



Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Робототехники и комплексной автоматизации

КАФЕДРА Системы автоматизированного проектирования (РК-6)

ОТЧЕТ О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

Студент Никифорова Ирина Андреевна

Группа РК6-61б

Тип задания лабораторная работа

Тема лабораторной работы Численное интегрирование с помощью
тригонометрических интерполянтов.
Полиномиальная регрессия.

Студент _____ **Никифорова И. А.**
подпись, дата фамилия, и.о.

Преподаватель _____ **Соколов А. П.**
подпись, дата фамилия, и.о.

Преподаватель _____ **Першин А. Ю.**
подпись, дата фамилия, и.о.

Оценка _____

Москва, 2019 г.

Оглавление

Задание на лабораторную работу	2
Цель выполнения лабораторной работы	5
Задачи, выполненные в процессе реализации лабораторной работы	5
Численное интегрирование с помощью тригонометрических интерполянтов	6
Полиномиальная регрессия	10
Заключение	15
Список использованных источников	16

Задание на лабораторную работу

Задание 1:

Даны интегралы:

$$I_1 = \int_{-\frac{\pi}{4}}^{\pi} |x| dx, \quad (1)$$

$$I_2 = \int_{-\frac{\pi}{4}}^{\pi} (x \cos x^2 + e^x \cos e^x) dx \quad (2)$$

Требуется:

1. Вывести общее выражение для формулы численного интегрирования путем аналитического интегрирования тригонометрического ряда, заменяющего подынтегральную функцию.
2. Используя алгоритм Кули–Тьюки, написать функцию *fft_coeff(y_nodes)*, которая вычисляет и возвращает комплексные коэффициенты тригонометрического полинома, интерполирующего узлы *y_nodes*, равномерно распределенные на отрезке $[-\pi; \pi]$.
3. Написать функцию *spectral_integral(f, N)*, которая вычисляет значение интеграла функции *f*, интерполируемой в *N* узлах с помощью тригонометрического ряда, на интервале $[-\frac{\pi}{4}; \pi]$. Функция *spectral_integral* должна использовать внутри себя функцию *fft_coeff*.
4. Для каждого из интегралов I_1 и I_2 провести следующий анализ:
 - Найти точное значение интеграла.
 - Найти приближенное значение интеграла с помощью функции *spectral_integral* для $N = 2^{\tilde{n}}$, $\tilde{n} \in 1, \dots, 8$.
 - Для каждого *N* найти относительную погрешность вычислений δ и вывести на экран график зависимости δ от *N*, где δ следует отображать в логарифмической шкале
5. Объяснить, как можно использовать полученные логарифмические графики для оценки порядка точности интегрирования.

6. Ответить, различаются ли порядки точности интегрирования в случае вычисления интегралов I_1 и I_2 и, если различаются, объяснить, с чем это связано.

Задание 2:

Дана функция

$$f(x) = -10x^2 + 1.5x + 1 + \sigma X,$$

где $x \in [-1;1]$ и X - случайная величина, нормально распределенная на интервале $[-1;1]$.

Требуется:

1. Написать функцию *poly_regression*(*x_nodes*, *y_nodes*, *degree*), которая возвращает коэффициенты многочлена степени *degree*, наилучшим образом приближающегося к точкам с абсциссами *x_nodes* и ординатами *y_nodes*.

2. Для каждого N из множества $\{2^3, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9\}$ с помощью функции $f(x)$ сгенерировать начальный набор данных $D_{regr}^{(N)}$ и проверочный набор данных $D_{test}^{(N)}$, где N - число точек в наборе данных. Затем, для каждого N , каждого σ из множества $\{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$, каждого p из множества $\{1, 2, 3, 4, 5\}$ провести следующий анализ:

- С помощью набора данных $D_{regr}^{(N)}$ и функции *poly_regression* построить многочлен степени p , наилучшим образом приближающийся к данным.
- Вычислить среднеквадратичную погрешность аппроксимации $\varepsilon_{regr}^{(N,p)}$ данных $D_{regr}^{(N)}$ полученным многочленом.
- Вычислить среднеквадратичную погрешность аппроксимации $\varepsilon_{test}^{(N,p)}$ данных $D_{test}^{(N)}$ полученным многочленом.

3. Вывести на экран несколько характерных примеров графиков многочленов вместе с начальными и проверочными данными.

4. Ответить на следующие вопросы:

- Как влияет увеличение числа начальных и проверочных данных на $\varepsilon_{regr}^{(N,p)}$ и $\varepsilon_{test}^{(N,p)}$? Выведите на экран графики, иллюстрирующие ответ, и сделайте вывод.
 - Как влияет увеличение степени многочлена p на $\varepsilon_{test}^{(N,p)}$ и $\varepsilon_{regr}^{(N,p)}$? Выведите на экран графики, иллюстрирующие ответ, и сделайте вывод.
 - Как влияет увеличение числа начальных и проверочных данных на относительную погрешность коэффициентов многочлена при сравнении с $f(x)$ в случае фиксированного σ ? Что происходит при увеличении σ ? Выведите на экран графики, иллюстрирующие ответ, и сделайте вывод. Как может повлиять изменение функции распределения случайной величины X на сделанный вывод?
5. Исходя из ответов на предыдущие вопросы, сделать общий вывод о свойствах сходимости полиномиальной регрессии в случае зашумленных данных.

Цель выполнения лабораторной работы

Цель выполнения лабораторной работы – вывести общее выражение для формулы численного интегрирования, основанное на интерполяции подынтегральной функции тригонометрическими полиномами. Запрограммировать его вычисление, основанное на быстром преобразовании Фурье, реализованном с помощью метода Кули-Тьюки. Проверить точность полученных решений.

Запрограммировать нахождение полиномиальной регрессии для зашумленной функции и проверить, насколько уровень шума и количество точек интерполяции влияет на качество приближения.

Задачи, выполненные в процессе реализации лабораторной работы

1. Задание 1. Численное интегрирование с помощью тригонометрических интерполянтов.
2. Задание 2. Полиномиальная регрессия.

1. Численное интегрирование с помощью тригонометрических интерполянтов

Было получено общее выражение, далее оно обозначается как I , для формулы численного интегрирования путем аналитического интегрирования тригонометрического ряда, заменяющего подынтегральную функцию. Для этого некоторая неизвестная функция $f(x)$ была представлена в виде тригонометрического ряда и занесена под знак интегрирования. [2]

$$I = \int_a^b a_0 + \sum_{k=1}^n (a_k \cos(kx) + b_k \cos(kx)) dx \quad (3)$$

Далее интеграл суммы (3) был представлен в виде суммы интегралов, каждый из которых был найден отдельно.

$$I = \int_a^b a_0 dx + \sum_{k=1}^n \left[\int_a^b a_k \cos(kx) dx + \int_a^b b_k \cos(kx) dx \right] \quad (4)$$

$$\int_a^b a_0 dx = a_0(b - a) \quad (5)$$

$$\int_a^b a_k \cos(kx) dx = \frac{a_k}{k} (\sin(kb) - \sin(ka)) \quad (6)$$

$$\int_a^b b_k \cos(kx) dx = -\frac{b_k}{k} (\cos(kb) - \cos(ka)) \quad (7)$$

После объединения выражений (5), (6) и (7) была получена формула для итогового значения I .

$$I = a_0(b - a) + \sum_{k=1}^n \left[\frac{a_k}{k} (\sin(kb) - \sin(ka)) - \frac{b_k}{k} (\cos(kb) - \cos(ka)) \right] \quad (8)$$

Также была получена частная формула при $a = -\frac{\pi}{4}$ и $b = \pi$.

$$I = \frac{5\pi}{4} a_0 + \sum_{k=1}^n \left[-\frac{a_k}{k} \sin\left(-\frac{\pi}{4}k\right) - \frac{b_k}{k} (\cos(k\pi) - \cos\left(-\frac{\pi}{4}k\right)) \right] \quad (9)$$

Каждый из коэффициентов a_k , для $k \in [0; n]$ и b_k , для $k \in [1; n]$ (b_0 может иметь любое значение в силу того, что $\sin(0) = 0$) можно определить в зависимости от комплексных коэффициентов тригонометрического ряда.

$$a_0 = \hat{a}_0 \quad (10)$$

$$a_k = 2\Re(\hat{a}_k), k = 1, \dots, n \quad (11)$$

$$b_k = -2\Im(\hat{a}_k), k = 1, \dots, n \quad (12)$$

В формулах (10), (11), (12) коэффициенты \hat{a}_k для $k \in [0; n]$ являются комплексными коэффициентами тригонометрического ряда.

$$\hat{a}_k = \frac{(-1)^k}{2 \cdot m} \sum_{j=0}^{2m-1} y_j e^{-\frac{ikj\pi}{m}} \quad (13)$$

В результате подстановки значений (10), (11) и (12) в формулу (9) и преобразований знаков было получено выражение для интеграла в зависимости от комплексных коэффициентов.

$$I = \frac{5\pi}{4} \hat{a}_0 + \sum_{k=1}^n \left[\frac{2\Re(\hat{a}_k)}{k} \sin\left(\frac{\pi}{4}k\right) + \frac{2\Im(\hat{a}_k)}{k} (\cos(k\pi) - \cos(\frac{\pi}{4}k)) \right] \quad (14)$$

Для получения комплексных коэффициентов была написана функция *fft_coeff(y_nodes)*. Расчет коэффициентов был произведен по формуле (13). Он основан на быстром преобразовании Фурье, реализованном с помощью метода Кули-Тьюки.

Использование быстрого преобразования Фурье позволяет решить поставленную задачу за время $O(m \log_2 m)$, где m - половина размерности массива *y_nodes*, вместо $O(m)$ для решения с помощью дискретного преобразования Фурье. [2]

Алгоритм Кули-Тьюки основан на разделении исходного массива на два новых - с четными и нечетными индексами - и вычислении каждого из них рекурсивно по такому же принципу.

Была написана функция *spectral_integral*, реализующая вычисление формулы (14). Для проверки точности написанной функции была написана функция *integral_analysis*, где были найдены точные значения интегралов I_1 и I_2 по выражениям их аналитических решений.

$$I_1 = \int_{-\frac{\pi}{4}}^{\pi} |x| dx = \frac{x|x|}{2} \Big|_{-\frac{\pi}{4}}^{\pi} = \frac{17 \cdot \pi^2}{32} \quad (15)$$

$$\begin{aligned} I_2 &= \int_{-\frac{\pi}{4}}^{\pi} (x \cos x^2 + e^x \cos e^x) dx = \\ &= \frac{1}{2}(\sin(\pi^2) - \sin(\frac{\pi^2}{16})) + \sin(e^{\pi}) - \sin(e^{-\frac{1}{4}}) \end{aligned} \quad (16)$$

Были также вычислены значения функций *spectral_integral* для количества узлов $N \in \{2^1, 2^2, \dots, 2^8\}$ и для каждого N была найдена относительная погрешность измерения. На основе этих данных был построен график, изображенный на рисунке 1.

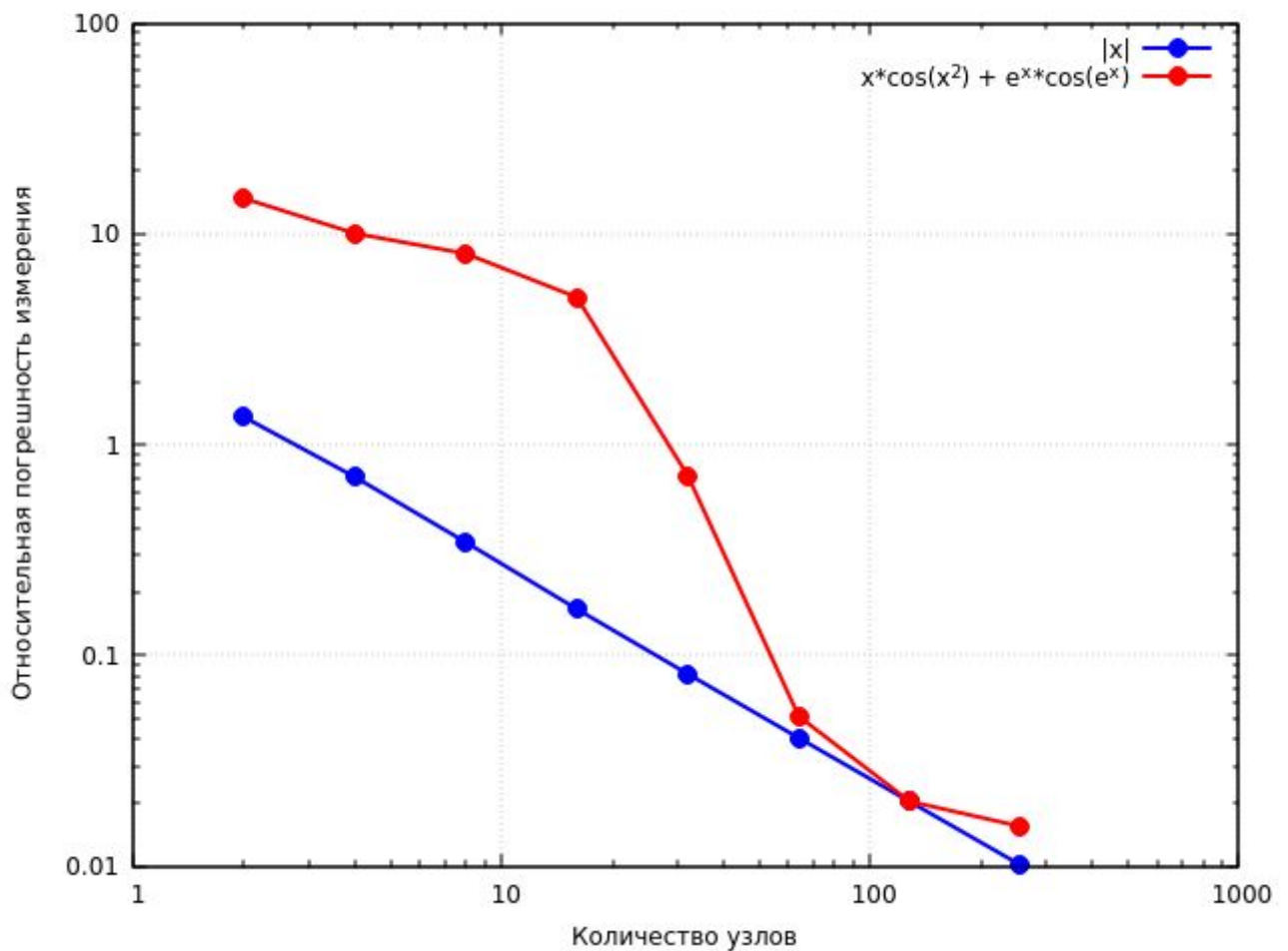


Рис. 1. График зависимости относительной погрешности интегрирования от количества узлов интерполяции для интегралов, представленных в легенде функций

Как видно на рисунке 1, зависимость погрешности интегрирования для функции $|x|$ представляет из себя линию, близкую к прямой, в логарифмических

координатах. Это означает, что при увеличении количества узлов в n раз, погрешность уменьшится приблизительно в такое же число раз, что указывает на порядок точности, приблизительно равный единице.

Похожая ситуация наблюдается и на втором графике - с каждым увеличением количества узлов, погрешность значительно убывает, хотя и не всегда равномерно для данной функции.

2. Полиномиальная регрессия

Для нахождения коэффициентов полинома, приближающего исходную функцию для заданных точек, была написана функция *poly_regression(x_nodes, y_nodes, degree)*. Вычисления для данной функции основываются на решении нормального уравнения для случая полиномиальной регрессии. [2]

$$a = (X^T X)^{-1} X^T y \quad (17)$$

$$X = \begin{vmatrix} 1 & x_{(1)} & \dots & x_{(1)}^n \\ 1 & x_{(2)} & \dots & x_{(2)}^n \\ \dots & \dots & \dots & \dots \\ 1 & x_{(m)} & \dots & x_{(m)}^n \end{vmatrix} \quad (18)$$

В данной формуле y - вектор значений ординат опорных точек, a - вектор коэффициентов полинома, X - матрица, имеющая вид, представленный в формуле (18), где x - элемент вектора абсцисс опорных точек.

При решении уравнения (17) был получен вектор коэффициентов полинома, наилучшим образом приближающего данную функцию.

$$f(x) = -10x^2 + 1.5x + 1 + \sigma X, \quad (17)$$

В качестве исходной функции была написана реализация формулы (17), где величина X представляет собой шум, нормально распределенный на участке рассмотрения функции $[-1; 1]$, а величина σ регулирует уровень шума.

Для проверки зависимости точности приближения от количества точек, уровня шума в исходной функции и степени полинома, была написана функция *analysis*. В ней с помощью функций *numpy* были созданы два набора данных: $D_{regr}^{(N)}$ - начальный, и $D_{test}^{(N)}$ - проверочный. Начальный набор данных необходим для расчета по нему полинома, наилучшим образом (по квадратичной погрешности) приближающего исходную функцию. Тестовый набор не

используется для расчета полинома, но его значения также принадлежат значениям функции на заданном интервале. Он используется для проверки того, насколько точно полином приближает не известные ему данные.

Для начального набора с помощью написанной функции *poly_regression* был определен полином, приближающий исходную функцию по данному набору точек наилучшим образом. После этого был проведен анализ: число точек N принимало каждое значение из множества $\{2^3, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9\}$, уровень шума σ - из множества $\{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$, p - из множества $\{1, 2, 3, 4, 5\}$. В результате описанного варьирования были получены характерные графики и выявлены некоторые особенности.

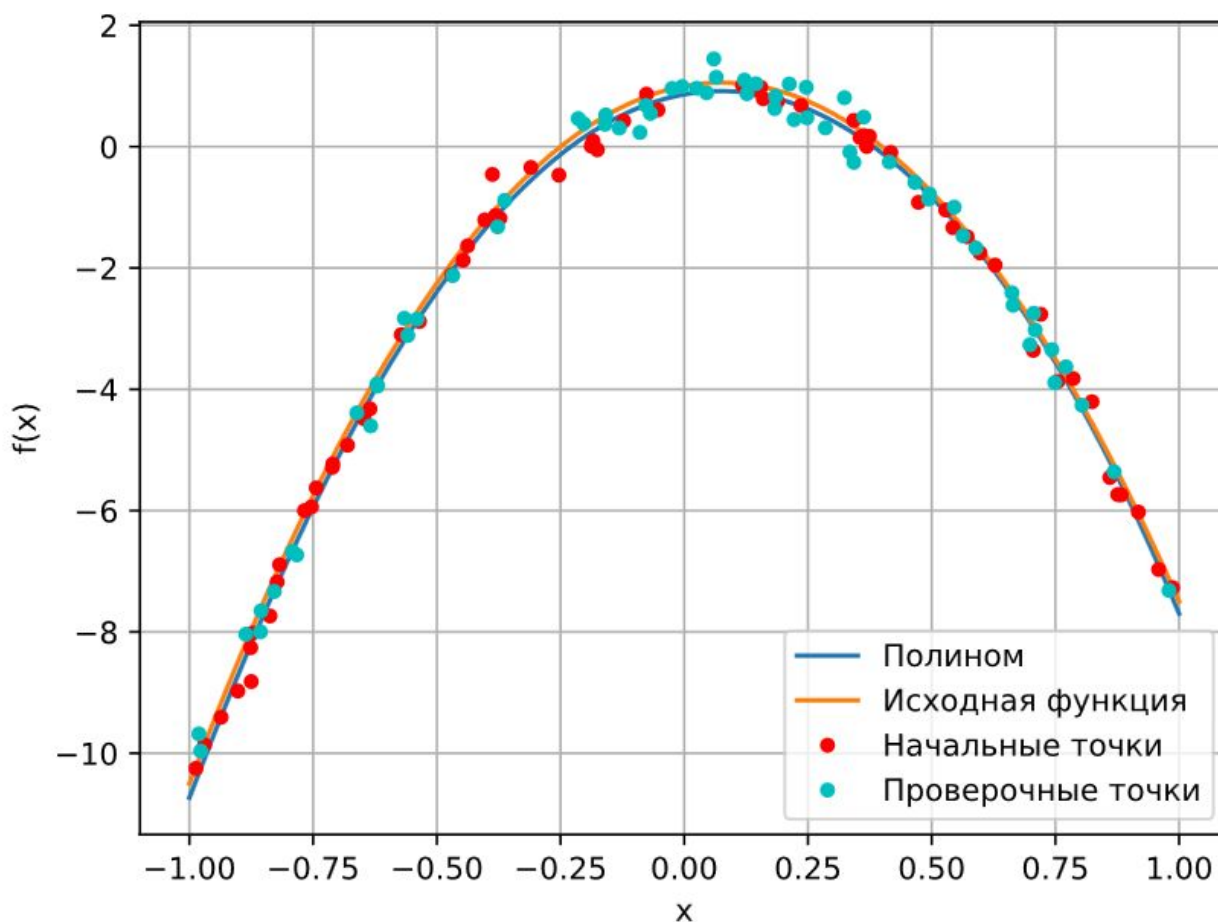


Рис. 2. Графики при $N = 64$, $\sigma = 10^{-1}$, $p = 4$

Для примера можно рассмотреть достаточно большое число узлов, средний уровень шума и высокую степень полинома - $N = 64$, $\sigma = 10^{-1}$, $p = 4$. На рисунке 2 можно увидеть, что при данных параметрах полином приближает

исходную функцию (она изображена без искажения шумом) с достаточно высокой точностью. Для проверки была подсчитана среднеквадратичная погрешность для начальной выборки - она составила 0.02408 - и для тестовой выборки - 0.02435. Разница между ними составляет всего 0.00027, что говорит о том, что тестовая выборка была приближена почти так же точно, как и начальная, на которой основывался расчет полинома.

Для определения вклада каждого параметра были проанализированы графики с варьированием только одного из параметров. Было выяснено, что при уменьшении числа узлов, точность приближения значительно сокращается. В таблице 1 приведены среднеквадратичные погрешности приближений в узлах тестовой выборки для различного числа узлов на примере фиксированных $\sigma = 0.01$, $p = 3$. Здесь видно, что с увеличением числа узлов уменьшается погрешность приближения. Для иллюстрации приведены рисунки (3) и (4), демонстрирующие приближение при минимальном и максимальном количестве узлов для $\sigma = 0.01$, $p = 5$.

Таблица 1. Зависимость погрешности от числа узлов интерполяции при $\sigma = 0.01$, $p = 3$

Число узлов интерполяции, N	Среднеквадратичная погрешность тестовой выборки, $\varepsilon_{test}^{(N,p)}$
8	0.00659
16	0.00630
32	0.00472
64	0.00250
128	0.00183
256	0.00117
512	0.00095

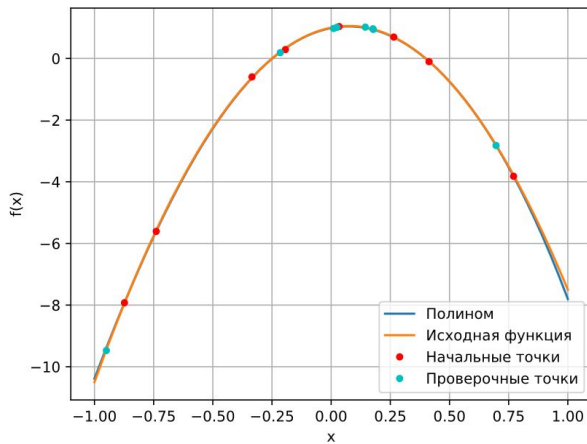


Рис. 3. $\sigma = 0.01$, $p = 5$, $N = 8$

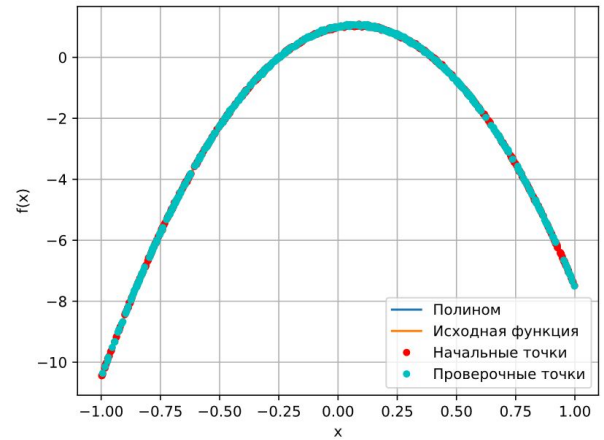


Рис. 4. $\sigma = 0.01$, $p = 5$, $N = 512$

Стоит отметить, что описанное выше правило работает даже в случае большой зашумленности данных. Это происходит в силу того, что шум распределен нормально - он располагается чаще близко к функции, чем далеко от нее. Это подтверждают графики на рисунках 5 и 6. На данных рисунках можно отметить, что погрешность уменьшается с увеличением количества узлов. Также можно подтвердить данное замечание численно: $\varepsilon_{test}^{(64, 5)} = 2.557$ и $\varepsilon_{regr}^{(64, 5)} = 2.776$, а $\varepsilon_{test}^{(512, 5)} = 0.862$ и $\varepsilon_{regr}^{(512, 5)} = 0.853$.

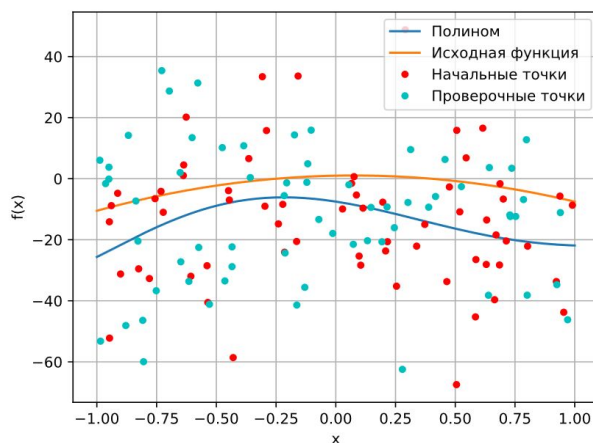


Рис. 3. $\sigma = 10$, $p = 5$, $N = 64$

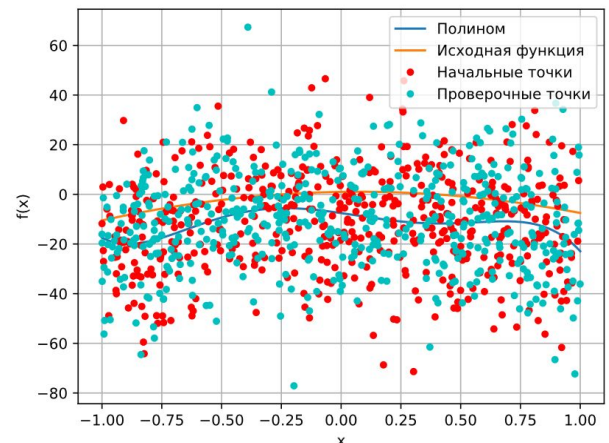


Рис. 4. $\sigma = 10$, $p = 5$, $N = 512$

Степень полинома также очень сильно влияет на погрешность приближения. Наилучшим образом приближает полином той же степени, что и исходная функция. При использовании полиномов высоких степеней и небольшого количества узлов на границах отрезка могут возникнуть паразитные осцилляции, даже при небольшом уровне шума. Примеры таких

графиков можно увидеть на рисунках 5 и 6. На рисунке 6 также видно, что при такой степени полинома, он начинает интерполировать шум больше, чем исходную функцию.

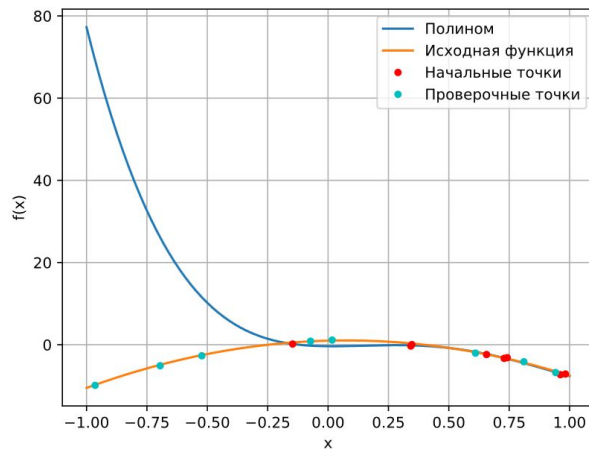


Рис. 5. $\sigma = 0.1$, $p = 4$, $N = 8$

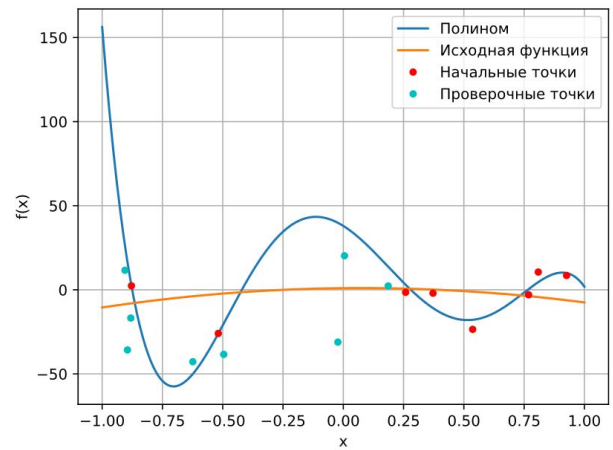


Рис. 6. $\sigma = 10$, $p = 5$, $N = 8$

Заключение

В результате первой части данной работы была выведена формула для численного интегрирования путем интерполирования подынтегральной функции тригонометрическими полиномами. Был реализован алгоритм Кули-Тьюки для быстрого преобразования Фурье, а также была выяснена зависимость погрешности интегрирования от количества узлов интерполяции.

В результате второй части работы была написана функция, осуществляющая поиск оптимальных коэффициентов полинома для приближения заданной функции. Было выяснено влияние зашумленности данных, их размерности и степени интерполирующего многочлена на погрешность приближения.

Список использованных источников

1. Першин А.Ю. *Лекции по вычислительной математике (черновик)* [Электронный ресурс] // Кафедра РК6 (Системы автоматизированного проектирования), МГТУ им. Н.Э. Баумана, 4 марта 2019 г.
2. Першин А.Ю. *Лекции по вычислительной математике. Презентации.* [Электронный ресурс]