Різні модифікації градієнтного методу

30 травня 2023 р.

Визначення

Визначення

Умова Ліпшиця для градієнта з константою L

$$\|\nabla f(x_1) - \nabla f(x_2)\| \geq L\|x_1 - x_2\|$$

Визначення

Гладка функція — це функція, що має неперервну похідну на всій області визначення.

Визначення

Визначення

m-сильно опукла функція - функція, що задовільняє нерівність:

$$f(y) \geq f(x) + \nabla f(x)(y - x) + \frac{m}{2}\|y - x\|^2$$

Формалювання проблеми

$\min\limits_{x \in \mathbb{R}^n} f(x)$, де f гладка і опукла функція. Часто ще додають сильну m-опуклість та умову Ліпшиця.

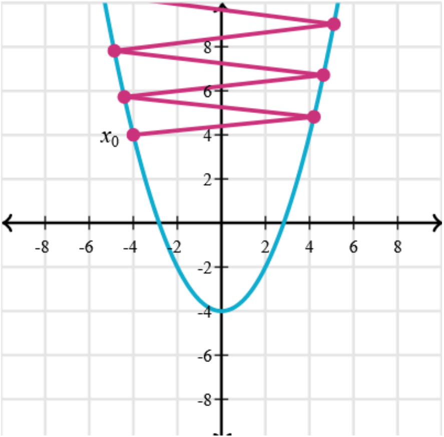
Градієнтний метод

Згадаємо базовий градієнтний методу

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

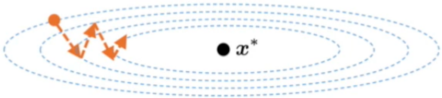
$\alpha = \frac{2}{L}, N = O(\frac{L}{m} \ln(\frac{\|x_0 - x^*\|^2}{\varepsilon}))$

Проблема незмінного кроку

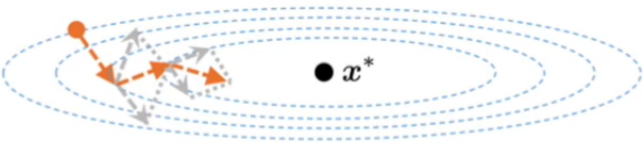


Змінний крок

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \alpha_k \to 0$$



Метод важкого шара Поляка



Метод важкого шара Поляка

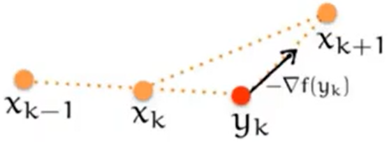
$x_{k+1} = x_k - \alpha_k \nabla f(x_k) + \beta(x_k - x_{k-1})$, β- масса шара

Метод Нестерова

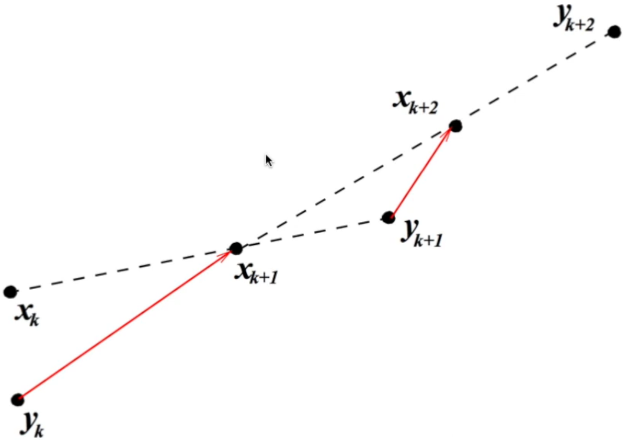
$$x_{k+1} = y_k - \alpha_k \nabla f(y_k)$$
$$y_{k+1} = x_{k+1} + \beta_k (x_{k+1} - x_k)$$

Або

$$x_{k+1} = x_k + \beta_k (x_k - x_{k-1}) - \alpha_k \nabla f(x_k + \beta_k (x_k - x_{k-1}))$$



Метод Нестерова



Метод Нестерова

Теорема

Для досягнення точності ε, отримання x_N, такого що $f(x_N) - f^* \leq \varepsilon$, методу Нестерова потрібно
- в опуломому випадку: $N = O\left(\frac{LR^2}{\sqrt{\varepsilon}}\right)$
- у сильно опуклому випадку $N = O\left(\sqrt{\frac{L}{\mu}} \log\left(\frac{1}{\varepsilon}\right)\right)$

Стохастичний градієнтний метод

Задача:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \to min, x \in \mathbb{R}^d, f_i : \mathbb{R}^d \to \mathbb{R}$$

1. f_i опуклі
2. f_i гладкі з L_i константою
3. F m-сильно опукла

Запозичення в ML

$$x_{k+1} = x_k - \alpha_k \nabla f_{i_k}(x_k) + \beta_k (x_k - x_{k-1})$$

Метод важкого шара Поляка

Теорема

Якщо взяти за параметри методу $\alpha_k = \frac{2\mu}{k+2}$, $\beta_k = \frac{k}{k+2}$ де $0 < \mu \leq \frac{1}{4L_{max}}$
Тоді для довільного $N \in \mathbb{N}$ має місце оцінка

$$M(F(x_N) - \min F) \leq \frac{\|x_0 - x^*\|^2}{\mu(N+1)} + \frac{2\mu}{n} \sum_{i=1}^{n} \|\nabla f_i(x^*)\|^2$$

Adam

$$m_i^k = \beta_1 m_i^{k-1} + (1-\beta_1)\, g_i^k, \quad \hat{m}_i^k = \frac{m_i^k}{1-\beta_1^k},$$
$$v_i^k = \beta_2 v_i^{k-1} + (1-\beta_2)\left(g_i^k\right)^2, \quad \hat{v}_i^k = \frac{v_i^k}{1-\beta_2^k}, \qquad (1)$$
$$x_i^{k+1} = x_i^k - \frac{\eta}{\sqrt{\hat{v}_i^k + \varepsilon}}\hat{m}_i^k, \quad i = \overline{1, d}, \quad \varepsilon = 10^{-8},$$