

Ускоренные методы нулевого порядка в гладкой выпуклой стохастической оптимизации

Фанис Адикович Хафизов

Московский физико-технический институт

Курс: Автоматизация научных исследований
(практика, В. В. Стрижов)/Группа 105

Эксперт: к.ф.-м.н А. Н. Безносиков

Консультант: А. И. Богданов

2024

Цель исследования

Цель:

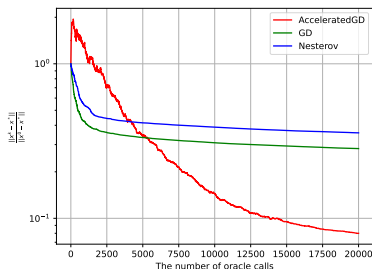
- ▶ Создать ускоренный безградиентный метод решения задачи безусловной гладкой стохастической оптимизации
- ▶ Доказать сходимость в случае детерминированного шума
- ▶ Поставить численный эксперимент и сравнить с методом Нестерова и градиентным спуском в случаях детерминированного и стохастического шума

Доклад с одним слайдом

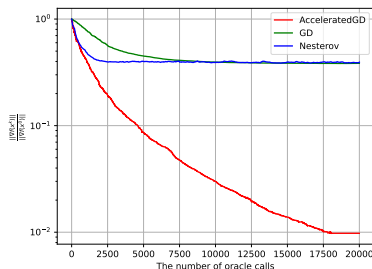
$$\min_{x \in \mathbb{R}^d} f(x)$$

$$f_\delta(x) = f(x) + \delta(x)$$

$$\tilde{\nabla} f_\delta(x) := d \frac{f_\delta(x + \tau e_i) - f_\delta(x - \tau e_i)}{2\tau} e_i, i \sim U\{1, d\}$$



Зависимость $\frac{\|x^k - x^*\|}{\|x^0 - x^*\|}$ от числа итераций, квадратичная задача минимизации



Зависимость $\frac{\|\nabla f(x^k)\|}{\|\nabla f(x^0)\|}$ от числа итераций, задача логистической регрессии

- ▶ Aleksandr Beznosikov, Sergey Samsonov, Marina Sheshukova, Alexander Gasnikov, Alexey Naumov, Eric Moulines. First Order Methods with Markovian Noise: from Acceleration to Variational Inequalities. 2023.
- ▶ Andrey Veprikov, Alexander Bogdanov, Vladislav Minashkin, Aleksandr Beznosikov, Alexander Gasnikov. New Aspects of Black Box Conditional Gradient: Variance Reduction and One Point Feedback. 2024.
- ▶ Eduard Gorbunov, Pavel Dvurechensky, Alexander Gasnikov. An Accelerated Method for Derivative-Free Smooth Stochastic Convex Optimization. 2020.

Постановка задачи. Детерминированный случай

Дано

$f : \mathbb{R}^d \rightarrow \mathbb{R}$ — μ -сильно выпуклая, L -гладкая функция. Доступ к оракулу нулевого порядка $f_\delta(x) = f(x) + \delta(x)$: $|\delta(x)| \leq \Delta$.

Требуется

Построить алгоритм, решающий задачу

$$\min_{x \in \mathbb{R}^d} f(x), \tag{1}$$

обращаясь к оракулу f_δ .

Постановка задачи. Стохастический случай

Дано

$f : \mathbb{R}^d \rightarrow \mathbb{R}$ — μ -сильно выпуклая, L -гладкая функция. Доступ к оракулу нулевого порядка $f_\delta(x) = f(x) + \delta(x, \xi)$:
 $\mathbb{E}[|\delta(x, \xi)|^2] \leq \Delta^2$.

Требуется

Построить алгоритм, решающий задачу

$$\min_{x \in \mathbb{R}^d} f(x) := \mathbb{E}_{\xi \sim \pi}[f(x, \xi)],$$

обращаясь к оракулу f_δ .

Предложенный метод

Algorithm Accelerated Gradient Descent

Require: stepsize $\gamma > 0$, momentums θ, η, β, p , number of iterations N , approximation parameter $\tau > 0$.

Initialization: choose $x^0 = x_f^0$

for $k = 0, 1, \dots, N - 1$ **do**

$$x_g^k = \theta x_f^k + (1 - \theta)x^k$$

Sample $i \sim U\{1, d\}$

$$g^k = d \frac{f_\delta(x_g^k + \tau e_i) - f_\delta(x_g^k - \tau e_i)}{2\tau} e_i$$

$$x_f^{k+1} = x_g^k - p\gamma g^k$$

$$x^{k+1} = \eta x_f^{k+1} + (p - \eta)x_f^k + (1 - p)(1 - \beta)x^k + (1 - p)\beta x_g^k$$

end for

Скорость сходимости

Theorem

Ускоренный градиентный спуск (Algorithm 1) имеет скорость сходимости на задаче (1):

$$\begin{aligned} \mathbb{E} \left[\|x^N - x^*\|^2 + \frac{6}{\mu}(f(x_f^N) - f(x^*)) \right] &\leq \\ &\leq \exp \left(-N \sqrt{\frac{p^2 \mu \gamma}{3}} \right) \left(\|x^0 - x^*\|^2 + \frac{6}{\mu}(f(x_f^0) - f(x^*)) \right) + \quad (2) \\ &\quad + \frac{3\sqrt{3}\gamma}{\mu^{3/2}} \cdot d \left(\frac{L\tau}{2} + \frac{\Delta}{\tau} \right)^2 \left(2 + \sqrt{\frac{3}{\gamma\mu}} \right), \end{aligned}$$

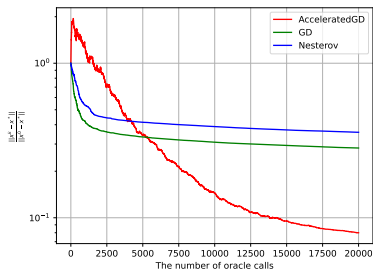
где $\gamma \in (0, \frac{3}{4L}]$, β, θ, η, p такие, что:

$$p \simeq (2(1 + \gamma L))^{-1}, \beta \simeq \sqrt{p^2 \mu \gamma}, \eta \simeq \sqrt{\frac{1}{\mu \gamma}}, \theta \simeq \frac{p\eta^{-1} - 1}{\beta p \eta^{-1} - 1}. \quad (3)$$

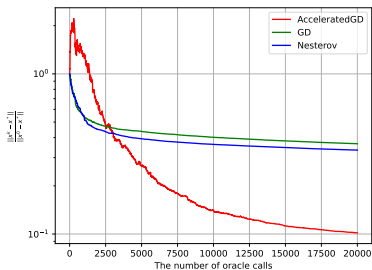
Вычислительный эксперимент. Квадратичная задача

$$\min_{x \in \mathbb{R}^d} f(x) = x^T A x + b^T x + c$$

$$A \in \mathbb{S}_d : \mu \preceq A \preceq L, \mu = 1, L = 1000, d = 100$$



Детерминированный шум,
 $\Delta = 10^{-6}$, $\tau = 10^{-4}$

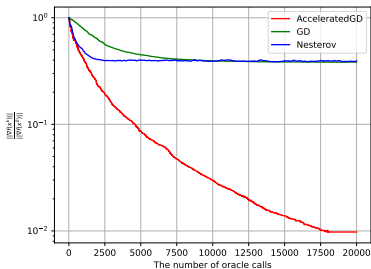


Стохастический шум, $\Delta = 10^{-6}$,
 $\tau = 10^{-4}$

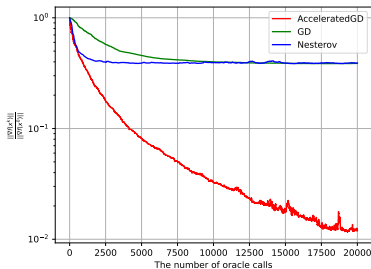
Вычислительный эксперимент. Задача логистической регрессии

Датасет mushrooms, $d = 112$, $m = 8124$, $y_i \in \{0, 1\}$, $\lambda = 0, 1$.

$$\min_{w \in R^d} f(w) = \frac{1}{m} \sum_{k=1}^m \log(1 + \exp(-y_k \cdot (Xw)_k)) + \lambda \|w\|_2^2$$



Детерминированный шум,
 $\Delta = 10^{-6}$, $\tau = 10^{-4}$



Стохастический шум, $\Delta = 10^{-6}$,
 $\tau = 10^{-4}$

- ▶ Предложен метод нулевого порядка, решающий поставленную задачу
- ▶ Получена скорость сходимости на описанном классе функций в случае детерминированного шума
- ▶ Показано превосходство предложенного метода над методом Нестерова и градиентным спуском

Будущая работа

Развить теорию на стохастический случай.