

Санкт-Петербургский государственный университет
Математико-механический факультет
Информационно-аналитические системы

Ким Юния Александровна
18.Б07-мм

Вычислительный практикум

Отчёт по заданию №14

Преподаватель:
Евдокимова Т.О.

Санкт-Петербург
2021

Содержание

1. Ссылка на код	3
2. Постановка задачи	3
3. Теоретическая часть	3
3.1. Градиентный спуск	3
3.2. Метод Нестерова	3
4. Численный эксперимент	4
4.1. Описание	4
4.2. Результаты	4
4.3. Анализ	7

1. Ссылка на код

<https://github.com/yuniyakim/MethodsOfComputation/pull/22>

2. Постановка задачи

Задача – реализация двух методов нахождения минимума сильно выпуклой функции: градиентного спуска и метода Нестерова, а также сравнение количества итераций.

3. Теоретическая часть

Пусть $f : H \mapsto \mathbb{R}$, где H – Гильбертово, f – сильно выпуклая функция, обладающая следующим свойством: $\forall x, y \in H \exists L > 0 : \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$.

Исходная задача – нахождение минимума заданной функции f .

3.1. Градиентный спуск

1. Выбираем начальное приближение $\theta_0 \in H$, коэффициент скорости обучения λ и параметр ϵ .
2. $\theta_{n+1} = \theta_n - \lambda \nabla f(\theta_n)$.
3. Повторяем пункт 2 до тех пор, пока $\|\theta_{n+1} - \theta_n\| > \epsilon$ или пока не достигнется максимальное количество итераций.

3.2. Метод Нестерова

1. Выбираем начальное приближение $y_0 \in H$ и точку $z \in H : z \neq y_0$ и $\nabla f(y_0) \neq \nabla f(z)$.
2. Полагаем:
 - $a_0 = 1$,
 - $x_{-1} = y_0$,
 - $a_{-1} = \frac{\|y_0 - z\|}{\|\nabla f(y_0) - \nabla f(z)\|}$.
3. На k -ом шаге метода:
 - (a) Выбираем наименьший индекс $i : f(y_k) - f(y_k - 2^{-i} a_{k-1} \nabla f(y_k)) \geq 2^{-1-i} a_{k-1} \|\nabla f(y_k)\|^2$.
 - (b) Полагаем:
 - $a_k = 2^{-i} a_{k-1}$,
 - $x_k = y_k - a_k \nabla f(y_k)$,
 - $a_{k+1} = \frac{1 + \sqrt{4a_k^2 + 1}}{2}$,
 - $y_{k+1} = x_k + \frac{(a_k - 1)(x_k - x_{k-1})}{a_{k+1}}$.
4. Повторяем пункт 3 до тех пор, пока $\|y_{k+1} - y_k\| > \epsilon$ или пока не достигнется максимальное количество итераций.

4. Численный эксперимент

4.1. Описание

Для численного эксперимента бралась функция $g(x) = 4x + 11$, $N = 100$ равноотстоящих точек x_1, \dots, x_{100} на отрезке $[-10; 10]$, функции $y_1 = g(x_1), \dots, y_{100} = g(x_{100})$, а также квадратичная функция потерь $L(a, b) = \frac{1}{2} \sum_{i=1}^N (ax_i + b - y_i)^2$.

Ставилась задача минимизации функции потерь $L(a, b)$, т.е. восстановления по N точкам параметров исходной функции $g(x)$.

4.2. Результаты

На рис. 4.1 представлены результаты работы градиентного спуска и метода Нестерова для данной функции, на рис. 4.2 – результаты работы метода Нестерова. Показаны значения параметров λ и ϵ , количество итераций, точность приближения коэффициентов функции $g(x)$ и значения функции потерь в получившейся точке минимума.

Lambda	Epsilon	Amount of iterations	a_ε-a	b_ε-b	Loss
0.01	0.01	6	5.53423e+09	1.5795e+06	5.10562e+22
0.01	0.0001	6	4.69324e+09	22781.8	3.67183e+22
0.01	1e-06	6	3.92064e+09	8.19216e+08	2.56898e+22
0.0001	0.01	235	0.00307557	0.994632	49.4498
0.0001	0.0001	691	3.08421e-05	0.00995826	0.00495686
0.0001	1e-06	1164	3.58101e-07	9.96425e-05	4.96288e-07
1e-06	0.01	0	2.86216	10.6761	19049.3
1e-06	0.0001	2000	0.0323557	8.61853	3712.91
1e-06	1e-06	2000	0.0277376	7.11481	2530.34

Рисунок 4.1. Результаты работы градиентного спуска

Epsilon	Amount of iterations	a_ε-a	b_ε-b	Loss
0.01	0	4.71134	10.5814	42101.7
0.0001	512	0.000399411	0.129155	0.833796
1e-06	944	1.7307e-06	0.000543544	1.47676e-05
1e-08	1710	6.45327e-08	4.70157e-06	1.10915e-09

Рисунок 4.2. Результаты работы метода Нестерова

Для сравнения функцию $g(x)$ «зашумели», добавив к ней $N(0, 25)$ нормально распределённых точек, после чего минимизировали $L(a, b)$ с новыми значениями.

На рис. 4.3, 4.4 и 4.5 представлены результаты работы градиентного спуска, на рис. 4.6 и 4.7 – результаты работы метода Нестерова. Точки красного цвета соответствуют начальным точкам (x_i, y_i) , синего – точкам получившейся прямой. Показаны значения параметров ϵ и λ , количество итераций и значения функции потерь.

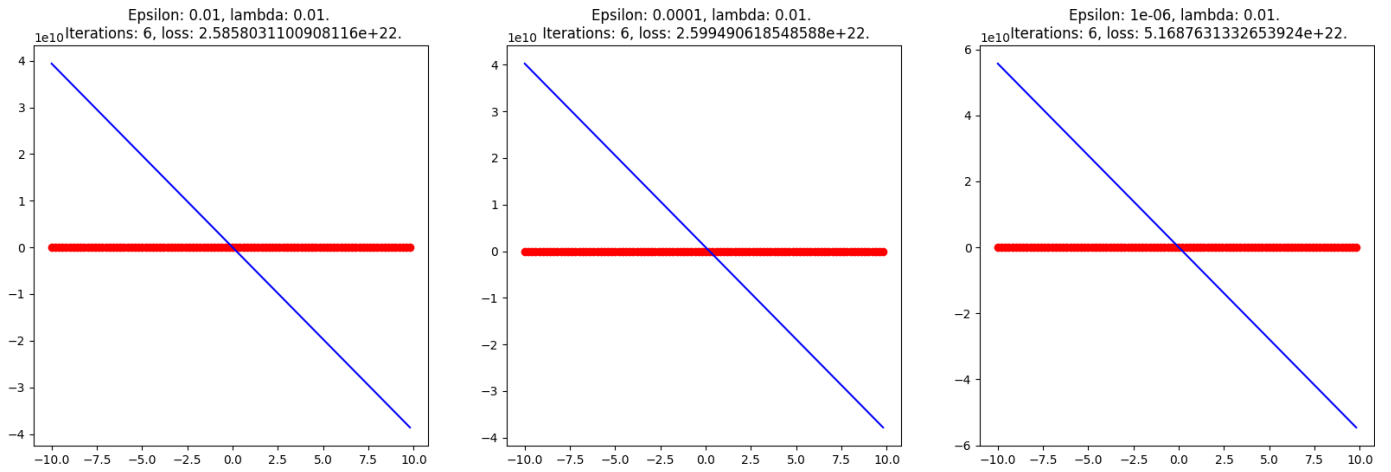


Рисунок 4.3. Результаты работы градиентного спуска с параметром $\lambda = 0.01$

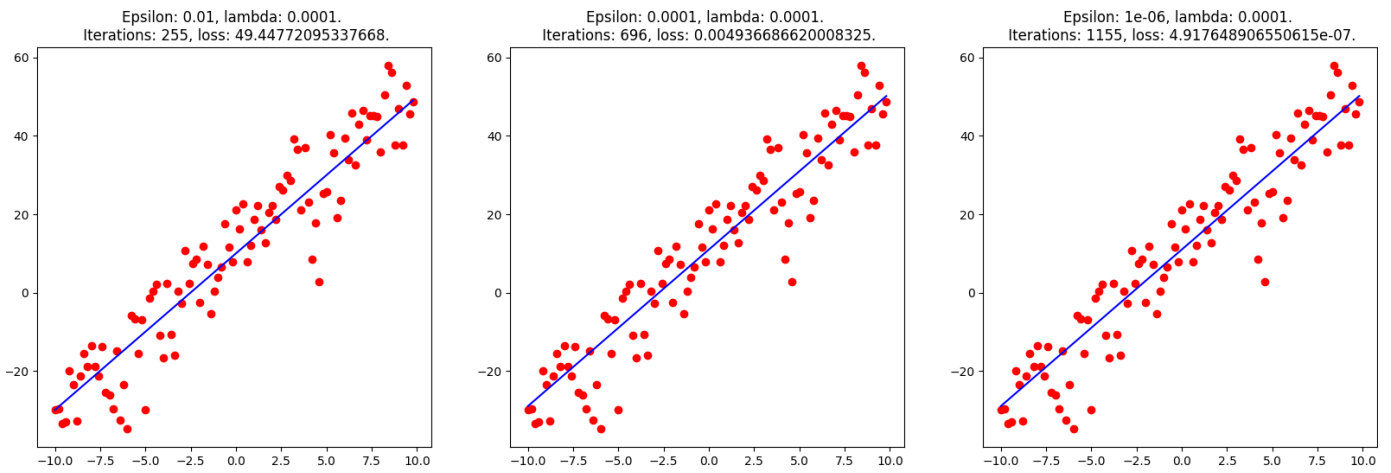


Рисунок 4.4. Результаты работы градиентного спуска с параметром $\lambda = 10^{-4}$

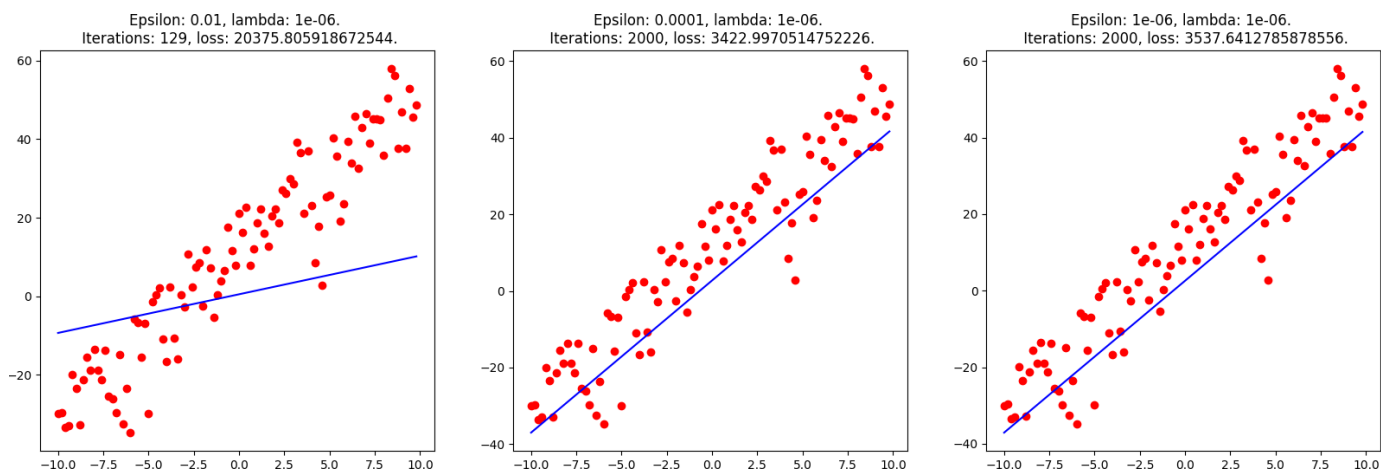


Рисунок 4.5. Результаты работы градиентного спуска с параметром $\lambda = 10^{-6}$

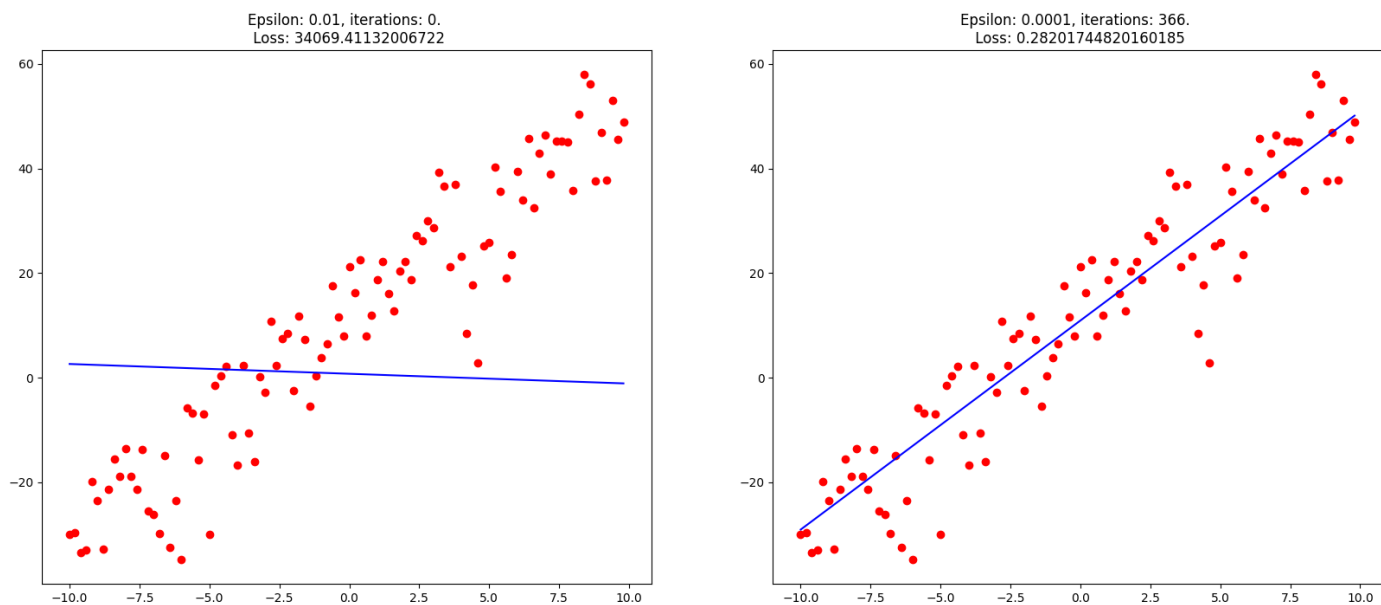


Рисунок 4.6. Результаты работы метода Нестерова

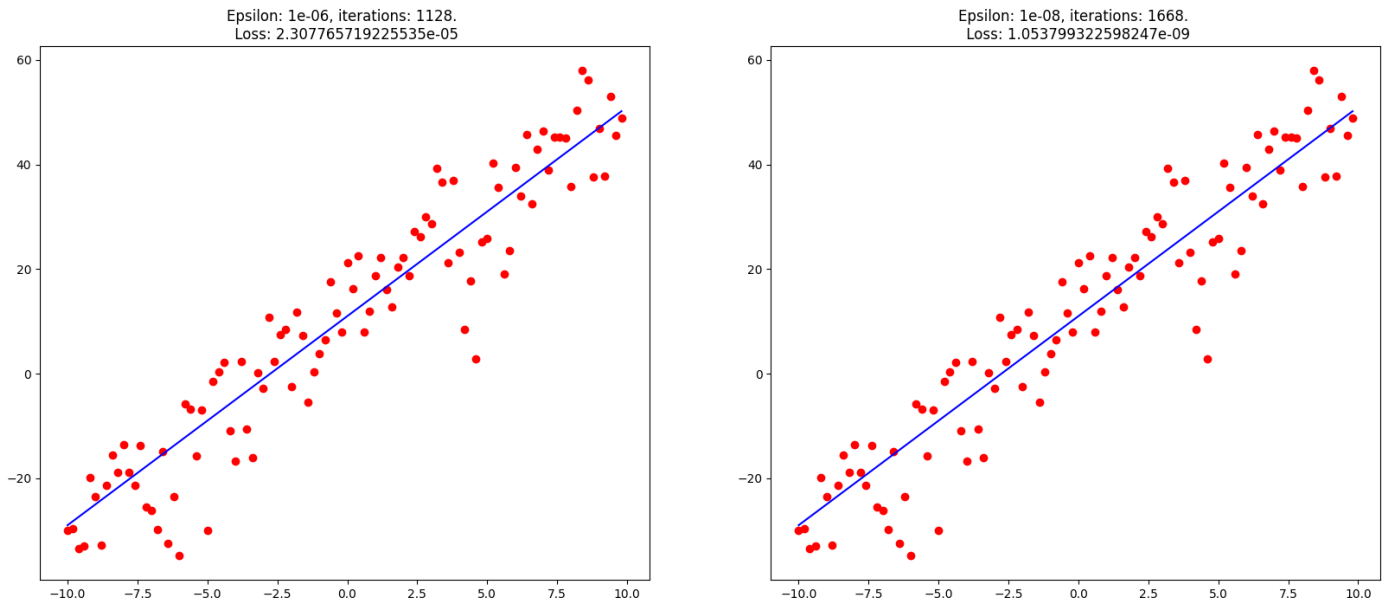


Рисунок 4.7. Результаты работы метода Нестерова

4.3. Анализ

В результате экспериментов было выявлено, что для данных функций потерь градиентный спуск:

- при $\lambda = 0.01$ расходится $\forall \epsilon$,
- при $\lambda = 10^{-4}$ лучшая точность достигается при $\epsilon = 10^{-6}$,
- при $\lambda = 10^{-6}$ сходится медленно.

Метод Нестерова расходится только при $\epsilon = 0.01$. При этом, однако, он способен достигать большей точности за меньшее количество итераций.