Сравнительный анализ методов математического моделирования для исследования стохастических данных

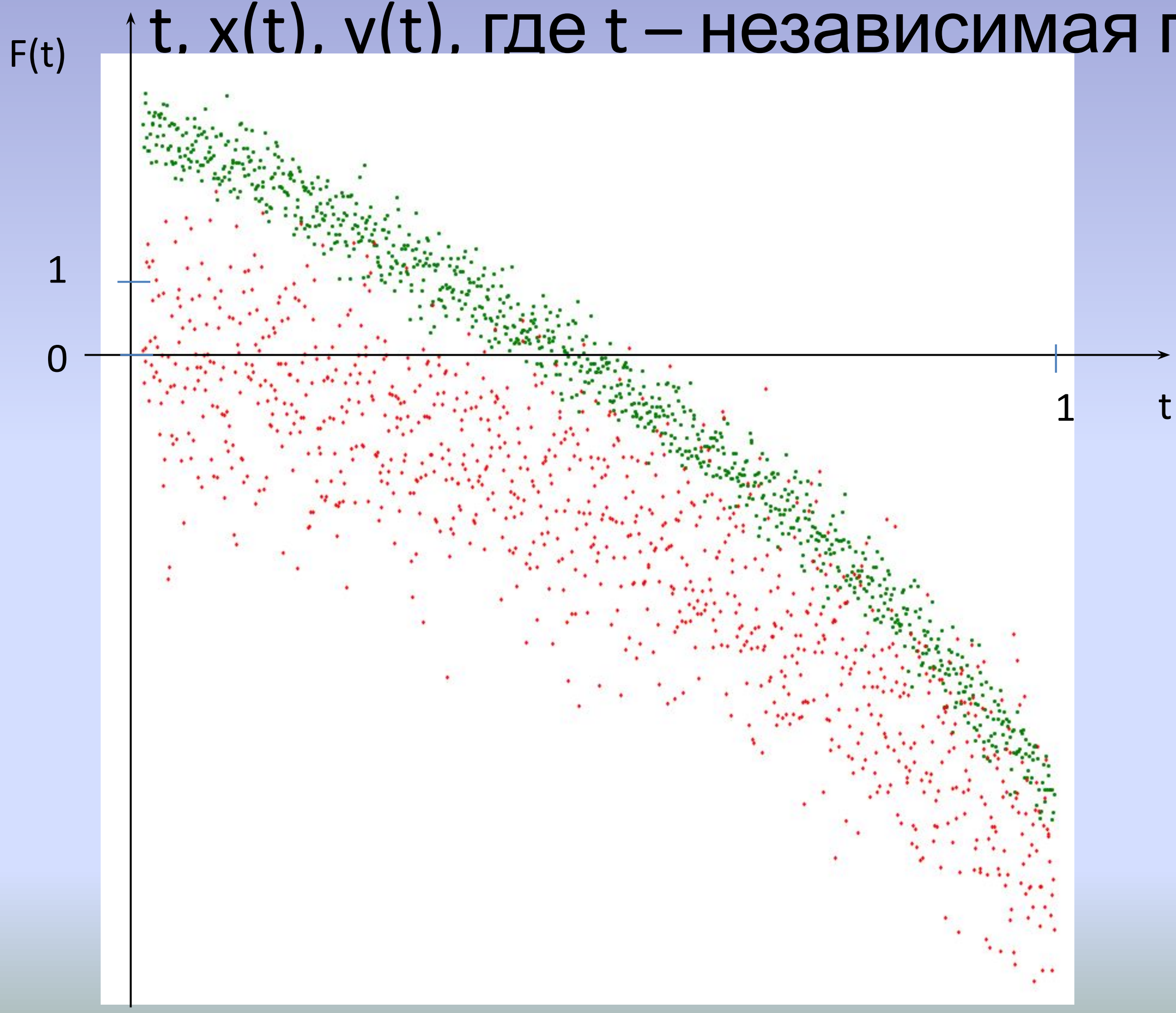
Выполнила:
Лазарева М.Р., гр.К3-31М
Научный руководитель:
п.д.ф.-м.н. Малашин А.А.

Постановка задачи

Исследовать применимость различных методов анализа для определения функциональной зависимости стохастических данных и сравнить их эффективность

Входные данные: наборы стохастических данных

t, x(t), y(t), где t – независимая переменная



Где зелеными точками обозначены - x(t),
а красными точками - y

Методы математического моделирования

1. Методы с использованием алгоритмов машинного обучения

2. Численный метод наименьших квадратов

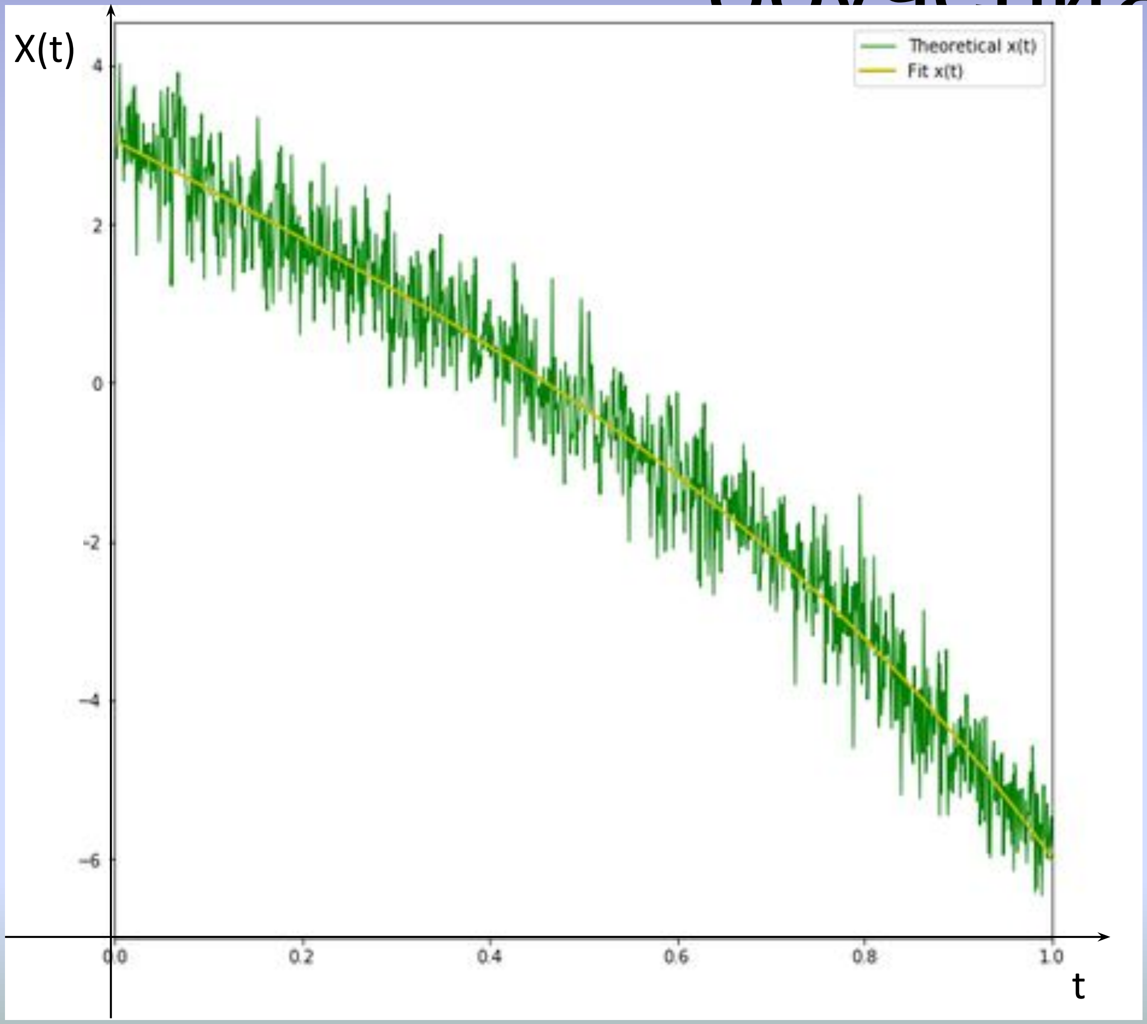
3. Метод конечно-разностной аппроксимации

(Пока добавить какие-то методы не успею)

Все методы реализованы на языке: Python

Для всех методов используются один и тот же набор данных и одна и та же модель системы дифференциальных уравнений.

Методы с использованием алгоритмов машинного обучения



- По начальным данным построена кривая, удовлетворяющая функции:

x(t) = $C_1 e^{\alpha t} + C_2 e^{\beta t}$,

Где были определены параметры:

C_1 = - 0,86256487

C_2 = 3,88362593

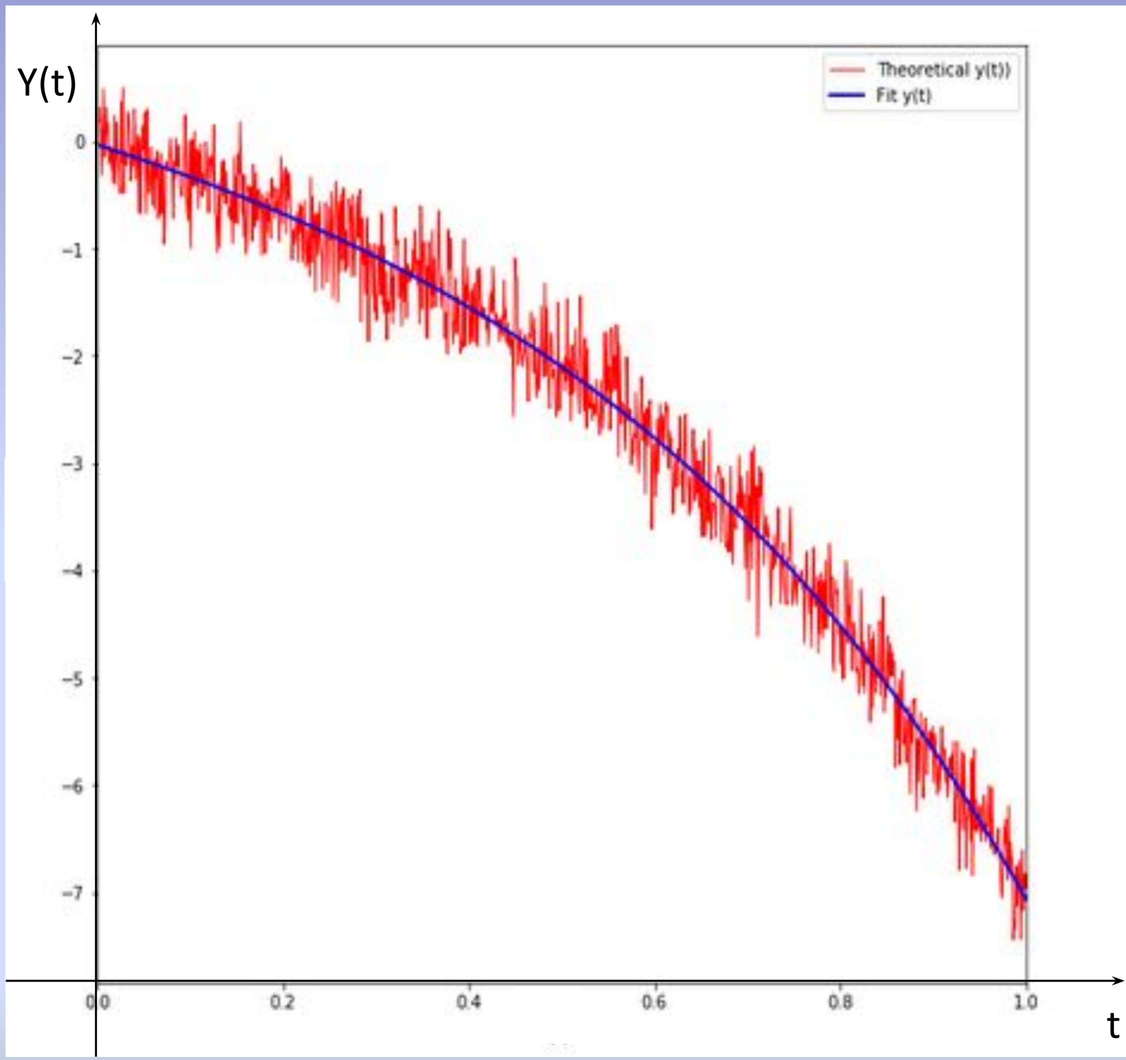
α = 2,11428402

β = - 1,1681938

Среднее отклонение исходных данных от найденной кривой:

```
--- mean deviation x(t) ---
0.38916182059105564
```

Методы с использованием алгоритмов машинного обучения



- По начальным данным построена кривая, удовлетворяющая функции:

y(t) = $C_1 e^{\alpha t} + C_2 e^{\beta t}$,

Где были определены параметры:

C_1 = - 0,98929539

C_2 = 0,99486478

α = 2,00797276

β = - 1,05145376

- Среднее отклонение исходных данных от найденной кривой:

```
--- mean deviation y(t) ---
0.4887707167132247
```

Модель системы дифференциальных уравнений

$$\begin{cases} \dfrac{dx}{dt} = Ax(t) + By(t) \\ \dfrac{dy}{dt} = Cx(t) + By(t), \end{cases}$$

Определяем параметры A,B,C,D. Для этого используем найденные функции x(t) и y(t).

Получае

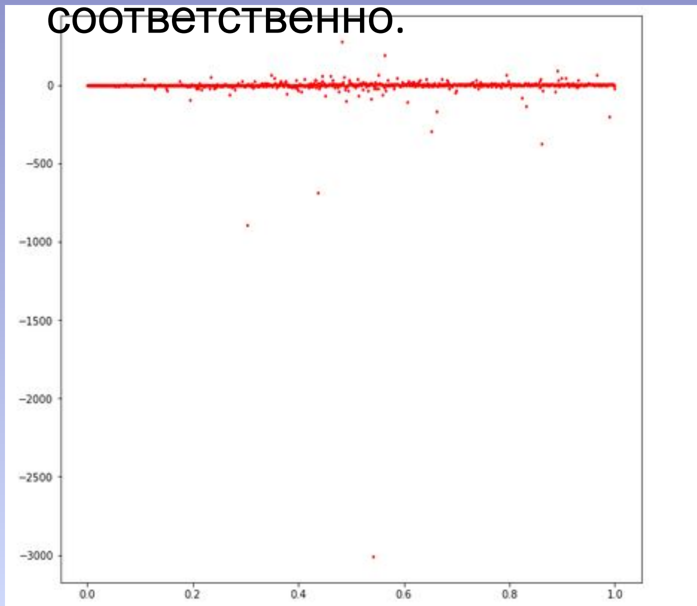
```
A
-2.11219100123443
B
3.68505557967532
C
-1.009120892129895
D
2.8878233578213712
```

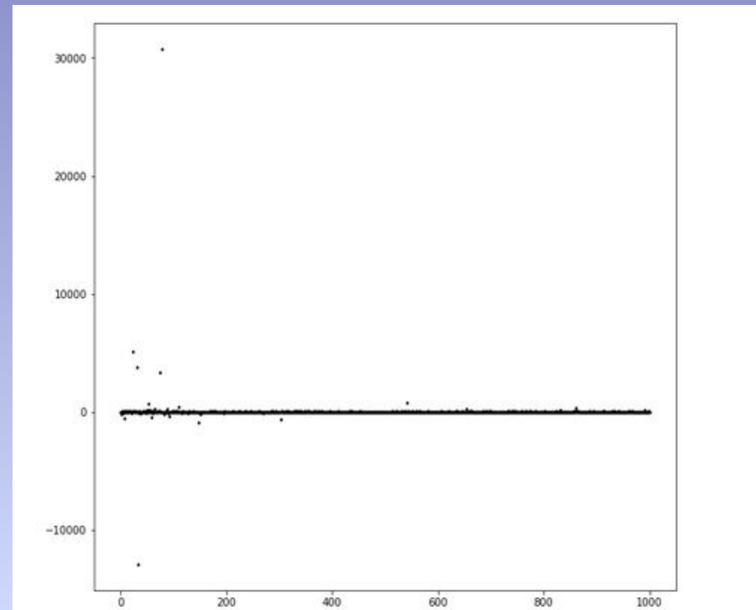
Для найденных коэффициентов найдем отклонение da, db, dc, dd соответственно, используя исходные данные и найденные параметры A,B,C,D

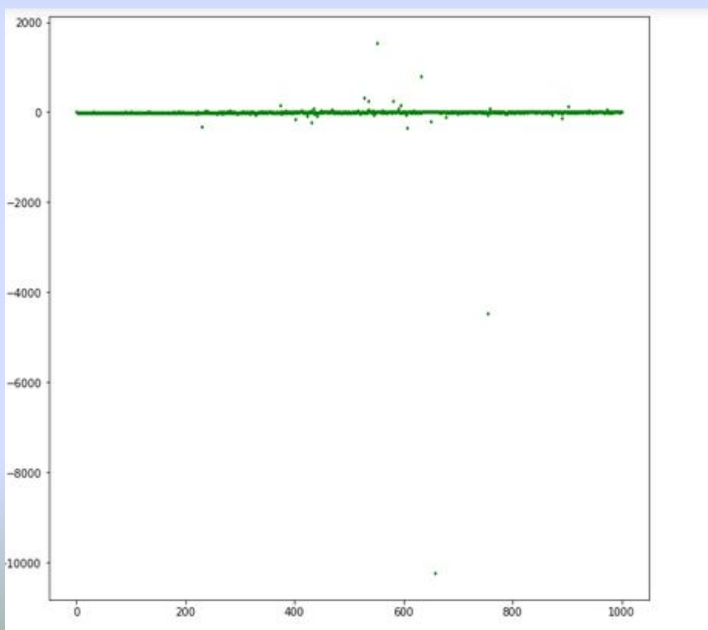
$$\begin{cases} \dfrac{dx}{dt} = (A+da)x(t) + (B+db)\,y(t) \\ \dfrac{dy}{dt} = (C+dc)\,x(t) + (D+dd)\,y(t), \end{cases}$$

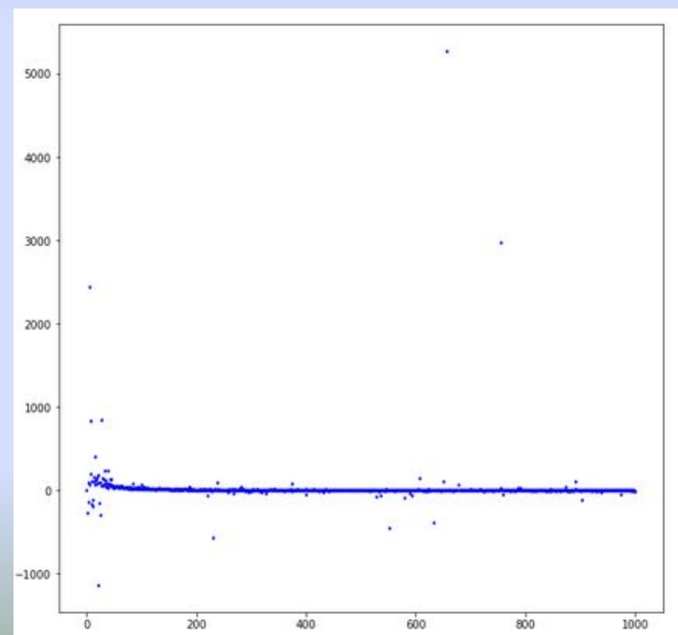
где отклонения будут определяться для каждой пары точек двух функций

Получили такое распределение значений da,db,dc,dd соответственно.

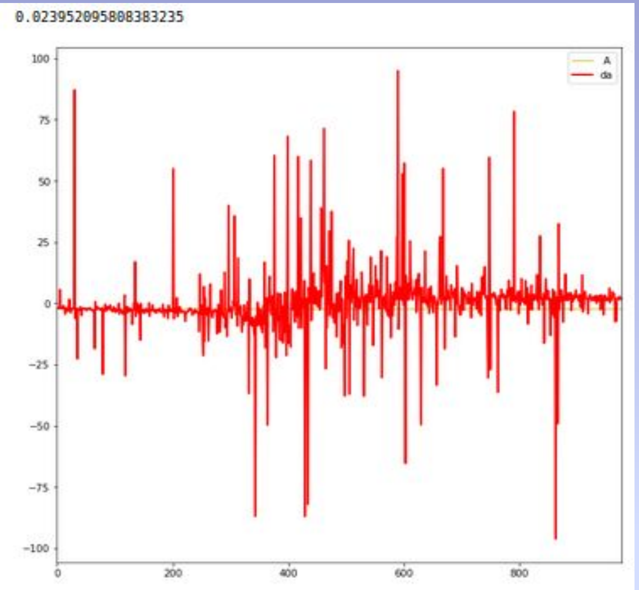


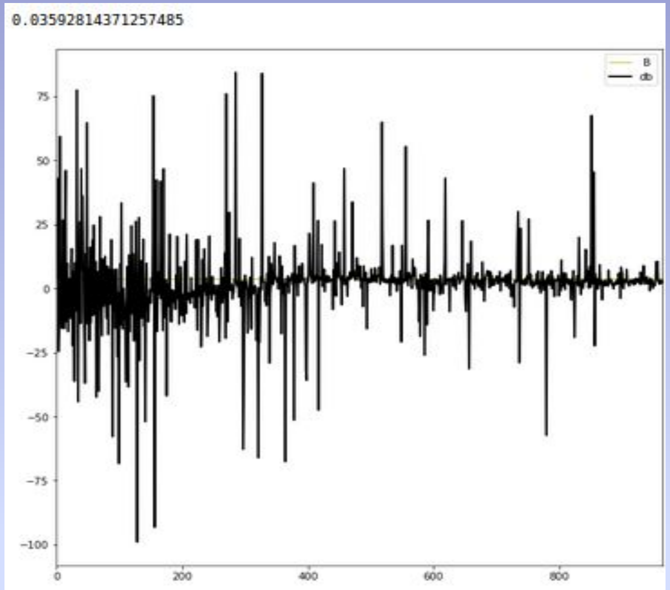


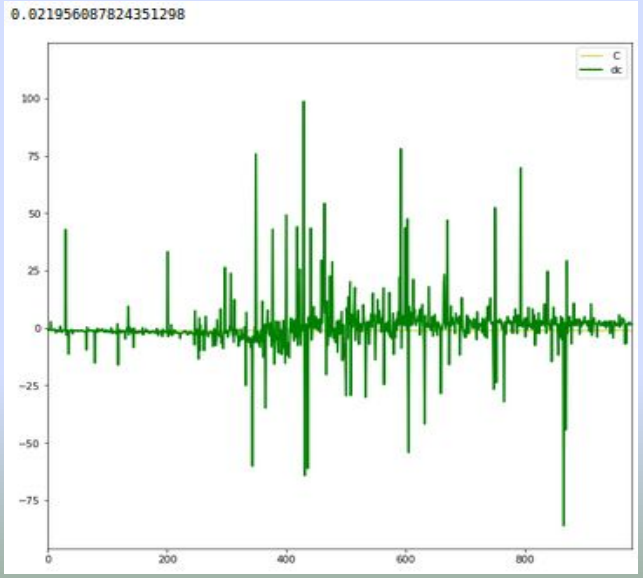


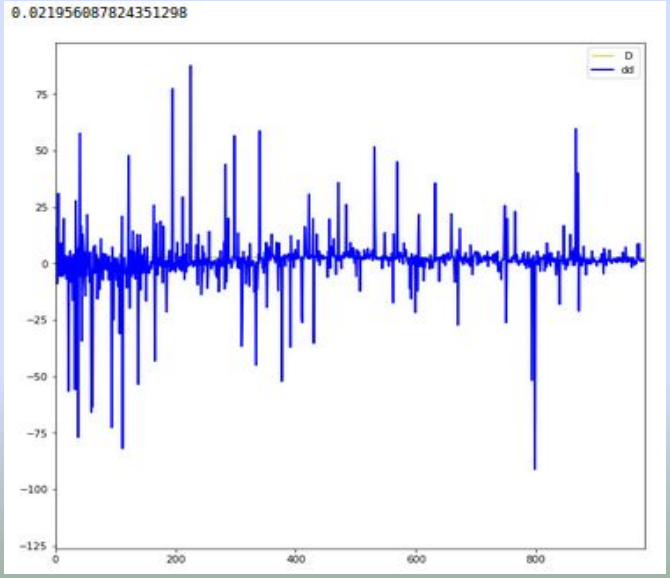


Наблюдаемые отклонение имеют грубые промахи, которые явно обусловлены шумом и далеки от найденного параметр. Отсеивая промахи, получаем:









Где числа над графиками показывают долю отсеянных значений

Численный метод наименьших квадратов.

Модель системы дифференциальных уравнений.

$$\begin{cases} \frac{dx}{dt} = Ax(t) + By(t) \\ \frac{dy}{dt} = Cx(t) + By(t), \end{cases}$$

Определяем параметры A,B,C,D. Для этого используем найденные функции x(t) и y(t).

Получае
м

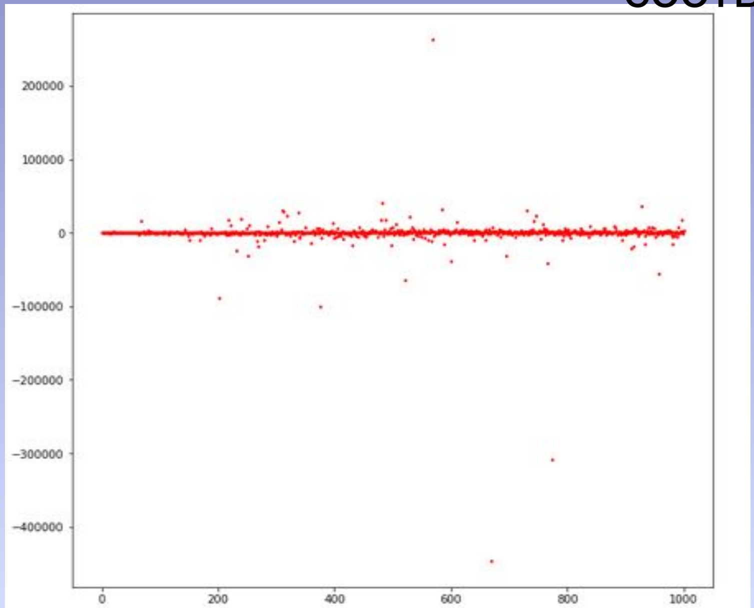
```
calculate A
-2.1796852333789127
calculate B
4.109976110835058
calculate C
-0.8034391680536414
calculate D
2.795862951488012
```

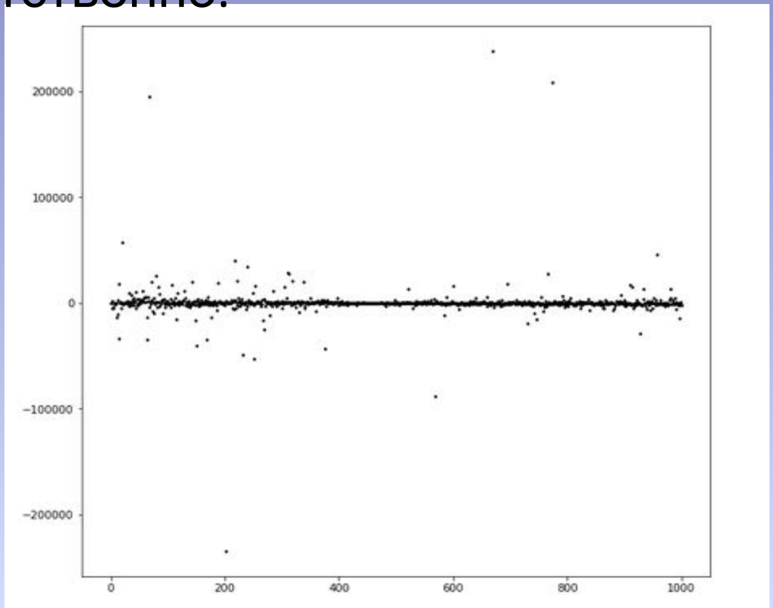
Для найденных коэффициентов найдем отклонение da, db, dc, dd соответственно, используя исходные данные и найденные параметры A,B,C,D

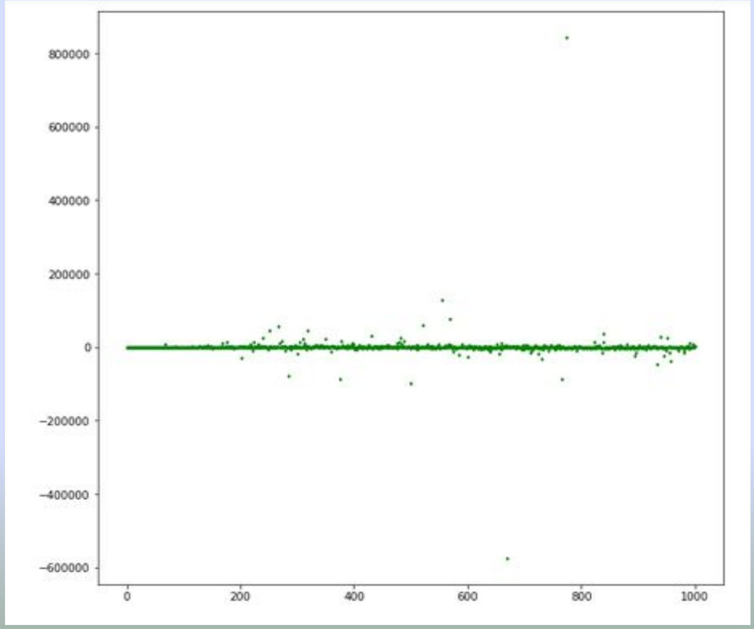
$$\begin{cases} \frac{dx}{dt} = (A+da)x(t) + (B+ db)\, y(t) \\ \frac{dy}{dt} = (C+dc)\, x(t) + (D+dd)\, y(t), \end{cases}$$

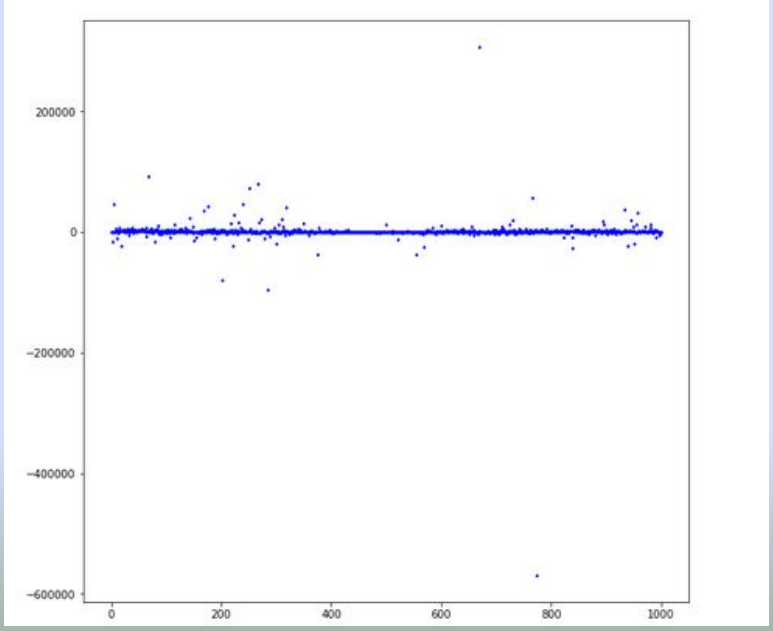
где отклонения будут определяться для каждой пары точек двух функций

Получили такое распределение значений da,db,dc,dd соответственно.

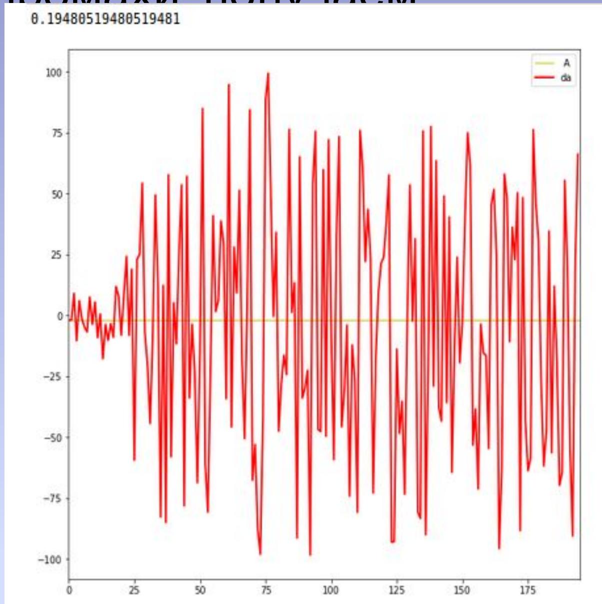


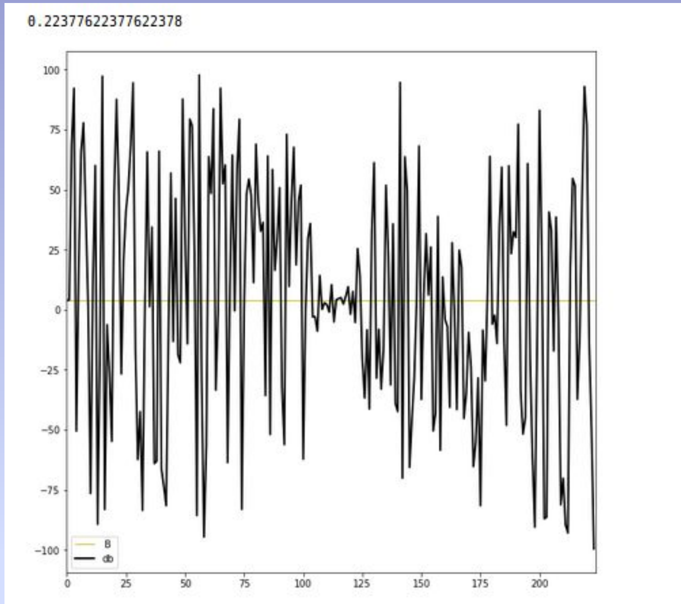


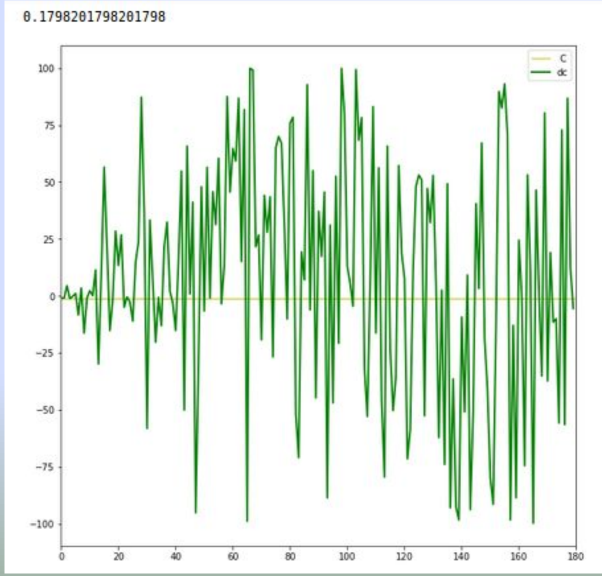


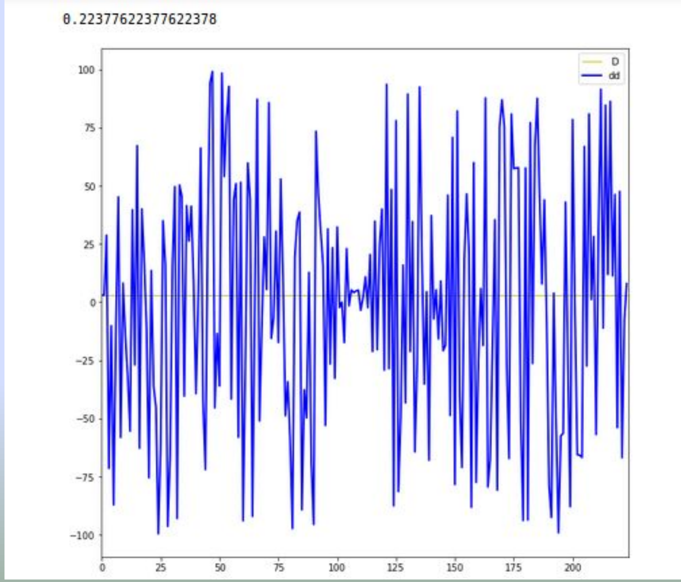


Наблюдаемые отклонение имеют грубые промахи, которые явно обусловлены шумом и далеки от найденного параметра. Отсеивая промахи, получаем:









Где числа над графиками показывают долю отсеянных значений.

Метод конечно-разностной аппроксимации.

Модель системы дифференциальных уравнений.

$$\begin{cases} \frac{dx}{dt} = \mathrm{A}x(t) + \mathrm{B}y(t) \\ \frac{dy}{dt} = \mathrm{C}x(t) + \mathrm{B}y(t), \end{cases}$$

Заменяем производные в левых частях уравнений конечно-разностной формулой.

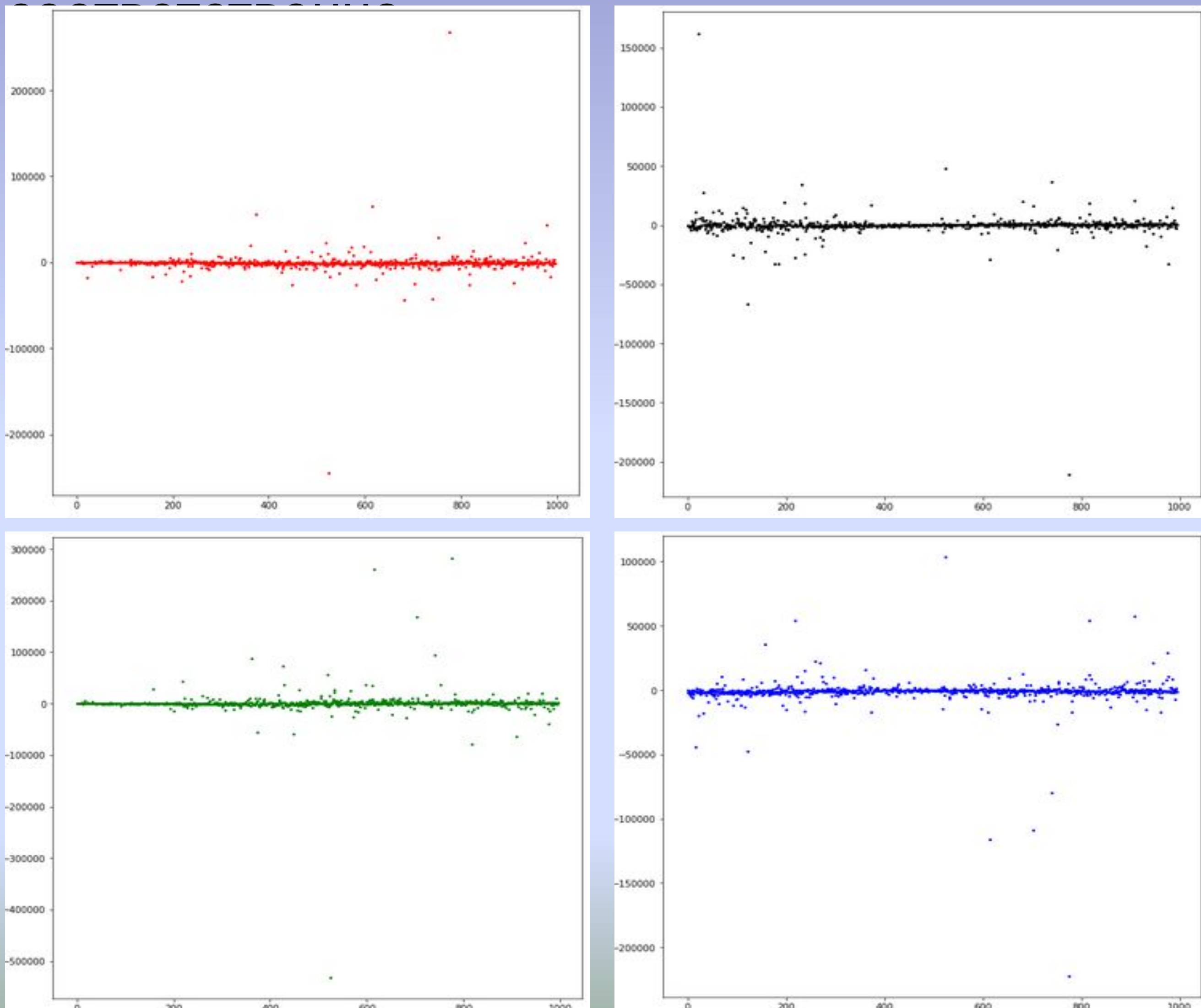
Получаем:

$$\begin{cases} \frac{x_{i+1} - x_i}{h} = \mathrm{A}x_i + \mathrm{B}y_i \\ \frac{y_{i+1} - y_i}{h} = \mathrm{C}x_i + \mathrm{C}y_i \end{cases}$$

Выражая параметры A,B,C,D, получили формулы для каждых трех точек из данных, так как для каждого уравнения имеется по две неизвестных. В отдельности для А, В из первого уравнения и для C,D – из второго

$$\begin{cases} \frac{x_{i+1} - x_i}{h} = \mathrm{A}x_i + \mathrm{B}y_i \\ \frac{x_{i+2} - x_{i+1}}{h} = \mathrm{A}x_{i+1} + \mathrm{B}y_{i+1}, \end{cases} \quad \text{и} \quad \begin{cases} \frac{y_{i+1} - y_i}{h} = \mathrm{C}x_i + \mathrm{D}y_i \\ \frac{y_{i+2} - y_{i+1}}{h} = \mathrm{C}x_{i+1} + \mathrm{D}y_{i+1} \end{cases}$$

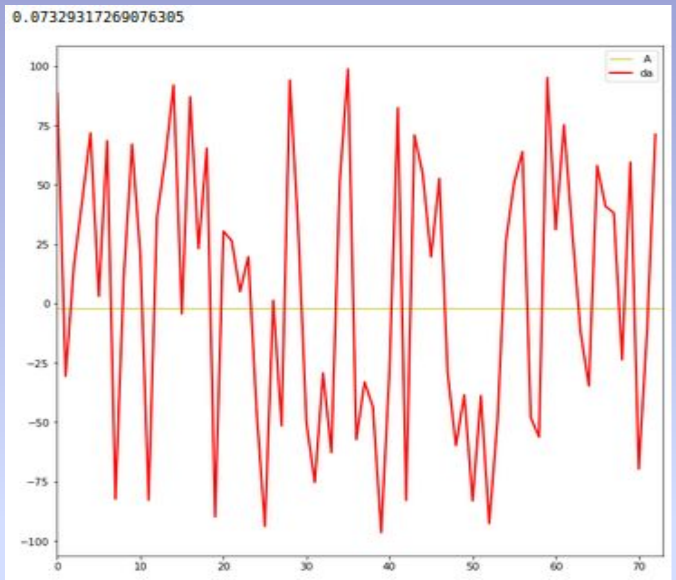
Решая в отдельности каждую систему мы получили определенные значения параметров для каждых трех точек. Получили такое распределение значений A,B,C,D соответственно

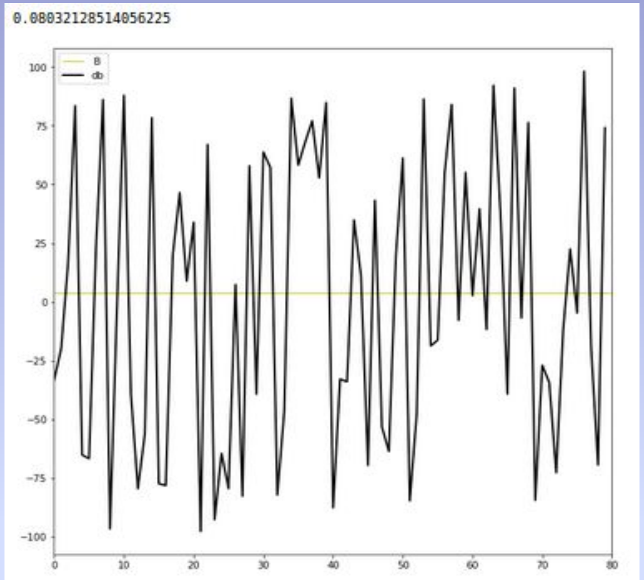


Где средние значения параметров :

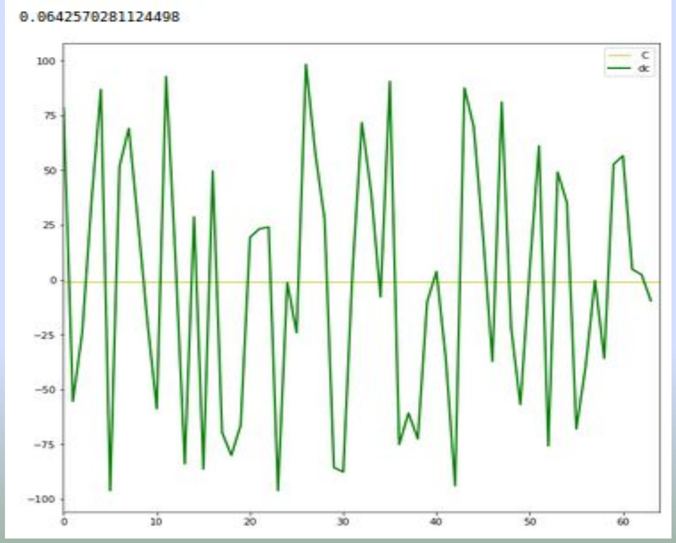
```
calculate A
-738.4608758285038
calculate B
-4.60249111197583
calculate C
673.5516695186013
calculate D
-1063.5325721295128
```

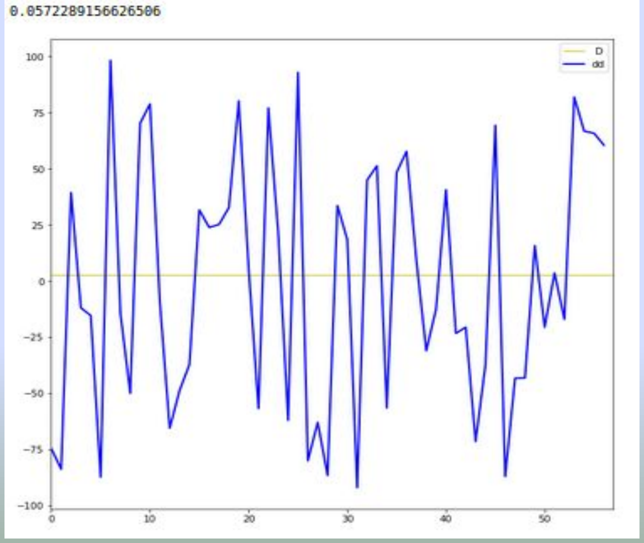
Наблюдаемые значения имеют грубые промахи, которые явно обусловлены шумом. Возможно, если удалить эти значения, то точность значений окажется выше. Отсеивая промахи, получаем:





Где числа над графиками показывают долю оставшихся значений. А средние величины:






```
calculate A
4.7367836187436625
calculate B
0.689579973166721
calculate C
-1.9260362951633567
calculate D
-1.0214729239831604
```

Удаление грубых промахов не привело к предполагаемым результатам, которые были определены в прошлых методах.

Следовательно, для метод конечно-разностной аппроксимации требуется исследование на устойчивость, так как при наличии шума наблюдаются значительные отклонения найденных значений.

Исследование устойчивости

(Еще ищу информацию, пытаюсь вывести какие-то формулы)

Сравнение методов

- По времени работы программы:

Методы с использованием алгоритмов машинного обучения:

```
--- working hours ML:  5.025323152542114 second ---
```

Метод наименьших квадратов:

```
--- working hours OLS:  3.6307873725891113 second ---
```

Метод конечно-разностной аппроксимации

```
--- working hours FDA:  3.1627113819122314 second ---
```

Сравнение методов

- По точности определяемых параметров:

Методы с использованием алгоритмов машинного обучения

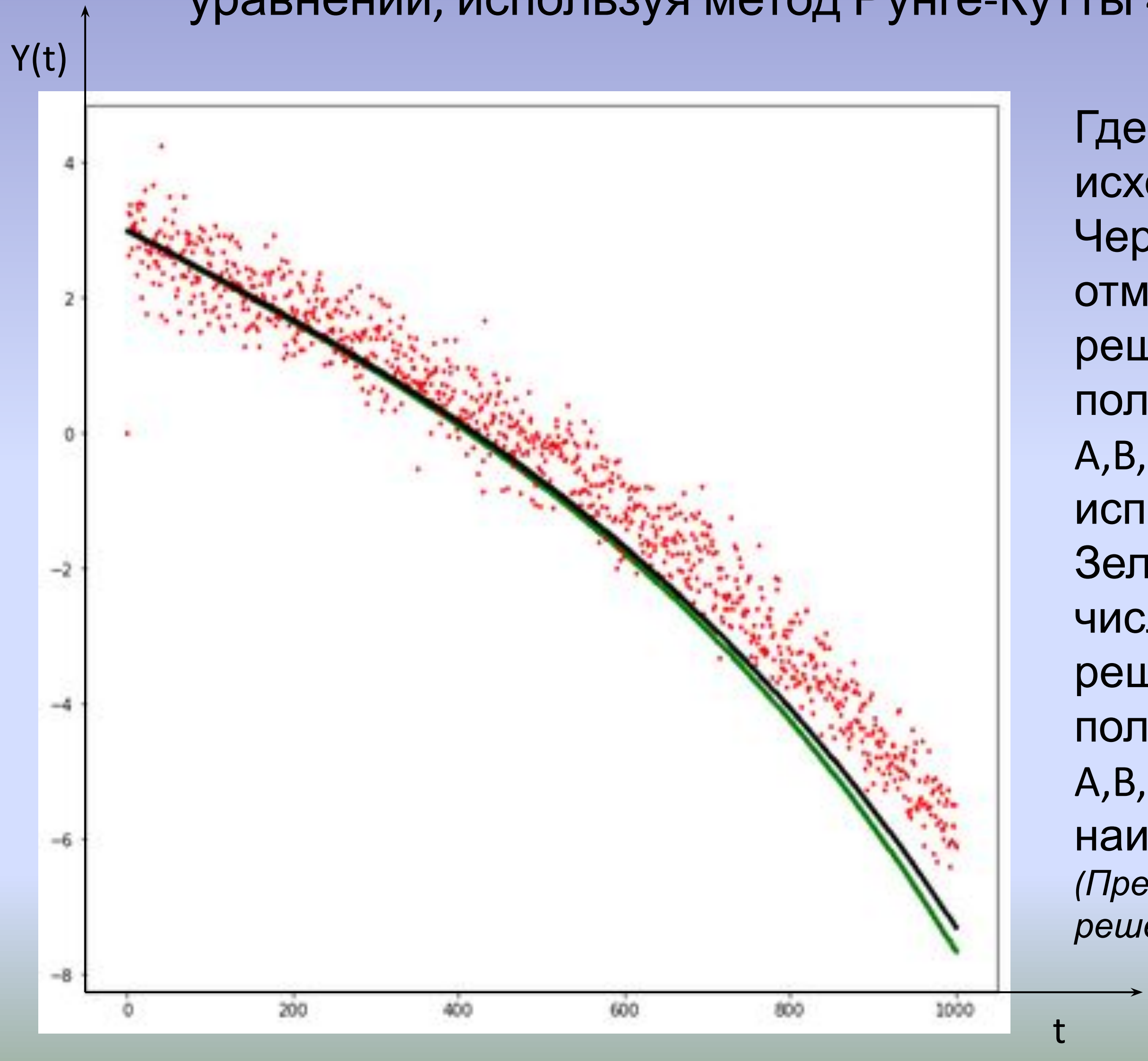
Метод наименьших квадратов

Метод конечно-разностной аппроксимации

(Предполагается, что мы рассматриваем найденные значения и «коридор», но я пока не могу подобрать формулировку)

Сравнение методов.

По численному решению полученной системы дифференциальных уравнений, используя метод Рунге-Кутты 4 порядка.



Где красным отмечены исходные данные y(t). Черным цветом отмечено численное решение с использование полученных параметров A,B,C,D из метода с использованием ML. Зеленым цветом отмечено численное решение с использование полученных параметров A,B,C,D из метода наименьших квадратов *(Предполагается еще кривая из решения К-Р аппроксимации)*