*) \le LR^2 \min\left\{\frac{1}{N}, \underbrace{\exp\left(-\frac{\mu}{L}N\right)}\right\}$$

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

$$N = \frac{L}{\mu}\ln\left(\frac{LR^2}{\varepsilon}\right)$$

$$\begin{cases} x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1})\right) + \\ \qquad + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1}) \\ x^1 = x^0 - \frac{1}{L}\nabla f(x^0) \end{cases}$$

$$f(x^N) - f(x_*) \leq \varepsilon$$

$$N \simeq \sqrt{\frac{L}{\mu}} \ln\left(\frac{LR^2}{\varepsilon}\right)$$

$$f(x^N) - f(x_*) \leq LR^2 \left\{ \frac{4}{N^2}, \exp\left(-\sqrt{\frac{\mu}{4L}}N\right) \right\}$$

Метод сопряженных градиентов

$$f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle \to \min_{x \in \mathbb{R}^d}$$

$$A \succeq 0$$
$$\langle x, Ax\rangle \geq 0 \quad \forall x$$

$$\nabla f(x) = 0 \iff \underbrace{Ax = b}$$

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

$$(\alpha_k, \beta_k) = \underset{(\alpha,\beta)}{\operatorname{argmin}}\, f\left(x^k - \alpha\nabla f(x^k) + \beta(x^k - x^{k-1})\right)$$

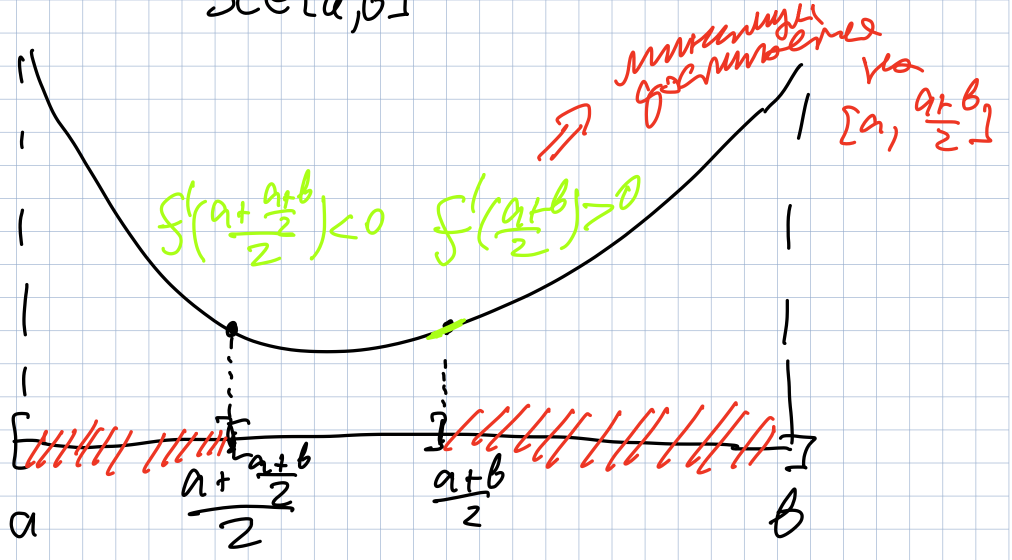
$$x^{k+1} = \underset{x \in x^0 + \operatorname{Lin}\{\nabla f(x^0), \nabla f(x^1), \ldots, \nabla f(x^k)\}}{\operatorname{argmin}} f(x)$$

$$f(x^N) - f(x_*) \le$$

$$\le LR^2 \min\left\{\frac{1}{2(2N+1)^2},\ 2\exp\left(-2\sqrt{\frac{\mu}{L}}N\right)\right\}$$

Одномерная оптимизация

$$\min_{x\in[a,b]} f(x)$$



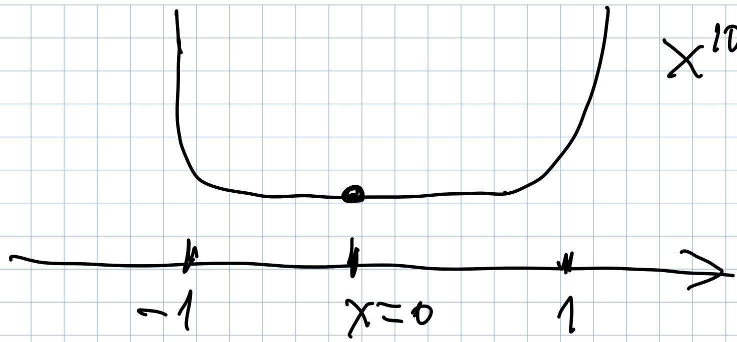
ускор. град. спуск

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

$$N \sim \frac{1}{\sqrt{\varepsilon}}$$ → точность по ф-ции

Метод деления отрезка пополам:

$$N \sim \log_2 \frac{(b-a)}{\varepsilon}$$ → точность по аргументу



$x^{1000} \to \min_{x}$

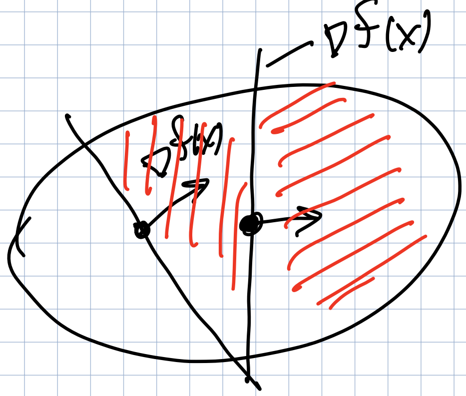
$$\min_{x \in Q \subseteq \mathbb{R}^d} f(x)$$

$\|\nabla f(x)\|_2 \le M$

$\operatorname{diam} Q \le R$

$$N = \tilde{O}\left(d \ln \frac{MR}{\varepsilon}\right)$$

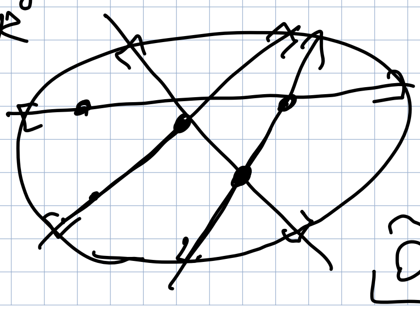
Методы типа центров тяжести



объём $\ge \frac{1}{e}$ убираем.

$$\left(1 - \frac{1}{e}\right)^N$$

$\mathbb{R}^d$



Hit and Run

2003

Lovaz & Vampala

$\tilde{O}(d^5) \to \tilde{O}(d^3)$ метод Вайды

$\nabla f(x)$

Ax

$O(d^2)$

стоимость итер. град методов $O(d^2)$

$d^{2.37}$

$d^{2.7}$

$O(d^3)$

$$N = \min\left\{\sqrt{\frac{LR^2}{\varepsilon}},\ \sqrt{\frac{L}{\mu}}\ln\frac{LR^2}{\varepsilon}\right\}$$

град. спуски

$$N = d\ln\frac{MR}{\varepsilon}$$

методы типа центр. тяжести

Регуляризация (А.Н. Тихонов)

$$f_\mu(x) := f(x) + \frac{\mu}{2}\|x - x^0\|_2^2$$

есть алгоритм, который умеет решать сильн. вып. задачи

$$N = \sqrt{\frac{L}{\mu}}\ln\left[\frac{\mu R^2}{\varepsilon}\right] \xrightarrow[\mu \sim \varepsilon/R^2]{} N = \sqrt{\frac{LR^2}{\varepsilon}}$$

$$\min f(x) \qquad x_*$$
$$\min f_\mu(x) \qquad x_*^\mu$$

Лемма. Если $R^2 = \|x^0 - x_*\|_2^2$, и $\mu \le \varepsilon/R^2$,

$$f_\mu(x^N) - f_\mu(x_*^\mu) \le \varepsilon/2,$$ то

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

Резюме: Достаточно решить задачу

$$\min_{x \in Q} \; f(x) + \frac{\mu}{2}\|x - x^0\|_2^2, \quad \mu = \varepsilon/R^2$$

нужно решить эту задачу с точностью $\varepsilon/2$

$\Downarrow$

Это решение будет ε-реш. задачи

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

Рестарты (А.С. Немировский, 1979)

$\min_{x \in Q} f(x) \to f(x)$ – μ-сильно-вып

$$\frac{\mu}{2}\|x^{\bar N}-x_*\|_2^2 \le f(x^{\bar N}) - f(x_*) \le \frac{4L\|x^0-x_*\|_2^2}{\bar N^2}$$

есть алгоритм, который не учитывает сильн. вып.

$$\|x^{\bar N}-x_*\|_2^2 \le \underbrace{\frac{8L}{\mu \bar N^2}}_{1/2}\|x^0-x_*\|_2^2$$

$$\bar N = \sqrt{8\frac{L}{\mu}}$$

Идея: рестартовать метод.

$$x^0 := x^{\bar N}$$

$$\|x^{\bar N}-x_*\|_2^2 = \frac{1}{2}\|x^0-x_*\|_2^2$$

$\bar N = \sqrt{8\frac{L}{\mu}}$ - длина рестарта

длина рестарта

$$N \simeq \sqrt{\frac{L}{\mu}}\,\underbrace{\ln\left(\frac{\mu R^2}{\varepsilon}\right)}$$

число рестартов

найду ε-решение

$$Ax = b$$

$$\Downarrow$$

$$f(x) = \frac{1}{2}\|Ax-b\|_2^2 \to \min_x$$

$$\nabla f(x) = A^T(Ax-b)$$