Оракульная сложность численных методов (1-го порядка) выпуклой оптимизации

$$f(x) \to \min_{x \in Q} \quad (*)$$

выпуклая; $Q \subseteq \mathbb{R}^n$ ← выпуклое

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

x^N — ε-решение задачи $(*)$

N — число итераций (число вызовов оракула) $\# \nabla f(x)$

$$x^{k+1} = x^0 + \underbrace{Lin\{\nabla f(x^0), \dots, \nabla f(x^k)\}} \quad // \quad x^{k+1} = x^k - h \nabla f(x^k)$$

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

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

① $N \ge n$ $\quad N \simeq n \ln\left(\frac{\Delta f}{\varepsilon}\right)$ ← нижняя оценка; ← достигается на Методе центров тяжести.

Методы типа деления отрезка пополам ← Вайды (сложность итерации $\tilde{O}(n^{\log_2 7})$)

② $N \le n$ $\quad$ в 2-норме

$R^2 \ge \|x^0 - x_*\|_2^2$

$N(\varepsilon)$	$\lvert f(y)-f(x)\rvert \le M\|y-x\|_2$	$\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y-x\|_2$
$f(x)$ - выпуклая	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}}$
$f(x)$ - μ- сильно выпуклая	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil \; (\forall N)$

рестарты; $\mu \sim \varepsilon/R^2$

* $R^2 = V(x^0, x_*)$, $\lvert f(y)-f(x)\rvert \le M\|y-x\|$, $\|\nabla f(y) - \nabla f(x)\|_* \le L\|y-x\|$

$$f(x^N) - f(x_*) \le \frac{MR}{\sqrt{N}} = \varepsilon \Rightarrow \varepsilon \Big\} \frac{MR}{\sqrt{N+1}}$$

нижняя оценка

Субградиентный метод, $Q = \mathbb{R}^n$

$$x^{k+1} = x^k - \frac{\varepsilon}{M^2}\cdot\nabla f(x^k), \quad k \le N$$

$$x^{k+1} = \Pi_Q\left(x^k - \frac{\varepsilon}{M^2}\nabla f(x^k)\right)$$

Поляк–Шор 60-ые

Ускорен. методы

Нестеров '83

$Q = \mathbb{R}^n$

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

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{k-1}{k+2}(x^k - x^{k-1})\right) + \underbrace{\frac{k-1}{k+2}(x^k - x^{k-1})}_{\text{моментный член}}, \quad k \geq 2$$

$$f(x^N) - f(x_*) \leq \frac{4LR^2}{N^2} = \varepsilon \simeq \frac{LR^2}{N^2}$$

нижняя оценка

сильно вып. случай

ускор. метод

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

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\left(x^k + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1})\right) + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x^k - x^{k-1}), \quad k \geq 2$$

$$f(x^N) - f(x_*) \leq LR^2 \exp\left(-\frac{1}{2}\sqrt{\frac{\mu}{L}}N\right)$$

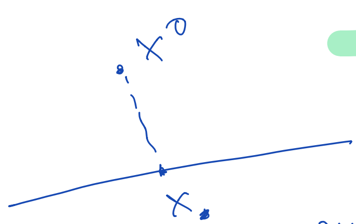


Схема рестартов

$$R^2 = \|x^0 - x_*\|_2^2$$

оценка для вып. задачи $\Longrightarrow$ оценка для μ-сильно вып. задачи

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

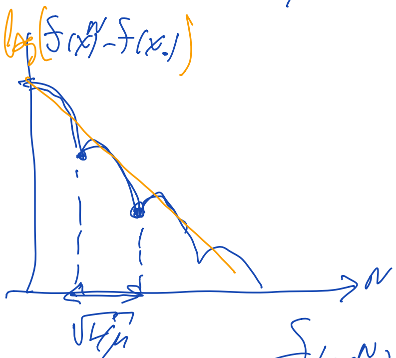
$$f(x) \geq f(x_*) + \underbrace{\langle \nabla f(x_*), x - x_*\rangle}_{\geq 0} + \frac{\mu}{2}\|x - x_*\|_2^2 \geq f(x_*) + \frac{\mu}{2}\|x - x_*\|_2^2$$

Выберем в $(**)$ $N = \bar{N}$: $(***)$ $\|x^{\bar{N}} - x_*\|_2^2 \leq \frac{1}{2}\|x^0 - x_*\|_2^2$

Из (**) $\Rightarrow$ $\dfrac{8L}{\mu \bar{N}^2} = \dfrac{1}{2} \Rightarrow \bar{N} = 4\sqrt{\dfrac{L}{\mu}}$

Число рестартов $r = \left\lceil \log_2\left(\dfrac{\mu R^2}{\varepsilon}\right)\right\rceil$, то $f(\tilde{x}^r) - f(x_*) \le \varepsilon$

$$\# \nabla f(x) \longrightarrow 4\sqrt{\frac{L}{\mu}}\left\lceil \log_2\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil$$



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

$\downarrow$

при рестартах не меняется

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

Регуляризация

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

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

μ-сильно выпуклая задача

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

$$\boxed{f_\mu(x^N) - \min_{x \in Q} f_\mu(x) \le \varepsilon/2 \qquad \mu = \varepsilon/R^2}$$

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

$$\Downarrow$$

$$\boxed{f(x^N) - \min_{x \in Q} f(x) = f(x^N) - f(x_*) \le \varepsilon}$$

Поскольку R – не известно, то делаем рестарты по R: $R_0 \to x^N$ ($\|x^N - x^0\|_2 \le R_0 \to$ ok) $\to R_1 = 2R_0 \to x^N (\|x^N - x^0\|_2 \le R_1)$...

Уход от евклидовой нормы

Пример. $f(x) \to \min\limits_{x \in S_n(1)}$ ← симплекс

$\|\cdot\| = \|\cdot\|_p \quad 1 \le p \le 2$ $\quad \cancel{p > 2}$

$f(x)$ – $\underbrace{\text{1-сильно-вып.}}$ $\underbrace{\text{отн. } \|\cdot\|}$ не нужно! Кроме $p=2$

прокс. ф-ция → $d(x)$

$\|\cdot\|_2 \to d(x) = \frac{1}{2}\|x\|_2^2$

$\|\cdot\|_1 \to \underbrace{\sum_{i=1}^{n} x_i \ln x_i}_{d(x)}, \; S_n(1)$

$$V(y,x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \ge \frac{1}{2}\|y - x\|^2$$

$$x^{k+1} = \operatorname*{argmin}_{x \in Q} \{ \underbrace{f(x^k)}_{\text{можно не писать}} + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2}\|x - x^k\|_2^2 \}$$

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

$$= \operatorname*{argmin}_{x \in Q} \{ \underline{h}\langle \nabla f(x^k), x - x^k \rangle + V(x, x^k) \}$$

$\frac{L}{2}\|x - x^k\|_2^2 \to L\, V(x, x^k)$

$$\underset{Q = \mathbb{R}^n}{=} x^k - \underline{h}\nabla f(x^k)$$

$d(x) = \frac{1}{2}\|x\|_2^2$

$V(y,x) = \frac{1}{2}\|y - x\|_2^2$

$R^2 = V(x^0, x_*)$

$|f(y) - f(x)| \le M\|y - x\|$

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

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

$f(x) - f(x_*) \ge \frac{\mu}{2}\|x - x_*\|^2$

Рестарты

$$\frac{1}{2}\|x^{\bar N}-x_*\|_2^2 \le \frac{4L\,V(x^0,x_*)}{N^2} \to \frac{4L\varkappa\|x^0-x_*\|_2^2}{N^2}$$

$$\underbrace{V(y,x_*) \le \varkappa\|y-x_*\|^2}$$

μ-сильно вып. в 1-норме на

Пример.

$$F(x) := \underbrace{\frac{1}{2}\|Ax-b\|_2^2}_{f(x)} + \underbrace{\mu\sum_{i=1}^n x_i\ln x_i}_{h(x)} \to \min_{x\in S_n(1)}$$

$\underbrace{Ax=b}$ → много решений

$$10^5\ \underbrace{\boxed{A}}_{10^8}\ \boxed{x} = \boxed{b}$$

Композитная

$$x^{k+1} = \arg\min_{x\in S_n(1)}\left\{ \cancel{\langle\nabla F(x^k), x-x^k\rangle} + \cancel{L_F}\,\underbrace{V(x,x^k)}\right\}$$

композитная оптимизация

$$\langle\nabla f(x^k), x-x^k\rangle + h(x) + L_f\,\underbrace{V(x,x^k)}$$

$$\tilde O(n) \to \quad \frac{1}{2}\underbrace{A^T(\underbrace{Ax^k}-b)} + \underbrace{\mu\sum_{i=1}^n x_i\ln x_i}$$

Самое дорогое [?]

$O(nnz(A))$

$$\max_k \|A^{\langle k\rangle}\|_2^2$$

$$F(x^N)-F(x_*) \le \frac{4L_f R^2}{N^2}, \quad R^2 \to \ln n$$

$$(x\ln x)' = 1+\ln x, \quad x\to 0+, \quad \text{то } (x\ln x)'\to\infty$$

Стох. оптимизация

x_* – реш. $\to$ $f(x) := \mathbb{E}_\xi [f(x,\xi)] \to \min\limits_{x\in Q}$

Оракул: $\nabla_x f(x,\xi)$

1) $\mathbb{E}_\xi [\nabla_x f(x,\xi)] \equiv_x \nabla f(x)$

2) $\mathbb{E}[\|\nabla_x f(x,\xi) - \nabla f(x)\|_2^2] \leq \sigma^2$

или

2)' $\mathbb{E}[\|\nabla_x f(x,\xi)\|_2^2] \leq M^2$

$\boxed{\mathbb{E}[f(x^N)] - f(x_*) \leq \varepsilon \quad R^2 = \|x^0 - x_0\|_2^2}$

3) $\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y-x\|_2$

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

$|f(y) - f(x)| \leq M\|y-x\|_2$

$N(\varepsilon)$	1)+2)' $\mathbb{E}[\|\nabla_x f(x,\xi)\|_2^2] \leq M^2$	1), 2), 3)
$f(x)$ – вып.	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}} + \frac{\sigma^2R^2}{\varepsilon^2}$
$f(x)$ – μ-сильно вып	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}}\lceil \ln\frac{\mu R^2}{\varepsilon}\rceil + \frac{\sigma^2}{\mu\varepsilon}$

отличие от детерм. случая

$\nabla_x f(x,\xi) \to \frac{1}{r}\sum_{j=1}^{r} \nabla_x f(x,\xi^j)$

подставлять в детер. метод.

$\sigma^2 \to \sigma^2/r$

Гладкость $\Leftrightarrow$ не гладкость

(A) $f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|_2^2$ $\leftarrow$ работаем тут с этим

$$\Downarrow$$

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

(II) в негладком случае ($\|\nabla f(x)\|_2 \le M$

$$\Uparrow$$

$|f(y)-f(x)| \le M\|y-x\|_2$)

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

$$f(x^N) - f(x_*) \le \underbrace{\frac{LR^2}{2N}} + \underbrace{\delta}$$ град. спуск.

$$f(x^N) - f(x_*) \le \frac{4LR^2}{\underbrace{N^2}} + \underbrace{N\cdot\delta}$$ ускор. метод.

Град. спуск.

$$L = \frac{M^2}{\delta}$$

$$\frac{M^2R^2}{2\delta N} \sim \frac{\varepsilon}{2}$$

$$\underbrace{\delta \sim \frac{\varepsilon}{2}}$$

$$\Rightarrow \boxed{N \approx \frac{M^2R^2}{\varepsilon^2}}$$

Ускор. град. метод

$$\frac{M^2R^2}{\delta N^2} \sim \frac{\varepsilon}{2}$$

$$N\cdot\delta \sim \frac{\varepsilon}{2}$$

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

$$\Rightarrow \boxed{\delta \sim \varepsilon^{3/2}}$$