- Град. спуск

f - L-гладкая, μ-сильно-выпуклая

$O\left(\frac{L}{\mu}\log\frac{\|x^0-x^*\|^2}{\varepsilon}\right)$ итераций / оракул. вызовов,

чтобы получить решение с точн. ε : $\|x^k-x^*\|^2 \le \varepsilon$

А можно лучше? Другой метод?

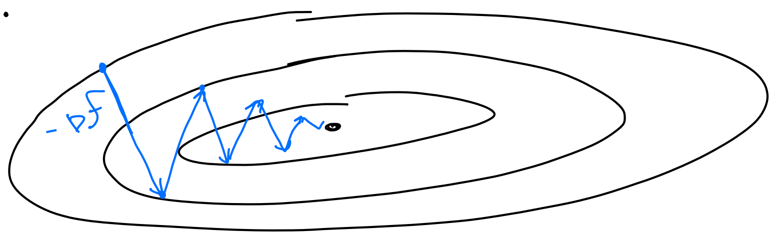
- <u>Метод тяжелого шарика</u> (Heavy Ball / HB)

Б.Т. Поляк 1964 г.

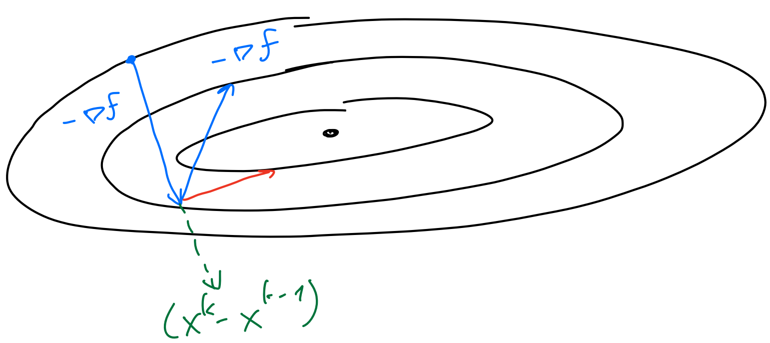
$$x^{k+1} = x^k - \gamma \nabla f(x^k) + \underbrace{\tau(x^k - x^{k-1})}_{\text{инерция / моментум}}$$

<u>Физика метода</u>:

град. спуск:



тяжелый шарик:



Плюсы:

+ на практике лучше, чем Г.С.
+ уменьшение осцилляций
+ легкая имплем.

Особенность:

·

Минусы:

– нужно подбирать 2 параметра: γ и τ ($\tau \in (0{,}9; 1)$)

– в теории не лучше, чем Г.С.

- GD с моментумом (pytorch):

$$v^{k+1} = \beta v^k + \nabla f(x^k) \qquad \beta \in (0;1)$$

$$x^{k+1} = x^k - \gamma v^{k+1}$$

$$\Downarrow$$

$$x^{k+1} = x^k - \gamma \nabla f(x^k) - \gamma \beta v^k$$

$$x^k = x^{k-1} - \gamma v^k \Rightarrow -\gamma v^k = x^k - x^{k-1}$$

$$\Downarrow$$

$$\underline{x^{k+1} = x^k - \gamma \nabla f(x^k) + \beta(x^k - x^{k-1})} \qquad HB$$

Физика метода (ML): HB моментум = запомин. старых град. с уменьш. весом

- Ускоренный градиентный метод (Nesterov)

Ю. Е. Нестеров 1983.

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

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

HB: $x^{k+1} = x^k + \tau(x^k - x^{k-1}) - \gamma \nabla f(x^k)$

Nesterov: $x^{k+1} = x^k + \tau(x^k - x^{k-1}) - \gamma \nabla f(x^k \underbrace{+ \tau(x^k - x^{k-1})}_{\text{инерция в точке подсч. град.}})$

• Сходимость метода Нестерова:

$$O\left(\sqrt{\frac{L}{\mu}}\log\frac{\|x^0-x^*\|^2}{\varepsilon}\right)$$ итераций / градиентных вызовов

для достижения точности ε: $\|x^k-x^*\|^2 \le \varepsilon$.

лучше, чем у град. спуска $O\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$

Плюсы:

+ как и у HB
+ оптим. оц. сход. (быстрее, чем град. спуск)

Минусы:

+ как и у HB ($\tau_k = \frac{k}{k+3}$; $\frac{k}{k+2}$)

• А можно ли еще лучше?

Нижние оценки

З М

μ - сильн. в.
L - гладкие

"память" M_k

• иниц. $M_0 = \{x^0\} = \{\vec{0}\}$

• обновл. $M_k = \mathrm{span}\{x', \nabla f(x'')\}$
$x', x'' \in M_{k-1}$

• выход: $x \in M_k$

GD, HB, Nesterov удовл.

Плохая функция (задача)

$$f(x) = \frac{L-\mu}{8} x^T A x + \frac{\mu}{2}\|x\|^2 - \frac{L-\mu}{4} e_1^T x$$

$$A = \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & -1 & \ddots & & \\ & & & 2 & -1 \\ 0 & & & -1 & \xi \end{pmatrix} \qquad e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ξ — подбир. далее

Можно показать:

- $A \succeq 0$, тогда f — μ-сильно выпукла
- $\|A\|_2 \le 4$, тогда f — L-гладкой

1) Решение $\min\limits_{x \in \mathbb{R}^d} f(x)$

$$\nabla f(x^*) = 0 \Rightarrow \frac{L-\mu}{4} A x^* + \mu x^* - \frac{L-\mu}{4} e_1 = 0$$

по компонентам:

- первая коорд.:

$$\frac{2(L+\mu)}{L-\mu} x_1^* - x_2^* = 1$$

нач. усл.

- остальные (кроме первой и послед.)

$$-x_{k-1}^* + \frac{2(L+\mu)}{L-\mu} x_k^* - x_{k+1}^* = 0$$

линейная рекурр.

- послед. коорд.: ... (завис. от ξ)

нач. усл.

Решение (при прав. ξ):

$$x_k^* = q^k \qquad q = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}$$

$\cancel{C\lambda^k}$ за счет ξ

$x^{k+1} = A x^k + B x^{k-1}$

$\lambda^2 = A\lambda + B$

λ_1, λ_2 — корни

$\lambda_1 \ne \lambda_2$

$x^k = C_1 \lambda_1^k + C_2 \lambda_2^k$

C_1, C_2 — нач. услов.

2) Как работает метод из класса выше:

$$\nabla f(x) = \frac{L-\mu}{4} Ax + \mu x - \frac{L-\mu}{4} e_1$$

$$\sim Ax + x + e_1$$

- стартуем из 0 — один вызов ∇f

  $\mathrm{span}\{0;\ \nabla f(0)\} = \mathrm{span}\{e_1\}$ — первая коорд не нулевая

- 1ая коорд $\neq 0$ — второй вызов ∇f

  $\mathrm{span}\{e_1,\ \underbrace{\nabla f(x)}_{x \in \mathrm{span}\{e_1\}}\} = \mathrm{span}\{e_1, e_2\}$ — 1ая и 2ая коорд ненул.

  $\underbrace{x \in \mathrm{span}\{e_1\}}_{\mathrm{span}\{e_1, e_2\}}$

- k вызовов ∇f $\Rightarrow$ макс $k \neq 0$ коорд. в выходе метода

3) Размерность задачи $d = 2\bar{K}$ (кол-во вызовов ∇f)

В лучшем случае

- $\bar{K}$ первых коорд угадали идеально
- $\bar{K}$ последних менее 0

$$\|x^{\bar K} - x^*\|^2 \geq \underbrace{\sum_{k=1}^{\bar K} 0^2}_{} + \sum_{k=\bar K+1}^{2\bar K} (0 - q^k)^2$$

выход $\quad q^k \quad$ оценка все угад. сверху

$$= \sum_{k=\bar K+1}^{2\bar K} (q^k)^2 = q^{2\bar K} \sum_{k=1}^{\bar K} (q^k)^2$$

$$\|x^0 - x^*\|^2 = \|x^*\|^2 = \sum_{k=1}^{\bar K} q^k + \sum_{k=\bar K+1}^{2\bar K} q^k = \sum_{k=1}^{K} q^k + q^{2\bar K} \sum_{k=1}^{\bar K} (q^k)^2$$

x^0: 0

$$\|x^{\bar{k}} - x^*\|^2 \geq q^{2\bar{k}} \cdot \frac{\|x^0 - x^*\|^2}{1 + q^{2\bar{k}}}$$

$$\geq \frac{q^{2\bar{k}}}{2} \|x^0 - x^*\|^2$$

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

- <u>Ф-ла нижн. оценки</u>

f - L-гладкая, μ - сильно вып., любой метод из класса найдет решение ($\|x^k - x^*\|^2 \leq \varepsilon$) не быстрее, чем за

$$\Omega\left(\sqrt{\frac{L}{\mu}} \log \frac{\|x^0 - x^*\|^2}{\varepsilon}\right) \text{ вызов оракула}$$

Нестеров оптимален!