Метод условного градиента

$$f(x) \to \min_{x \in Q} \quad \Big\} \quad x_* \text{ — решение задачи}$$

$$x^{k+1} = \arg\min_{x\in Q} \Big\{ \underbrace{\langle \nabla f(x^k), x - x^k \rangle}_{} + \frac{L}{2}\|x - x^k\|_2^2 \Big\}$$

h

Франк-Вульф (1955)

$$\langle \nabla f(x^k), y^k - x^k\rangle \le \langle \nabla f(x^k), x_* - x^k\rangle \Leftarrow \boxed{y^k \in \operatorname{Argmin}_{x\in Q} \langle \nabla f(x^k), x - x^k\rangle}$$

этого нет!

(2)

$$\boxed{x^{k+1} = (1-\gamma_k)x^k + \gamma_k y^k} \Rightarrow x^{k+1} - x^k = \gamma_k (y^k - x^k)$$

(1)

Теорема. Введём $\|\cdot\|$, $R = \sup_{x,y\in Q} \|x-y\|$

$$L = \max_{\substack{\|h\|\le 1 \\ x\in Q}} \langle h, \nabla^2 f(x) h\rangle$$

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

Пусть $\gamma_k = \frac{2}{k+1}$.

Тогда $f(x^N) - f(x_*) \le \frac{2LR^2}{N+1}$.

сходимость аналогич. град. спуск.

ускор. град. спуск $\sim \frac{LR^2}{N^2}$

Д-во

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

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

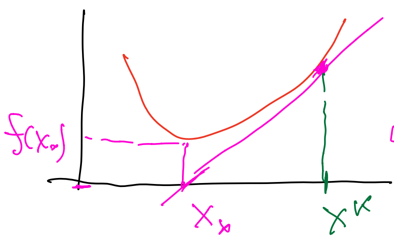
$$\Downarrow$$

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

$$\underset{(1)}{\le} \gamma_k \langle \nabla f(x^k), y^k - x^k \rangle + \frac{L}{2}\gamma_k^2 \underbrace{\|y^k - x^k\|^2}_{\le R^2} \underset{(2)}{\le}$$

$$x^{k+1} - x^k = \gamma_k (y^k - x^k)$$

$$\underset{(2)}{\le} \gamma_k \langle \nabla f(x^k), x_* - x^k \rangle + \frac{L}{2}\gamma_k^2 R^2 \le$$



$$\underbrace{f(x_*) \ge f(x^k) + \langle \nabla f(x^k), x_* - x^k \rangle}_{\text{по выпуклости } f}$$

$$\le \gamma_k \left(f(x_*) - f(x^k)\right) + \frac{L}{2}\gamma_k^2 R^2$$

$$\delta_k = f(x^k) - f(x_*)$$

$$f(x^{k+1}) - f(x^k) \le \gamma_k \left(f(x_*) - f(x^k)\right) + \frac{L}{2}\gamma_k^2 R^2$$

$$f(x^{k+1}) - f(x_*) \le f(x^k) - f(x_*) - \gamma_k\left(f(x^k) - f(x_*)\right) + \frac{L}{2}\gamma_k^2 R^2$$

$$\delta_{k+1} \le (1-\gamma_k)\delta_k + \frac{L}{2}\gamma_k^2 R^2$$

$\gamma_k = \frac{2}{k+1}$, $\delta_2 \le \frac{L}{2}R^2$ $\Rightarrow$ $\delta_k \le \frac{2LR^2}{k+1}$ ч.т.д.

(M. Jaggi '2013)

Пример. (упражнение 1.6 пособия МЦНМО)

$n \times n$ матрица, s ненулевых эл-тов с каждого столбца

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \underbrace{\langle b, x\rangle}_{0 \text{ для простоты}} \to \min_{x \in S_n(1)}$$

$S_n(1) = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ — вероятностный симплекс

Быстрый градиентный метод.

$d(x) = \sum_{i=1}^n x_i \ln x_i$ — прокс-функция

$d(x)$ — 1-сильно вып. отн. 1-нор на $S_n(1)$, $\|\cdot\| = \|\cdot\|_1$

$V(y,x) = \sum_{i=1}^n y_i \ln(y_i/x_i)$ — дивергенция Брэгмана

(не ускорен.) $x^{k+1} = \arg\min\limits_{x \in S_n(1)} \{\langle \nabla f(x^k), x - x^k\rangle + L\, V(x, x^k)\}$

$\langle \nabla f(x^k), x - x^k\rangle$ считается за $O(sn)$; решается за $O(n)$

$\max\limits_{x \in S_n(1)} V(x_*, x^0) \le \ln n = R^2$, $x^0 = (1/n, \ldots, 1/n)$

$L = \max\limits_{\substack{\|h\|_1 \le 1 \\ x \in S_n(1)}} \langle h, \nabla^2 f(x) h\rangle = \max\limits_{\|h\|_1 \le 1} \langle h, Ah\rangle$, $\nabla^2 f(x) = A$

$\|\nabla f(y) - \nabla f(x)\|_\infty \le L\|y - x\|_1$

$L = \max\limits_{i,j=1,\ldots,n} |A_{ij}|$

Для ускор. варианта (см. пособие Воронцова и др. конец главы 2)

$$f(x^N) - f(x_*) \le \frac{4LR^2}{N^2} = \varepsilon \quad \Rightarrow \quad N \sim \sqrt{\frac{LR^2}{\varepsilon}}$$

N — число итераций (число вычислений $\nabla f(x^k)$) $\to$ $Ax^k \to O(sn)$ стоимость

стоимость одной итерации: $O(sn)$

Общая сложность

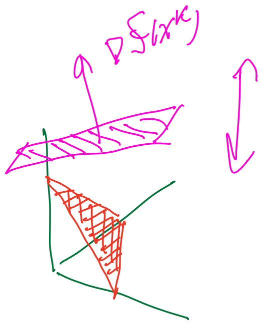
$$\boxed{\underbrace{O(sn)}_{\text{стоимость итерации}} \cdot O\Big(\underbrace{\sqrt{\frac{L \ln n}{\varepsilon}}}_{\text{число итераций}}\Big)}$$

Метод Ф-В для этой же задачи

$$\langle \nabla f(x^k), x - x^k \rangle \to \min_{x \in S_n(1)}$$

$$\langle \nabla f(x^k), x \rangle \to \min_{x \in S_n(1)}$$

$$\Downarrow$$



$$y^k = (\underbrace{0, \dots, 0, 1}_{i_k}, 0, \dots, 0)^T$$

$$i_k \in \operatorname{Argmin}_{i=1,\dots,n} \frac{\partial f(x^k)}{\partial x_i}$$

$$x^{k+1} = \underbrace{(1-\gamma_k)x^k}_{O(n)} + \underbrace{\gamma_k y^k}_{\text{обновление только одной коор.}}, \quad \gamma_k = \frac{2}{k+1} = \frac{2}{k+2}$$

$O(n)$

$\cancel{\tilde{x}^k}$

стоимость итерации $O(s \log_2 n)$

i-й столбец

$$A(x + \gamma e_{i_k}) = \underbrace{Ax}_{\text{известно}} + \gamma(Ae_i)$$

$\left(\vdots\right)$ – s ненул. коор.

Сложность метода:

$$O(n) \text{ – препроцессинг.} + \underbrace{O(s \log_2 n)}_{\text{стоимость итерации}} \cdot \underbrace{O\left(\frac{LR^2}{\varepsilon}\right)}_{\text{число итераций}}, \quad R^2 \le 2$$

$$O\left(n + \frac{L}{\varepsilon} s \log_2 n\right) \quad \text{vs} \quad O\left(sn\sqrt{\frac{L \ln n}{\varepsilon}}\right)$$

Франк-Вульф — ускоренн. метод Нестерова

Нестеров
2013
ЭБ пособие МФТИ

Универсальный градиентный метод

$$f(x) \to \min_{x \in \mathbb{R}^n} \qquad \left\{ \|\nabla f(y) - \nabla f(x)\|_2 \le L\|y-x\|_2 \right.$$

Адаптивный градиентный метод.

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

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

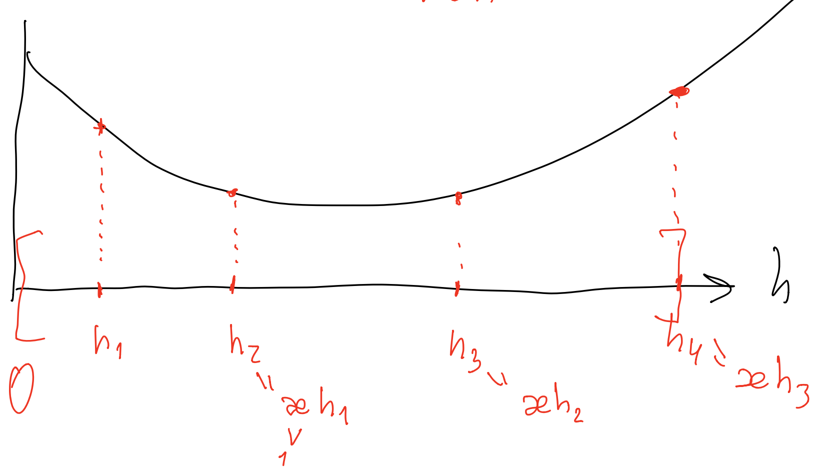
$$L I_n \succeq \nabla^2 f(x) \succeq \mu I_n, \quad \mu \ge 0$$

L - неизвестно

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

$$h_k \in \operatorname*{Argmin}_{h \ge 0} \underbrace{f(x - h\nabla f(x^k))}_{\Phi(h)}$$

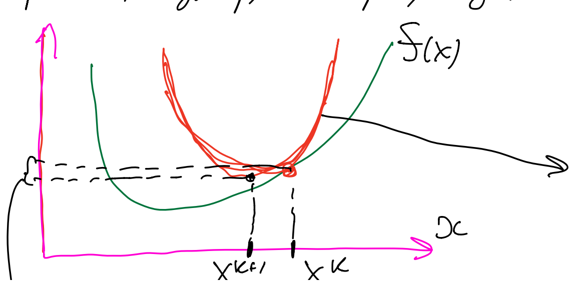
наискорейший спуск.



$$\Phi(h_1) \ge \Phi(h_2) \ge \Phi(h_3) < \Phi(h_4)$$

СТОП!

Армихо, Вульф, Нестеров, Гольдштейн.



$$\underbrace{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2}\|x - x^k\|_2^2}_{\ge f(x)} \quad \forall x \in \mathbb{R}^n$$

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{ f(x^k) + \langle \nabla f(x^k), x - x^k\rangle + \frac{L}{2}\|x - x^k\|_2^2 \right\}$$

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

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

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

$$\Downarrow$$

$$\boxed{f(x) \le f(x^k) + \langle \nabla f(x^k), x - x^k\rangle + \frac{L}{2}\|x - x^k\|_2^2 + \underbrace{\delta}_{=\varepsilon/2}}$$

x^{k+1}

Идея!!!

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

If $\left\{ f(x^{k+1}) \le f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{L}{2}\|x^{k+1} - x^k\|_2^2 + \delta \right\}$, $+\delta = \varepsilon/2$

THEN следующая итерация $L := L/2$

Else $\{ \ldots > \ldots \}$

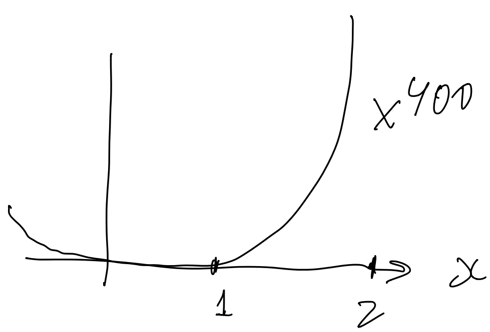
THEN $L := 2L$ go to 1

≤ 3 трех go to в среднем на 1 итер

§5 пособия МЦНМО

Сходится как градиентный метод с

$$L = 2L$$



Негладкая функция

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

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

$\frac{L}{2}$

$$f(x^N) - f(x_*) \le \frac{LR^2}{2N} + \delta \simeq \varepsilon, \quad \frac{LR^2}{2N} \le \frac{\varepsilon}{2}$$

$$L \sim \frac{M^2}{\varepsilon}$$

$$\underbrace{\frac{M^2R^2}{\varepsilon N} + \frac{\varepsilon}{2}}_{\Downarrow}$$

$$\varepsilon$$

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