МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Лабораторная работа №5  
по дисциплине «Вычислительная математика»

Сплайн-аппроксимация и обработка экспериментальных данных

Вариант 13

Группа: АВТ-809  
Студент: Семёнов Б.В.  
Преподаватель: Балакин В.B.

НОВОСИБИРСК 2020

# Цели и задачи работы

Целью работы является изучение и приобретение практических навыков численного решения задач приближения (аппроксимации) функций аналитическим выражением, оценки достигаемой точности и необходимых ресурсов, применения изучаемых методов для обработки экспериментальных данных методами восстановления сеточных функций.

Задачи:

* изучение методов кусочно-глобальной аппроксимации сплайнами, оценок точности получаемых приближений, границ применимости;
* изучение и применение алгоритмов обработки экспериментальных данных и глобальной аппроксимации функций, заданных таблично, на основе метода наименьших квадратов, оценок точности получаемых приближений, границ применимости.

# Задание

Для заданной аналитической функции создать псевдоэмпирические исходные данные – «зашумленные» сетки значений y различным числом узлов (например, n1=10, n2=20). Результаты измерений содержат «экспериментальный шум» - случайные ошибки исследуемой величины y. Для формирования числовых наборов исходных данных использовать функции MathCAD генерации случайных величин, распределенных по равномерному и нормальному закону распределения с параметрами, определяющими погрешности соответствующих величин не более 10 процентов.

1. Ряд «экспериментальных» измерений функции построить в двух вариантах:

1) равноотстоящих точках;

2) в случайной последовательности с не равноотстоящими точками.

2. Сгладить результаты измерений, используя алгоритмы (по выбору студента) линейного и нелинейного сглаживания.

3. Сравнить полученные результаты созданных псевдоэмпирических данных с заданными значениями «незашумленной», истинной функции.

4. Составить план проведения вычислительного эксперимента, отвечающий задачам исследования, перечисленным ниже.

4.1. Исследование интерполирование сплайн-функциями.

4.1.2 Провести сравнение качества построения интерполяционного полинома различными методами (критерий - уклонение в узлах интерполяционной сетки и на заданном интервале). (Доп. задание: при построении полиномов методом неопределенных коэффициентов зафиксировать изменения значений числа обусловленности интерполяционной матрицы от порядка полинома.)

4.1.2 Провести исследование погрешности интерполирования для модельной функции в зависимости от свойств гладкости и монотонности на интервале интерполирования; выявить зависимость погрешности от порядка интерполяционного полинома, от величины интервала, от типа сетки и предварительной обработки «псевдоэкспериментальных» данных.

4.2. Исследование среднеквадратичного приближения функций.

4.2.1. Для заданной функции на интервале приближения решить задачу приближения полиномами 2-й, 3-й, 4-й и максимально достижимой степени; обратить внимание на изменение значения числа обусловленности матрицы нормальной системы уравнений от ее порядка; при двух различных порядках приближающего полинома исследовать зависимость среднеквадратичной погрешности решения от величины массива исходных данных, от порядка полинома и типа сетки.

4.2.2 Исследовать устойчивость решения задачи среднеквадратичного приближения к погрешности исходных данных. Критерием может служить отличие величин среднеквадратичной погрешности аппроксимации, полученных при наличии и отсутствии возмущений, вызываемых «зашумлением» данных.

Исходные данные

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

# Теоретические сведенья



## Приближение функции сглаживанием

**Суть процедуры сглаживания** состоит в подмене данной функции на каждом из рассматриваемых отрезков наилучшим линейным среднеквадратичным приближением.

На первом этапе. Для таблично заданной функции найти такую функцию S(x), составленную из линейных функций, чтобы для всех ***х*** в смысле минимума квадрата отклонений, т.е. . В результате решается задача нахождения коэффициентов ai, bi методом наименьших квадратов:

Второй этап состоит в пересчете данной таблицы, для

. Доопределим новую табличную функцию значениями и . В результате этого получаем новую табличную функцию, в которой сохраняется характер поведения исходной функции. Описанная процедура называется осреднением по трем точкам и является простым частным случаем линейного фильтра.

Некоторые задачи, возникающие при анализе и интерпретации опытных данных, не требуют построения единой аналитической формулы во всем диапазоне изменения переменной ***х***. Например, для численного дифференцирования (интегрирования) важно лишь устранить «шум» эксперимента, сохранив информацию об истинной функции. Для этой цели применяется сглаживание эмпирических данных, т. е. замена исходной таблицы опытных точек другой таблицей близких к ним точек, лежащих на достаточно гладкой кривой.

При сглаживании часто используется метод наименьших квадратов и аппроксимирующие многочлены различных степеней. Если используется многочлен первой степени, сглаживание называется *линейным*, в противном случае – *нелинейным*.

Количество точек для сглаживания берут нечетным, а группы точек – «скользящими» вдоль всей таблицы. Например, при линейном сглаживании по трем точкам последовательность действий такова. Сначала выбирают первые три точки , по которым находят линейный многочлен наилучшего среднеквадратичного приближения, вычисляют значение этого многочлена в средней точке и заменяют сглаженным значением . Затем берут следующую группу точек , и после соответствующих вычислений производят сглаживание среднего в данной группе значения , (т. е. заменяют значение на ) и т. д. до конца таблицы. После этого сглаживают две первых и две последних точки по особым (менее точным) формулам.

Линейное сглаживание по трем точкам приводит к формулам:

, ,

,

. (4)

## Интерполирование сплайнами

Одним из способов кусочно-глобальной аппроксимации на всем отрезке является интерполирование с помощью сплайн-функций. Сплайн-функцией или сплайном называют кусочно-полиномиальную функцию, опре­деленную на отрезке [a, b] и имеющую на этом отрезке некоторое число непрерывных производных.

Преимуществом сплайнов перед обычной интерполяцией явля­ется, во-первых, их сходимость и, во-вторых, устойчивость процесса вычислений. Рассмотрим частный случай (часто используемый на практике), когда сплайн определяется многочленом третьей степени

## Построение кубического сплайна

Пусть на отрезке  в узлах сетки заданы значения некоторой функции , т.е. , (i= 0,1,…, n)*.*

Сплайном, соответствующим этим узлам функции  называется функция *S(х)*, которая:

1) на каждом частичном отрезке является многочленом третьей степени;

1. функция  и ее первые две производные непрерывны на ;
2. .

На каждом частичном отрезке  будем искать сплайн , где  многочлен третьей степени

. (5)

То есть для  нужно построить такую функцию, где  подлежат определению. Для всего отрезка интерполирования , таким образом, необходимо определить ***4⋅ n*** неизвестных коэффициента.

Доопределим . Требование непрерывности функции *S(x)* приводит к условиям  (i=0, 1,…,n-1).

Отсюда из (5) получаем следующие уравнения:

(i= 1,2,…,n-1).

Введем шаг интерполирования . Тогда последнее равенство можно переписать в виде (i= 1,2,…,n). Из непрерывности первой производной следует (i= 2,3,…,n), а из непрерывности второй производной (i= 2,3,…,n).

Объединив все три вида уравнений, получим систему из 3n-2 уравнений относительно 3n неизвестных . Два недостающих уравнения получим, задав граничные условия для функции S(x). Для этого воспользуемся граничными условиями для сплайн-функции в виде (концы гибкой линейки свободны).

Тогда получим систему уравнений

(6)

Решая систему методом подстановки (исключаем из (6) неизвестные bi,di), получим систему:

(7)

(i= 1,2,…,n-1).

Система (7) имеет трехдиагональную матрицу. Эта система может быть решена методом прогонки или Гаусса. После ее решения коэффициенты сплайна определим через коэффициенты *сi* с помощью явных формул

, (i= 1,2,…,n).

## Сходимость процесса интерполирования кубическими сплайнами

Доказывается, что при неограниченном увеличении числа узлов на одном и том же отрезке . Оценка погрешности интерполяции зависит от выбора сетки и степени гладкости функции f(x).

При равномерной сетке  (i= 0,1,…,n)

где .

## Аппроксимация функцией методом наименьших квадратов

Для экспериментальных данных, представленных множеством {(***xi, yi***)} требуется подобрать вид аппроксимирующей зависимости *y=f(x),* связывающей переменные *х* и *у ,* и оценить её параметры*.* К такой задаче приходят при статистической обработке экспериментальных данных с помощью регрессионного анализа. Возможны следующие случаи:

во-первых - значения функции *f(x)* могут быть заданы в достаточно большом количестве узлов;

во-вторых - значения таблично заданной функции отягощены погрешностями. В этих случаях проводить приближения функции с помощью интерполяционных многочленов нецелесообразно, т.к.

- это неудобно делать, поскольку число узлов велико и пришлось бы строить несколько интерполяционных многочленов или один, но очень большого порядка;

- построив интерполяционные многочлены, мы повторили бы те же самые ошибки, которые присущи экспериментальным данным.

Поэтому, а также по ряду другим причинам, в том числе и вычислительного характера, приближающую функцию обычно ищут из следующих соображений:

1. приближающая функция не проходит через узлы сеточной эмпирической функции и не повторяет ошибки аргументов и значений функции;
2. чтобы она удовлетворяла бы одному из заданных критериев (2), (3), (4) близости аппроксимирующей к эмпирической сеточной функции.

Обычно для этих целей выбирают критерий минимума сумма квадратов отклонений приближающей функции от узловых значений эмпирических данных.

Пусть даны функции , назовем их базисными функциями. Будем искать приближающую (аппроксимирующую) функцию в виде линейной комбинации

. (8)

Такая аппроксимация называется линейной, а Фm(х) – обобщенный многочлен. Согласно критерию метода наименьших квадратов вычислим сумму квадратов отклонений таблично заданной функции от искомого многочлена в узлах:

. (9)

Степень обобщенного многочлена неизвестна. Она выбирается так, чтобы было наименьшим и удовлетворяла условиям:

- аппроксимирующая кривая не проходила бы через все узлы таблицы;

- получить приближение с заданной степенью точности.

Выражение можно рассматривать как функцию от неизвестных . Требуется найти, при каких значениях , значение будет минимально.

Из условием существования экстремума , частные производные от по всем переменным приравниваются нулю. Таким образом получается система уравнений вида:

(10)

Система (10) - система линейных уравнений относительно. Для компактной записи системы (10) можно применить определение скалярного произведения функций. (**Определение.** Скалярным произведением функции f на g на множестве точек называется выражение ). Тогда систему (10) можно записать в виде:

(10а)

Системы (10) и (10а) принято называть нормальными системами уравнений метода наименьших квадратов.

Решения этих СЛАУ являются искомые коэффициенты , и следовательно, аппроксимирующий многочлен будет полностью определен. Это возможно обычными методами решения нормальной системы уравнений при соблюдении двух основных условий: узлы сеточных эмпирических функций не равноотстоящие и базисные функции аппроксимирующего многочлена линейно не зависимы.

Степень аппроксимирующего многочлена ‘’**m**’’ обычно выбирается методом последовательного приближения. В случае, когда **m = n** становится равным нулю и мы получим интерполяционный многочлен со всеми его недостатками для приближения эмпирических данных. Для того чтобы избежать этого, достаточно выбирать m из условия **m<n** и задавать диапазон изменения числами *ε1*и *ε2,* учитывая следующее:

1. *ε1*>0 и *ε2*>0 должны быть такими, чтобы находилось между ними;
2. первоначально **m** выбирают произвольно, но учитывая условие, что **m<<n**;
3. выбрав **m**, строят системы (10) и (10a), решив которые находят ;
4. используя найденные коэффициенты вычисляется и проверяется, попала ли она в промежуток между *ε1* и *ε2*. Если попала, то степень многочлена выбрана правильно, иначе

а) если > *ε1*, то степень необходимо уменьшить хотя бы на единицу;

б) если <*ε2*, то степень необходимо увеличить хотя бы на единицу.

1. затем строить приближающую функцию.

Очень часто для приближения по методу наименьших квадратов используются алгебраические многочлены степени **m < n**, т.е. . Тогда нормальная система (10) принимает следующий вид:

(k= 0,1,…,m). (11)

Запишем систему (11) в развернутом виде в двух наиболее простых случаях m=1 и m=2. В случае многочлена первой степени *P1(x)=c0+c1x*, нормальная система имеет вид

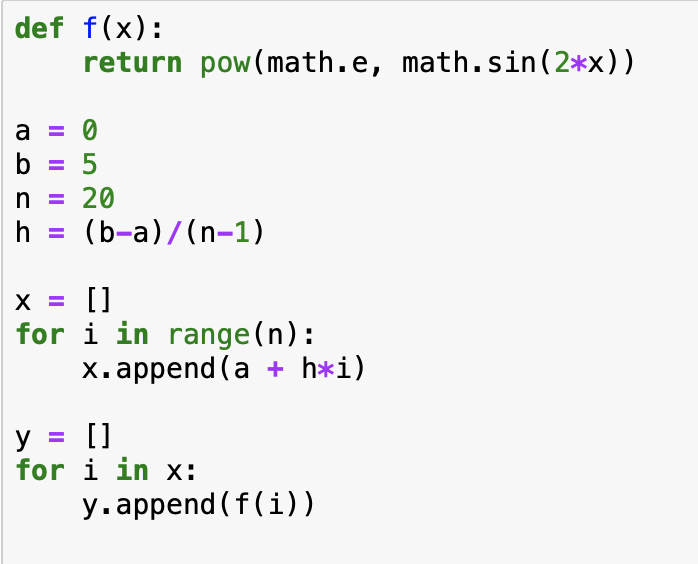
(12)

Для многочлена второй степени *P2(x)=c0+c1x+c2x2*, нормальная система имеет вид

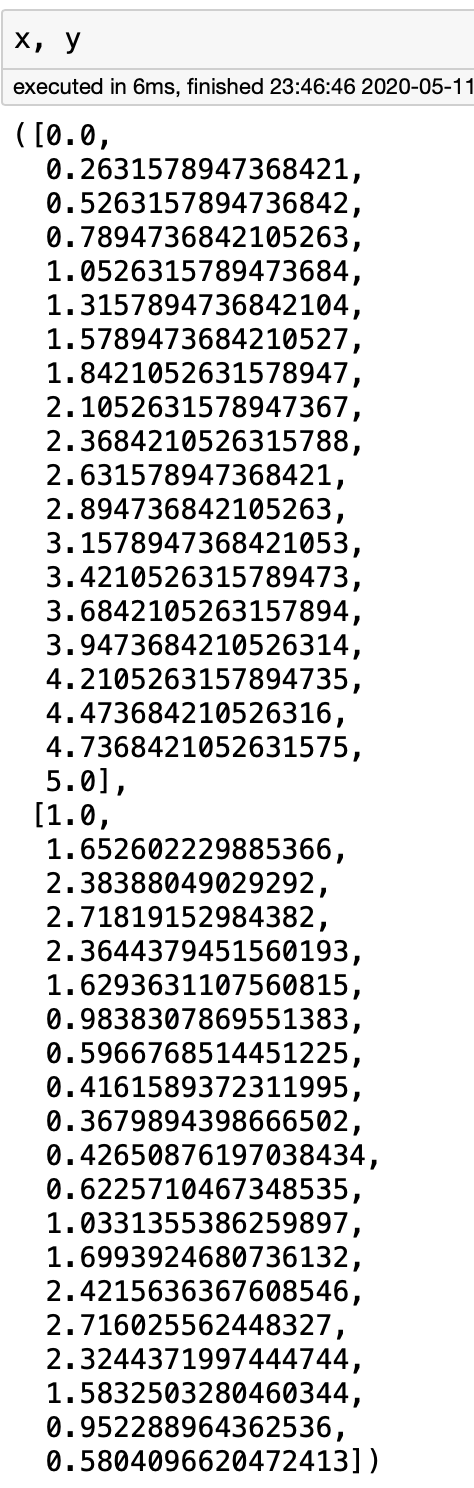
(13)

# Работа с вектором псевдоэмпирических данных

## Формирование вектора псевдоэмпирических данных и данных с «зашумлением»



Псевдоэмперические данные:



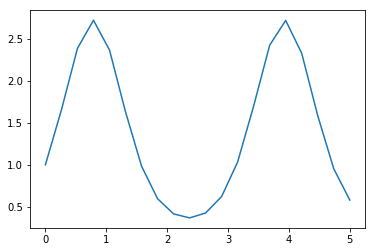
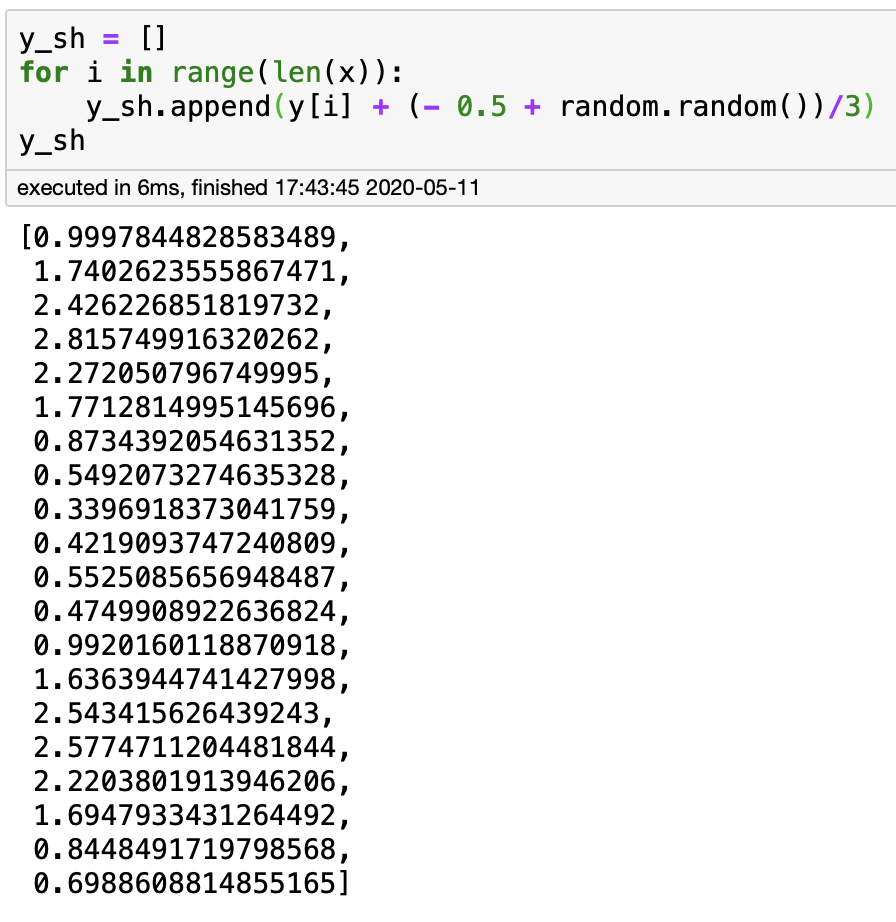


Рис. 1. График исходных данных

Зашумленные псевдоэмперические данные:



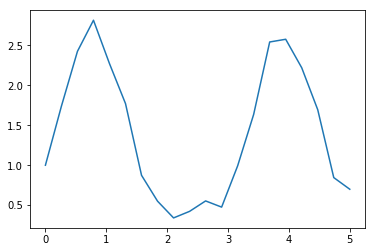
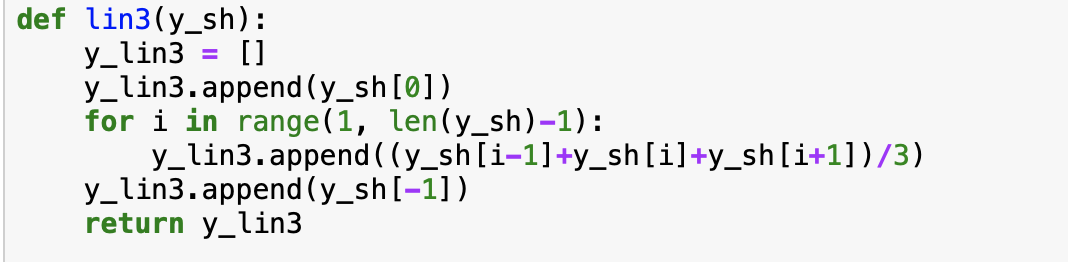


Рис. 2. График зашумленных исходных данных

Теперь произведём сглаживание тремя алгоритмами: линейным сглаживанием по трём и пяти точкам, а также нелинейным сглаживанием по семи точкам.

# Ход работы

## Сглаживание по 3 точкам



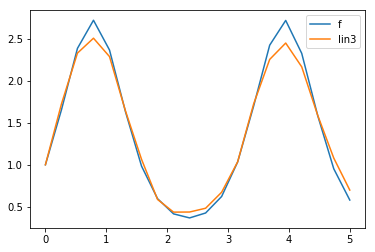
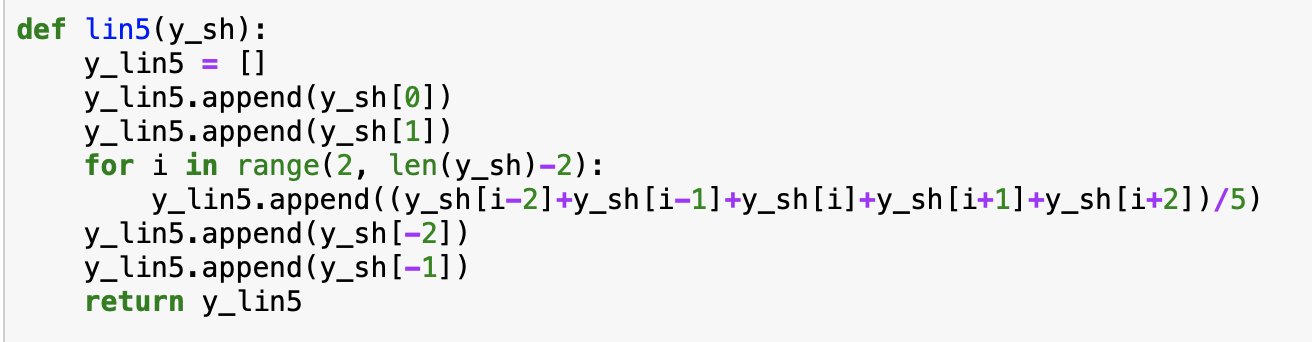


Рис. 3. Линейное сглаживание по 3-м точкам

## Сглаживание по 5 точкам



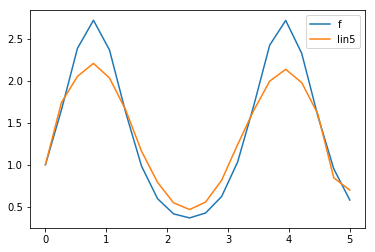
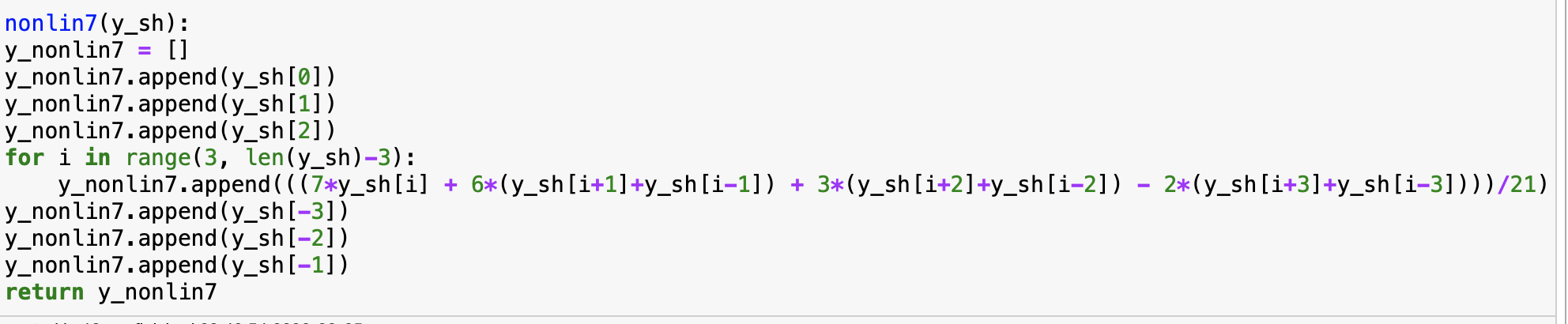


Рис. 4. Линейное сглаживание по 5 точкам

## Нелинейное сглаживание по 7 точкам



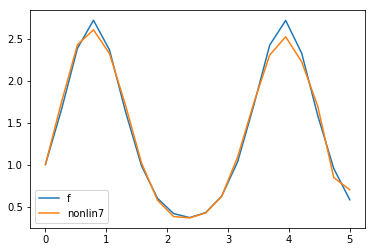
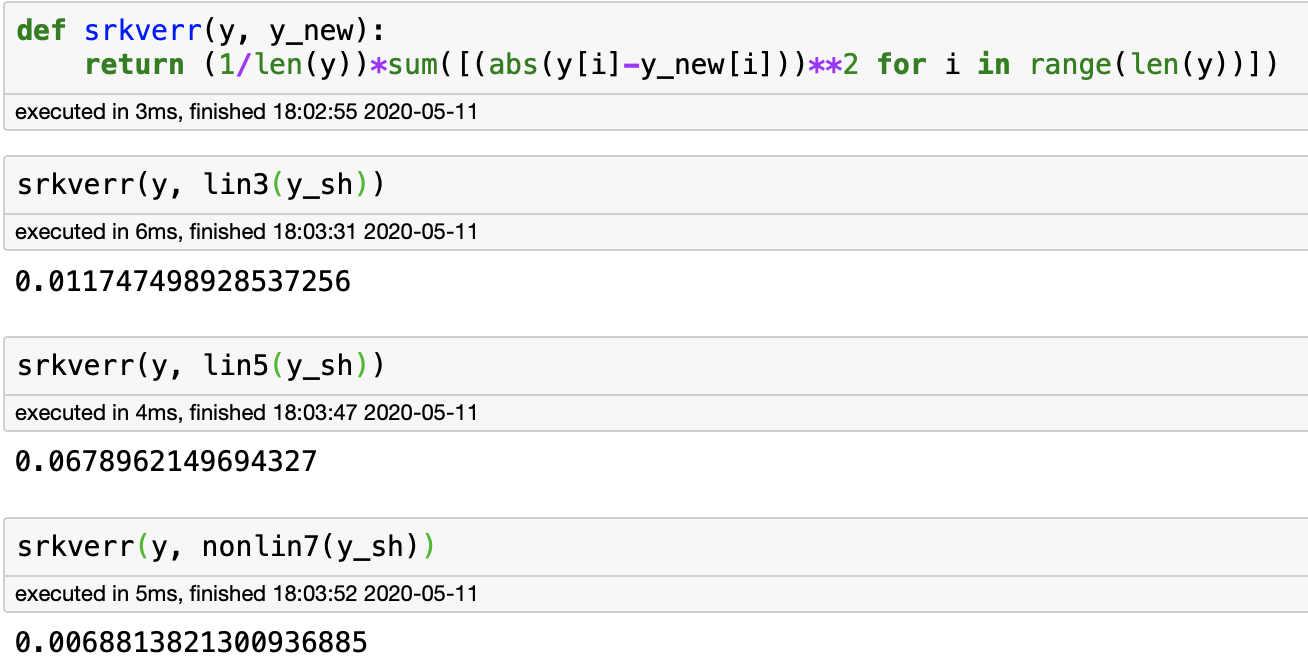


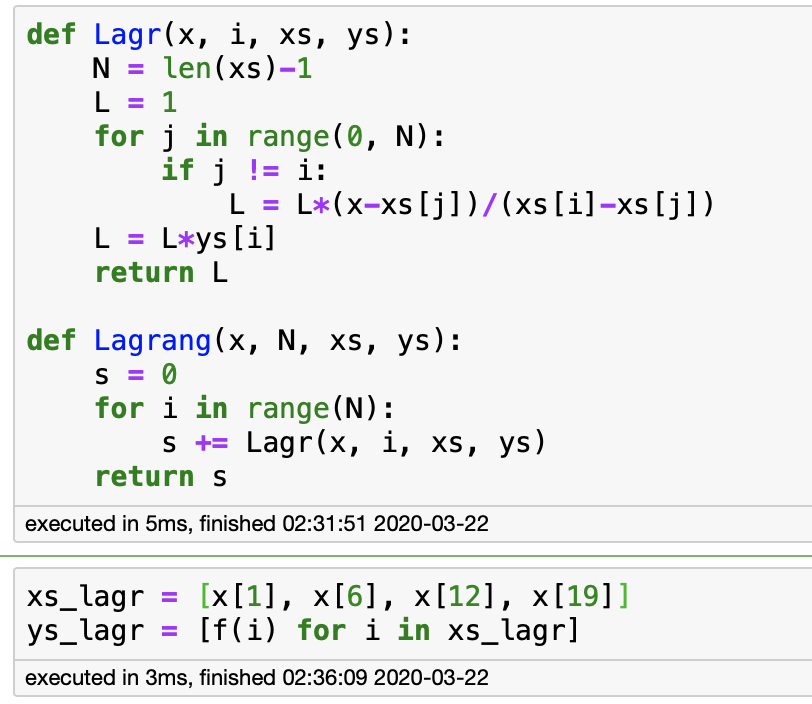
Рис. 5. Нелинейное сглаживание по 7 точкам

Теперь вычислим среднеквадратичную ошибку для каждого из методов:



Заметно, что линейное сглаживание по 5 точкам показало себя хуже, чем линейное по 3 и нелинейное по 7.

## Интерполяция методом Лагранжа



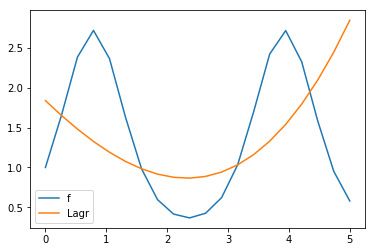
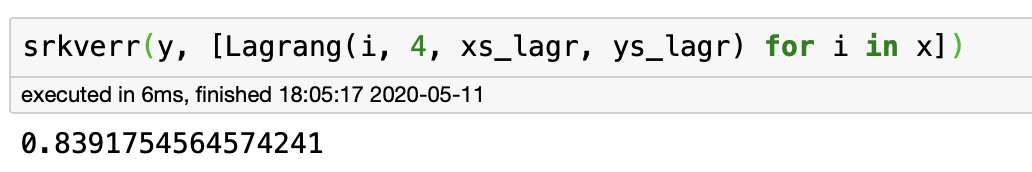


Рис. 6. Интерполяция методом Лагранжа

СКО равно:



Возможно, стоит использовать большее кол-во точек, распределенных по всей кривой для лучшего результата.

## Интерполяция кубическими сплайнами

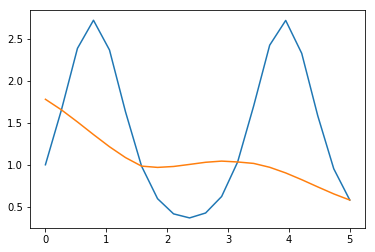
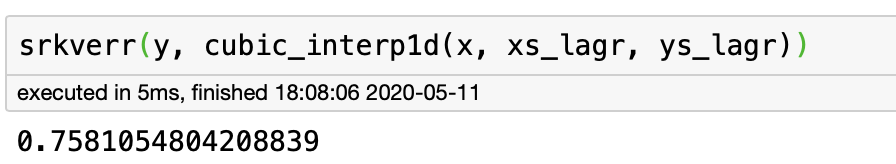


Рис. 7. Интерполяция кубическими сплайнами



Так же как и с методом Лагранжа стоит использовать большее кол-во узлов.

## Интерполяция по методу наименьших квадратов

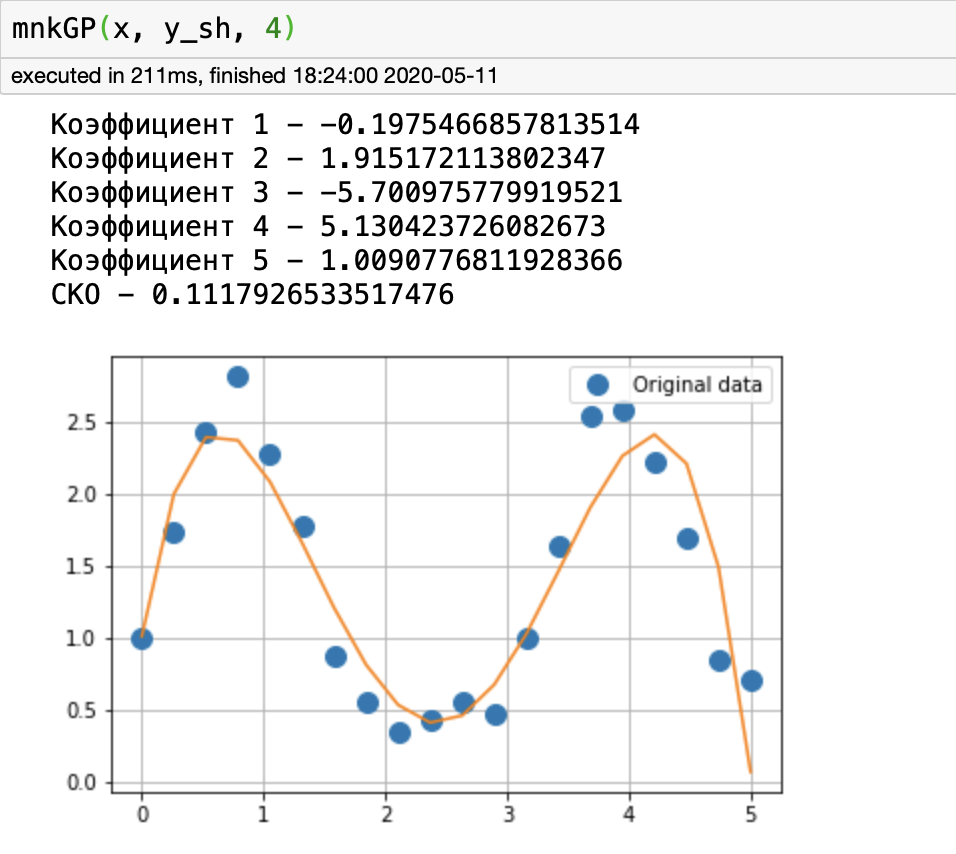


Рис. 8. Интерполяция по методу наименьших квадратов и полиномом четвертой степени

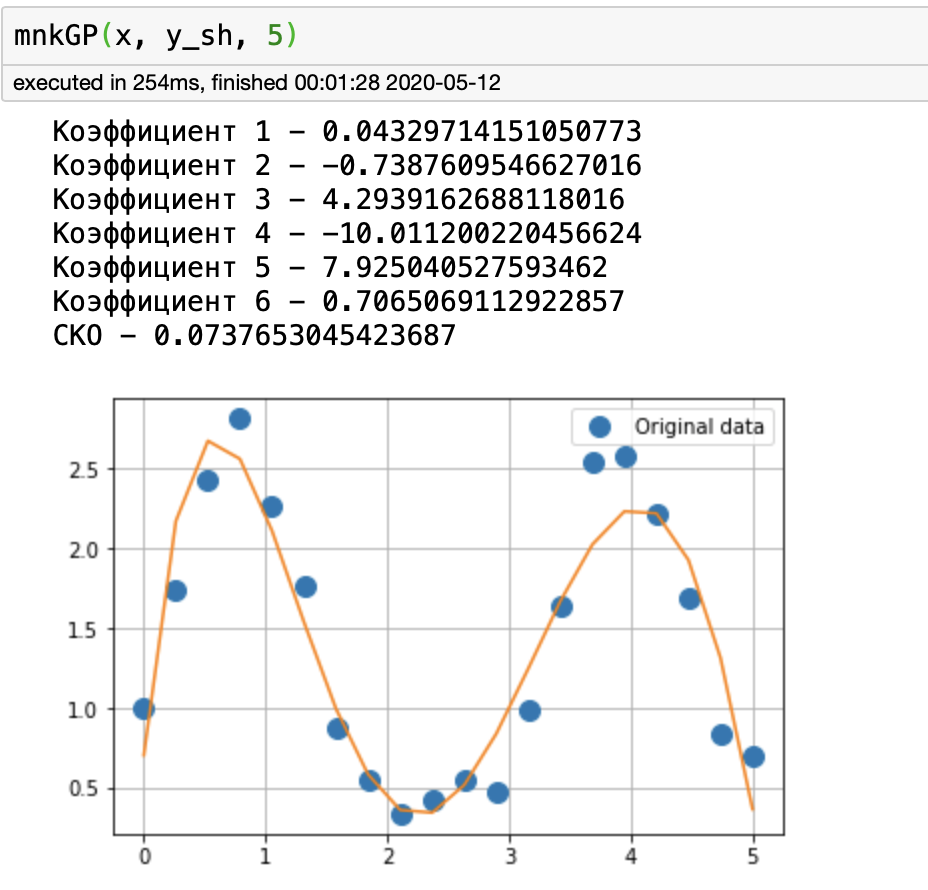


Рис. 9. Интерполяция по методу наименьших квадратов и полиномом пятой степени

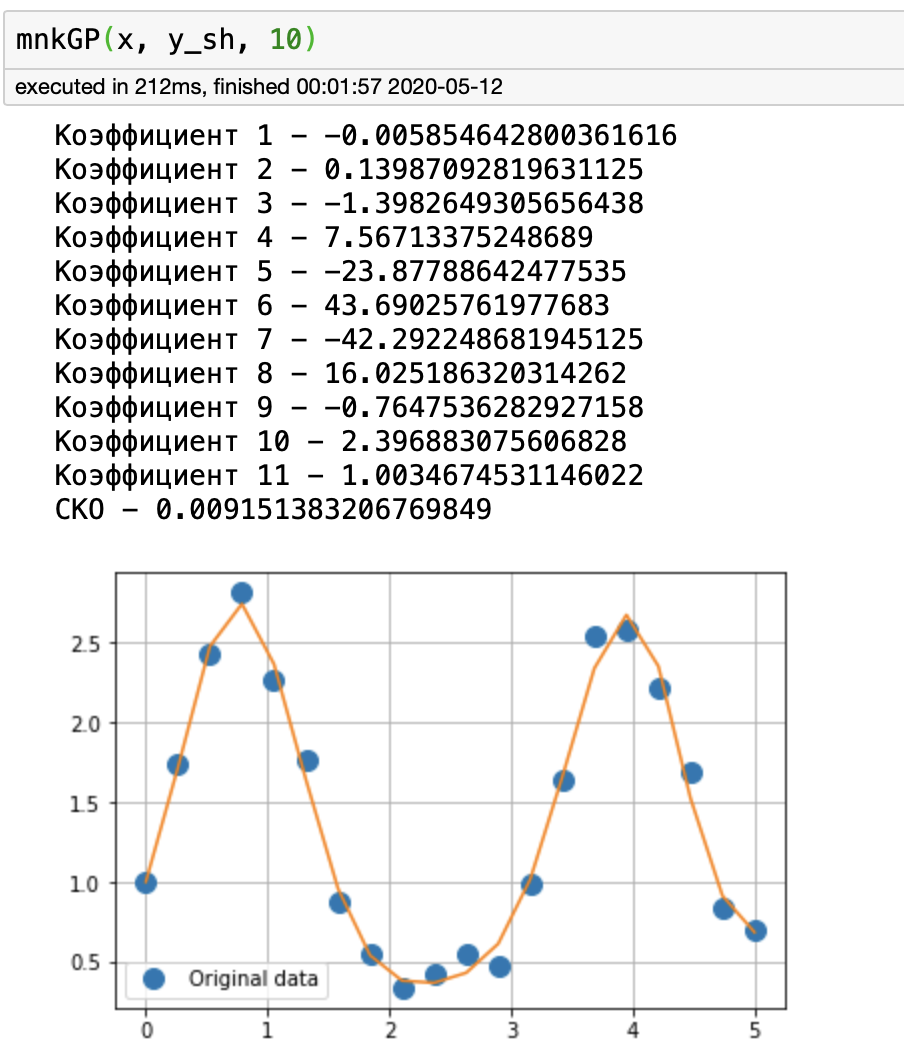


Рис. 9. Интерполяция по методу наименьших квадратов и полиномом десятой степени

# Вывод

В ходе выполнения лабораторной работы для заданной аналитической функции были созданы псевдо-эмпирические исходные данные – «зашумленные» сетки значений y различным числом узлов. Результаты измерений содержат «экспериментальный шум» - случайные ошибки исследуемой величины y. Для формирования числовых наборов исходных данных использовались функции Python для генерации случайных величин.

Далее было установлено, что наиболее точным из кусочно-линейных и нелинейных оказалось нелинейное сглаживание по семи точкам. Данное приближение было выбрано в качестве результата сглаживания зашумлённой функции. Для вычислений значения полинома Лагранжа не требуется предварительно определять коэффициенты полинома путем решения системы уравнений. Однако этот метод выдал совершенно неудовлетворительный результат.

Кубические сплайны дают более-менее близкий к истинному результат, но не более того.

Метод наименьших квадратов в нашем случае является самым лучшим, так как можно добавлять степени полиному до бесконечности и получать довольно точные результаты.

В общем можно заключить, что, во-первых, четырёх точек недостаточно для интерполяции даже весьма гладкого отрезка функции, а во-вторых, даже при этом условии все методы дали довольно посредственные результаты.