МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского

**Н.Р. Strongina**

# КУРС «ЧИСЛЕННЫЕ МЕТОДЫ»

## Методы приближения функций и обработки экспериментальных данных, основанные на решении задач оптимизации

## (Модуль 14.2)

Учебно-методическое пособие

Рекомендовано методической комиссией
Института информационных технологий, математики и механики
для студентов ННГУ, обучающихся по направлению
01.03.02 «Прикладная математика и информатика»

Нижний Новгород
2021

УДК 519.6
ББК 22.19
С-86

С-86 Стронгина Н.Р. Курс «Численные методы»: Методы приближения функций и обработки экспериментальных данных, основанные на решении задач оптимизации (Модуль 14.2): Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2021. – 59 с.

Рецензент: к.ф.-м.н., доцент **А.А. Перов**

Пособие является компонентом учебно-методического комплекса по дисциплине «Численные методы». В пособии представлены классические подходы к приближению функций, основанные на минимизации расстояния от функции до элемента, ее приближающего, в метрике бесконечномерного гильбертова пространства; отыскание наилучших равномерных приближений и улучшение свойств усеченных степенных рядов; обработка экспериментальных данных методом наименьших квадратов. В каждом случае представлено обоснование оптимального выбора, приведены примеры и пошаговый разбор решения задач.

Пособие предназначено для студентов университета, обучающихся по направлению 01.03.02 «Прикладная математика и информатика», а также для преподавателей.

Ответственный за выпуск:
председатель методической комиссии
Института информационных технологий, математики и механики ННГУ
к.ф.-м.н., доцент А.В. Грезина

УДК 519.6
ББК 22.19

## СОДЕРЖАНИЕ

# ВВЕДЕНИЕ

Развитие вычислительной техники и последующее развитие высокопроизводительных вычислительных систем открывают качественно новые возможности изучения сложных реальных объектов методами вычислительного эксперимента [11, 12].

Машинный вычислительный эксперимент как новый метод научного исследования предполагает дискретизацию исходной задачи. Он требует специальной проработки численного алгоритма: корректность, устойчивость, точность, сходимость. Поэтому на современном этапе подготовки выпускников по направлению «Прикладная математика и информатика» основной целью освоения дисциплины «Численные методы» является изучение фундаментальных принципов построения численных алгоритмов, подходов к анализу их свойств, подготовка студентов к разработке и применению эффективных вычислительных комплексов, необходимых для математического моделирования сложных систем.

В Институте информационных технологий, математики и механики ННГУ в системе подготовки бакалавров по указанному выше направлению дисциплина «Численные методы» изучается на 3-м курсе в течение двух семестров. Обучение включает лекции, практические и лабораторные занятия, самостоятельную работу, зачеты и экзамен. Содержание дисциплины соответствует требованиям федеральных государственных образовательных стандартов и обновляется с учетом проблематики научных исследований и технологий программирования. Фундаментальные основы курса соответствует требованиям типовой программы по направлению «Прикладная математика и информатика», разработанной под руководством академика РАН А.А. Самарского [13].

Курс содержит изучение основ машинной арифметики, анализ структуры погрешности, подходы и методы приближенного вычисления функций, численное дифференцирование и интегрирование, численное решение систем линейных алгебраических уравнений, задач на собственные значения, решение нелинейных алгебраических уравнений и систем. Особое внимание уделяется инструментам математического моделирования сложных систем: методам численного решения задачи Коши и краевых задач для обыкновенных дифференциальных уравнений (ОДУ), решению уравнений в частных производных, а также структуре соответствующих вычислительных комплексов.

В связи с успешным применением в ННГУ практико-ориентированного подхода и на основе принципа «образование как исследование», вытекающего из положения Гумбольдта «образование на основе исследований» [11], фундаментальный курс «Численные методы» имеет в ННГУ уровневую

структуру. С одной стороны, в нем представлены все основные разделы численного анализа. С другой стороны, актуальные приложения требуют одновременного использования разных методов. Поэтому основой курса является системное изучение модельных задач, описывающих свойства реальных объектов различной природы. Освоение разделов дисциплины построено таким образом, чтобы в течение каждого семестра студенты могли самостоятельно подготовить программную реализацию численного алгоритма для решения модельной задачи, провести вычислительный эксперимент и подготовить отчет.

Нижегородский государственный университет является участником Суперкомпьютерного консорциума университетов России [12]. Студентов 3-го курса, изучающих дисциплину «Численные методы», знакомят с подходами к организации параллельных вычислений. Глубокое изучение этих подходов опирается на тот же комплект модельных задач, но проводится на старших курсах после освоения дисциплин, посвященных технологиям и методам параллельного программирования.

При освоении курса «Численные методы» у студентов 3-го курса должны быть сформированы компетенции разработки и применения программных средств разного уровня сложности. Поэтому требования к программам, подготовленным студентами, также реализуют практико-ориентированный подход. Программа должна быть написана на алгоритмическом языке высокого уровня. Код, реализующий алгоритм, должен быть подготовлен студентом самостоятельно. Объектно-ориентированный подход приветствуется. Программа и способ работы с ней должны быть пригодны не только для выполнения конкретного расчета, но также для проверки корректной реализации метода и результатов вычислительного эксперимента, и затем для изучения свойств метода и свойств моделируемого объекта. Ряд заданий выполняются с помощью специальной программы-тренажера, затем – с помощью программы, подготовленной студентом.

Требования самостоятельной программной реализации алгоритма и последующего самостоятельного проведения вычислительного эксперимента предполагают, что при рассмотрении теоретического материала, проведении практических занятий, выполнении заданий в рамках самостоятельной работы необходимо уделить больше внимания анализу понятийного аппарата дисциплины, доказательной базе, рассмотрению «простых» примеров и разбору по шагам решений специально подобранных задач. Решение именно этой учебной задачи поддерживает предлагаемое пособие.

Изучение тематического модуля, представленного в пособии, опирается на дисциплины «Дифференциальные уравнения», «Математический анализ», «Геометрия и алгебра» и «Программирование на ЭВМ» и осуществляется

одновременно с изучением дисциплин «Уравнения математической физики» и «Функциональный анализ».

Нумерация разделов пособия соответствует установленной в настоящее время нумерации тематических модулей электронного учебного курса «Численные методы», представленного в системе электронного обучения ННГУ (СЭО ННГУ) на базе платформы Moodle. В период весеннего семестра 2019-20 учебного года и осеннего семестра 2020-21 учебного года дистанционная организация учебного процесса по дисциплине «Численные методы» выстраивалась на базе этого электронного курса [14].

Пособие предназначено для студентов университета, обучающихся по направлению подготовки 01.03.02 «Прикладная математика и информатика», изучающих курс «Численные методы», и преподавателей.

Материал пособия может быть полезен студентам, изучающим в вузе численные методы на различных направлениях подготовки, а также студентам магистратуры ИИТММ, изучающим параллельные численные методы на основе технологий параллельного программирования.

# Модуль 14.2. Методы приближения функций и обработки экспериментальных данных, основанные на решении задач оптимизации

## *14.2.1. Наилучшие приближения в гильбертовых пространствах*

### *Измерение расстояния между функциями в гильбертовом пространстве*

**С целью решения задачи об отыскании наилучшего приближения функции, заданной в гильбертовом пространстве, приведем сведения из функционального анализа.**

Пусть $H$ – гильбертово пространство, в общем случае - бесконечномерное.

Для любых элементов $f, g$ из $H$ определен функционал, именуемый **скалярным произведением**. Этот функционал обозначают символом

$$(f, g)_H \qquad (14.1)$$

и в каждом пространстве $H$ этот функционал должен соответствовать **аксиомам скалярного произведения**.

Именно этот функционал определяет «состав» и свойства своего гильбертова пространства.

Функционал, именуемый **нормой** элемента $f \in H$, обозначают символом

$$\| f \|_H$$

Для элементов гильбертова пространства $H$ **норму определяют на основе скалярного произведения** (14.1) как **корень из скалярного квадрата элемента**:

$$\| f \|_H = \sqrt{(f, f)_H} \qquad (14.2)$$

*В курсе ФА доказано, что функционал «норма», определяемый по правилу (14.2), будет соответствовать всем аксиомам нормы.*

Функционал, именуемый **расстоянием** между элементами $f, g$ из $H$, обозначают символом

$$\rho(f, g)_H$$

Для элементов гильбертова пространства $H$ **расстояние определяют как норму разности элементов**:

$$\rho(f, g)_H = \| f - g \|_H \qquad (14.3)$$

*В курсе ФА доказано, что функционал «расстояние», определяемый по правилу (14.3), будет соответствовать всем аксиомам расстояния.*

### *Принцип отыскания наилучшего приближения*

**Для того, чтобы решить задачу об отыскании наилучшего приближения некоторого элемента гильбертова пространства, нужно**

**– определить класс элементов, среди которых необходимо найти такое приближение;**

**– выбрать в качестве приближения элемент, наиболее близкий к заданному элементу по расстоянию.**

**Разность элементов называют погрешностью**, причем погрешность также является элементом пространства $H$ .

**Норму погрешности** используют для описания **качества приближения**: норма погрешности говорит о том, велика погрешность или мала.

**Пример пространства, нормы, расстояния**

**Рассмотрим в качестве примера $H = L_2[0;1]$ – гильбертово пространство функций, определенных на отрезке $[0;1]$ и «суммируемых на данном отрезке с квадратом». То есть функций, для которых существует конечное значение интеграла**

$$I = \int_0^1 f^2(x)dx$$

**Скалярным произведением элементов $f, g \in L_2[0;1]$ является функционал, обозначенный символом**

$$(f,g)_{L_2[0;1]}$$

**заданный формулой**

$$(f,g)_{L_2[0;1]} = \int_0^1 f(x)\, g(x)\, dx\,.$$

**Нормой элемента $f \in L_2[0;1]$ является функционал, обозначенный символом**

$$\| f \|_{L_2[0;1]}$$

**заданный формулой**

$$\| f \|_{L_2[0;1]} = \sqrt{(f,f)_{L_2[0;1]}} = \sqrt{\int_0^1 f^2(x)\, dx}\,.$$

**Расстоянием между элементами $f, g \in L_2[0;1]$ является функционал, обозначенный символом**

$$\rho(f,g)_{L_2[0;1]}$$

**заданный формулой**

$$\rho(f,g)_{L_2[0;1]} = \| f - g \|_{L_2[0;1]} = \sqrt{\int_0^1 [\, f(x) - g(x)\,]^2\, dx}\,.$$

**Расстояние $\rho(f,g)_{L_2[0;1]}$ между элементами $f, g$ можно интерпретировать как погрешность приближения одним из элементов другого элемента: например, погрешность приближения $f \in L_2[0;1]$ элементом $g \in L_2[0;1]$.**

**Числовой пример**

Рассмотрим функции $f(x)=1,\ g(x)=x$. Каждая из них является элементом пространства $H=L_2[0;1]$, и для них верно

$$(f,g)_{L_2[0;1]}=\int_0^1 1\cdot x\,dx=\frac{1}{2}$$

(скалярное произведение функций $f(x)=1,\ g(x)=x$ равно $\frac{1}{2}$)

$$\| f \|_{L_2[0;1]} == \sqrt{\int_0^1 1\cdot 1\cdot dx}=1$$

$$\| g \|_{L_2[0;1]} == \sqrt{\int_0^1 x\cdot x\,dx}=\frac{1}{\sqrt{3}}$$

(нормы функций $f(x)=1,\ g(x)=x$ равны $1$ и $\frac{1}{\sqrt{3}}$ соответственно)

$$\rho(f,g)_{L_2[0;1]}=\| f-g \|_{L_2[0;1]}=\sqrt{\int_0^1 [1-x]^2\,dx}=\frac{1}{3}$$

(расстояние между функциями $f(x)=1,\ g(x)=x$ в пространстве $L_2[0;1]$ равно $\frac{1}{3}$).

Можно сказать, что функция $g(x)=x$ приближает функцию $f(x)=1$ с погрешностью

$$f(x)-g(x)=1-x$$

и норма погрешности в пространстве $L_2[0;1]$ равна $\frac{1}{3}$.

Можно посмотреть иначе: функция $f(x)=1$ приближает $g(x)=x$ погрешностью

$$g(x)-f(x)=x-1$$

и норма погрешности в пространстве $L_2[0;1]$ также равна числу $\frac{1}{3}$.

### *Отыскание элемента наилучшего приближения в конечномерном классе H*

**Пусть $H$ – гильбертово пространство (в общем случае – бесконечномерное).**

**Пусть $K_n \subset H$ – его подпространство конечной размерности $n$.**

**Линейно независимые элементы, образующие базис $K_n$, обозначим**

$$\varphi_i \in H, i = 1,...n. \qquad (14.4)$$

**Тогда любой элемент $\varphi \in K_n$ можно единственным образом представить в виде линейной комбинации базисных элементов**

$$\varphi = \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + ... + \alpha_n \varphi_n \qquad (14.5)$$

а подпространство $K_n$ – записать как множество всех линейных комбинаций вида (14.5):

$$K_n = \{ \sum_{i=1}^{n} \alpha_i \cdot \varphi_i \mid \alpha_i \in R,\ \varphi_i \in H,\ i = 1,...n \} \qquad (14.6)$$

**Определение 1.** Пусть $f \in H$ – элемент гильбертова пространства $H$. Элемент $\varphi \in K_n$ называют **элементом наилучшего приближения $f$ в классе $K_n$**, если для $\forall \tilde{\varphi} \in K_n,\ \tilde{\varphi} \neq \varphi$, верно

$$\rho(f, \varphi)_H \leq \rho(f, \tilde{\varphi})_H \qquad (14.7)$$

Читается так: расстояние между $f$ и $\varphi$ не превышает расстояния между $f$ и любым другим элементом класса $K_n$ (все расстояния измерены по правилам пространства $H$).

Ответ на вопрос о существовании, единственности и способе построения элементов наилучшего приближения содержится в следующем утверждении.

**Утверждение 1.** Пусть $H$ – гильбертово пространство, $K_n \subset H$ – подпространство конечной размерности $n$, причем линейно независимые элементы $\varphi_i \in H,\ i = 1,...n$ образуют базис подпространства $K_n$.

Тогда для $\forall f \in H$ элемент $\varphi \in K_n$, обеспечивающий наилучшее приближение $f$ в классе $K_n$ **существует**, является **единственным** и может быть представлен в виде (14.5), где коэффициенты $\alpha_i, i = 1,...n$ являются решением СЛАУ

$$\begin{bmatrix} (\varphi_1,\varphi_1)_H & (\varphi_1,\varphi_2)_H & ... & (\varphi_1,\varphi_n)_H \\ (\varphi_2,\varphi_1)_H & (\varphi_2,\varphi_2)_H & ... & (\varphi_2,\varphi_n)_H \\ ... & ... & ... & ... \\ (\varphi_n,\varphi_1)_H & (\varphi_n,\varphi_2)_H & ... & (\varphi_n,\varphi_n)_H \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (f,\ \varphi_1)_H \\ (f,\ \varphi_2)_H \\ ... \\ (f,\ \varphi_n)_H \end{bmatrix} \qquad (14.8)$$

СЛАУ (14.8) называют **нормальной системой уравнений.**

**Доказательство**

**Шаг I**

**Рассмотрим задачу оптимизации, отвечающую за отыскание элемента $\varphi$.**

**Элемент $\varphi$, наилучшим образом приближающий заданный элемент $f \in H$ в классе $K_n$, должен соответствовать условию (14.7).**

**Поэтому элемент $\varphi \in K_n$ следует искать как решение оптимизационной задачи**

$$\rho(f, \varphi)_H \rightarrow \min \qquad (14.9)$$

**где поиск минимального значения проводится для заданного $f$ из пространства $H$ по всем элементам $\varphi$, принадлежащим подпространству $K_n$.**

Заменим (14.9) на эквивалентную задачу минимизации квадрата расстояния:

$$\rho^2(f, \varphi)_H \rightarrow \min \qquad (14.10)$$

Используем (14.3) и запишем квадрат расстояния между $f$ и $\varphi$ как квадрат нормы разности элементов:

$$\rho^2(f, \varphi)_H = \| f - \varphi \|_H^2$$

Затем используем (14.2) и запишем норму через скалярный квадрат:

$$\| f - \varphi \|_H^2 = (f - \varphi, f - \varphi)_H \qquad (14.11)$$

**Таким образом, для отыскания элемента $\varphi \in K_n$, наилучшим образом приближающего заданный элемент $f$ из пространства $H$, нужно решить оптимизационную задачу**

$$(f - \varphi, f - \varphi)_H \rightarrow \min \qquad (14.12)$$

**где поиск минимума ведется по всем $\varphi$ из подпространства $K_n$.**

**Шаг II**

**Выясним, как выглядит функционал задачи (14.2).**

С помощью заданных базисных функций

$$\varphi_i \in H, i = 1,...n$$

каждый элемент $\varphi \in K_n$ может быть представлен в виде

$$\varphi = \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + ... + \alpha_n \varphi_n$$

**Поэтому функционал задачи (14.12) должен зависеть зависит от аргументов** $\alpha_i, i=1,...n$.

**Обозначим эту зависимость** $S(\alpha_1,\alpha_2,...\alpha_n)$ **и запишем (14.12) в виде**

$$S(\alpha_1,\alpha_2,...\alpha_n) \underset{\alpha\in R^n}{\rightarrow} \min. \qquad (14.13)$$

**При таком способе записи задачи оптимизации поиск минимального значения проводится в пространстве аргументов размерности** $n$.

Используя (14.5) и (14.12), запишем формулу функционала $S(\alpha_1,\alpha_2,...\alpha_n)$:

$$S(\alpha_1,\alpha_2,...\alpha_n)=(f-\varphi, f-\varphi)_H=(f-\sum_{i=1}^{n}\alpha_i\ \varphi_i, f-\sum_{j=1}^{n}\alpha_j\ \varphi_j)_H.$$

Раскрывая скалярное произведение, запишем

$$(f,f)_H-2(f,\sum_{i=1}^{n}\alpha_i\varphi_i)_H+(\sum_{i=1}^{n}\alpha_i\ \varphi_i,\sum_{j=1}^{n}\alpha_j\ \varphi_j)_H$$

Далее используем линейные свойства скалярного произведения в гильбертовом пространстве $H$.

Сначала из-под знаков скалярных произведений выносим знаки суммирования, а затем за скобками скалярных произведений должны оказаться числовые коэффициенты $\alpha_i, i=1,...n$. В итоге получим

$$S(\alpha_1,\alpha_2,...\alpha_n)=$$
$$=(f,f)_H-2\sum_{i=1}^{n}\alpha_i(f,\varphi_i)_H+\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\cdot\alpha_j\ (\varphi_i,\varphi_j)_H \qquad (14.14)$$

**Доказано, что** $S(\alpha_1,\alpha_2,...\alpha_n)$ **является квадратичной функцией своих аргументов** $\alpha_i, i=1,...n$.

**Шаг III и далее**

**Далее доказательство Утверждения 1 аналогично доказательству утверждений Модуля 14.1 и включает следующие этапы.**

1) Точки, подозрительные на экстремум, находим из условий

$$\frac{\partial S}{\partial \alpha_i}=0,\ i=1,...n \qquad (14.15)$$

Систему уравнений (14.15) называют **нормальной системой уравнений**.

2) Линейная независимость элементов $\varphi_i\in K_n, i=1,...n$ обеспечивает существование и единственность решения нормальной системы уравнений (14.15).

3) В силу линейной независимости элементов $\varphi_i\in K_n, i=1,...n$, единственное решение системы (14.15) является точкой локального минимума.

4) В силу свойств квадратичного функционала $S(\alpha_1,\alpha_2,...\alpha_n)$. единственный локальный минимум является глобальным.

**Кратко пройдем эти этапы.**

Для функционала (14.14) нормальная система уравнений (14.15) принимает вид

$$\begin{cases} \dfrac{\partial S}{\partial \alpha_1} = -2(f,\varphi_1)_H + 2\alpha_1(\varphi_1,\varphi_1)_H + 2\sum\limits_{j=2}^{n}\alpha_j(\varphi_1,\varphi_j)_H = 0 \\ \dfrac{\partial S}{\partial \alpha_2} = -2(f,\varphi_2)_H + 2\alpha_2(\varphi_2,\varphi_2)_H + 2\sum\limits_{j=1,j\neq 2}^{n}\alpha_j(\varphi_2,\varphi_j)_H = 0 \\ ... \\ \dfrac{\partial S}{\partial \alpha_n} = -2(f,\varphi_n)_H + 2\alpha_n(\varphi_n,\varphi_n)_H + 2\sum\limits_{j=1}^{n-1}\alpha_j(\varphi_n,\varphi_j)_H = 0 \end{cases}$$

Это СЛАУ с неизвестными $\alpha_i, i=1,...n$. Если ее записать в векторном виде, получим (14.8):

$$\begin{bmatrix} (\varphi_1,\varphi_1)_H & (\varphi_1,\varphi_2)_H & ... & (\varphi_1,\varphi_n)_H \\ (\varphi_2,\varphi_1)_H & (\varphi_2,\varphi_2)_H & ... & (\varphi_2,\varphi_n)_H \\ ... & ... & ... & ... \\ (\varphi_n,\varphi_1)_H & (\varphi_n,\varphi_2)_H & ... & (\varphi_n,\varphi_n)_H \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (f,\varphi_1)_H \\ (f,\varphi_2)_H \\ ... \\ (f,\varphi_n)_H \end{bmatrix}$$

Матрица СЛАУ (14.8) является матрицей Грама линейно независимых элементов $\varphi_i \in K_n, i=1,...n$:

$$Gr(\varphi_1,\varphi_2,...\varphi_n) = \begin{bmatrix} (\varphi_1,\varphi_1)_H & (\varphi_1,\varphi_2)_H & ... & (\varphi_1,\varphi_n)_H \\ (\varphi_2,\varphi_1)_H & (\varphi_2,\varphi_2)_H & ... & (\varphi_2,\varphi_n)_H \\ ... & ... & ... & ... \\ (\varphi_n,\varphi_1)_H & (\varphi_n,\varphi_2)_H & ... & (\varphi_n,\varphi_n)_H \end{bmatrix}$$

Поэтому указанная матрица не вырождена и положительно определена:

$$\det Gr(\varphi_1,\varphi_2...\varphi_n) \neq 0$$

$$Gr(\varphi_1,\varphi_2...\varphi_n) > 0$$

**Отсюда следует, что для любого элемента $f$ гильбертова пространства $H$ решение СЛАУ (14.8) существует и единственно.**

**Функционал $S(\alpha_1,\alpha_2,...\alpha_n)$ имеет единственную точку, подозрительную на экстремум.**

**Аналогично Утверждениям из Модуля 14.1 доказывается:**

**Точка, подозрительная на экстремум, является точкой локального минимума функционала $S(\alpha_1,\alpha_2,...\alpha_n)$.**

**Решение нормальной системы уравнений (14.8), являясь точкой локального минимума функционала $S(\alpha_1,\alpha_2,...\alpha_n)$, является решением задачи минимизации (14.13), то есть глобальным минимумом $S(\alpha_1,\alpha_2,...\alpha_n)$.**

**Для любого элемента $f$ гильбертова пространства $H$ решение задачи оптимизации (14.13) существует, единственно и может быть найдено в виде (14.5), где коэффициенты $\alpha_i, i=1,...n$ являются решением СЛАУ (14.8).**

Считаем, что Утверждение 1 доказано.

**Определение 2. Погрешностью** приближения элемента $f$ гильбертова пространства $H$ элементом $\varphi$ конечномерного подпространства $K_n \subset H$ является элемент $z \in H$, определяемый как

$$z = f - \varphi \qquad (14.16)$$

**Качество приближения** характеризуется **нормой погрешности**, то есть значением

$$\| z \|_H$$

которое в данном случае является **корнем квадратным из минимального значения функционала** $S(\alpha_1,\alpha_2,...\alpha_n)$:

$$\| z \|_H = \| f - \varphi \|_H = \sqrt{S(\alpha_1,\alpha_2,...\alpha_n)}.$$

**Следствие. Пусть в условиях Утверждения 1 линейно независимые элементы $\varphi_i \in H,\ i=1,...n$, образующие базис подпространства $K_n$, ортогональны, то есть**

$(\varphi_i, \varphi_j)_H = 0,\ i, j = 1,...n, i \neq j..$

Тогда для $\forall f \in H$ элемент $\varphi \in K_n$, обеспечивающий наилучшее приближение $f$ в классе $K_n$ **существует**, является **единственным** и может быть представлен в виде

$$\varphi = \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + ... + \alpha_n \varphi_n$$

где **коэффициенты $\alpha_i, i = 1,...n$ являются решением СЛАУ с диагональной матрицей**

$$\begin{bmatrix} (\varphi_1, \varphi_1)_H & 0 & ... & 0 \\ 0 & (\varphi_2, \varphi_2)_H & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & (\varphi_n, \varphi_n)_H \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (f, \varphi_1)_H \\ (f, \varphi_2)_H \\ ... \\ (f, \varphi_n)_H \end{bmatrix} \qquad (14.17)$$

**Коэффициенты $\alpha_i, i = 1,...n$ вычисляются по формулам**

$$\alpha_i = \frac{(f, \varphi_i)_H}{(\varphi_i, \varphi_i)_H}, \quad i = 1,...n \qquad (14.18)$$

**и называются коэффициентами Фурье элемента $f$ по ортогональной системе линейно независимых элементов**

$\varphi_i \in H, \ i = 1,...n$.

## 14.2.2. Экономизация полиномов и степенных рядов

### *Цели понижения степени полинома*

Предположим, что значение функции $f(x)$ в точке $x$ вычисляется с помощью полинома, полученного усечением формулы Тейлора

$$f(x) \approx \underbrace{f(x^*)+f'(x^*)\cdot(x-x^*)+f''(x^*)\cdot\frac{(x-x^*)^2}{2!}+\ldots+f^{(n)}(x^*)\cdot\frac{(x-x^*)^n}{n!}}_{\text{Это усеченная формула Тейлора, то есть полином } S_n(x) \text{ степени не выше } n}$$

**Определение 1.** Погрешность замены функции $f(x)$ полиномом $S_n(x)$ обозначим $E(x)$, ее называют **погрешностью усечения**:

$$E(x)=f(x)-S_n(x) \qquad (14.1)$$

**Утверждение 1.** Погрешность усечения определяется **остаточным слагаемым** формулы Тейлора:

$$E(x)=f^{(n+1)}(\xi)\cdot\frac{(x-x^*)^{n+1}}{(n+1)!} \qquad (14.2)$$

Здесь остаток записан в форме Лагранжа, неизвестная точка $\xi\in[x^*;x]$.

Предположим, что для функции $f(x)$ в точке $x$ ряд Тейлора сходится.

Тогда, с одной стороны, чем **выше степень полинома**, тем **меньше (по модулю) погрешность усечения** и более точным должен быть результат вычисления функции.

С другой стороны, вычисление полиномов высоких степеней может приводить к накоплению вычислительной погрешности, потому что в одну сумму складываются и крупные, и малые, и совсем малые слагаемые ряда.

Поэтому при увеличении степени полинома, полученного на основе формулы Тейлора, общая погрешность вычисления функции может не убывать.

**Чтобы обеспечить**
**точность приближенного вычисления функции,**
**используют полиномы (усеченные ряды Тейлора)**
**высоких степеней,**
**а для того, чтобы при вычислении полиномов**
**избежать накопления вычислительной погрешности,**
**проводят их экономизацию.**

**Определение 2.** **Экономизацией полинома степени** $n$ называют такое понижение его степени, при котором **погрешность его замены полиномом степени не выше** $n-1$ является в том или ином смысле **оптимальной**.

### *Понижение степени полинома $x^n$ на основе наилучшего равномерного приближения полиномом меньших степеней*

Заменим на отрезке $[-1;1]$ полином $x^n$ полиномом меньшей степени $Q_{n-1}(x)$ так, чтобы разность указанных полиномов на отрезке $[-1;1]$ была (по модулю) как можно меньше.

Для этого запишем задачу оптимизации:

$$\max_{x\in[-1;1]} \left| x^n - Q_{n-1}(x) \right| \to \min \qquad (14.3)$$

Поиск минимального значения функционала (14.3) ведется по всем возможным полиномам степени $n-1$.

**Определение 3.** Задачу (14.3) называют **задачей об отыскании наилучшего равномерного приближения полинома $x^n$ полиномом меньших степеней на отрезке $[-1;1]$.**

**Утверждение 2. Решением задачи об отыскании наилучшего равномерного приближения полинома $x^n$ полиномом меньших степеней на отрезке $[-1;1]$ является**

$$Q_{n-1}(x) = x^n - T_n(x) \qquad (14.4)$$

Здесь $T_n(x)$ есть полином Чебышёва, наименее уклоняющийся от нуля на отрезке $[-1;1]$ в классе полиномов степени $n$ со старшим коэффициентом 1.

(рекуррентные формулы для вычисления полиномов Чебышёва можно найти в справочнике либо записать полином самостоятельно, используя формулы его корней, см. Доказательство).

Норма погрешности экономизации, то есть замены $x^n$ полиномом наилучшего равномерного приближения $Q_{n-1}(x)$ на отрезке $[-1;1]$, составит

$$\max_{x\in[-1;1]} \left| x^n - Q_{n-1}(x) \right| = \frac{1}{2^{n-1}} \qquad (14.5)$$

**Доказательство**

Под знаком модуля задачи (14.3) записан полином степени $n$ со старшим коэффициентом 1. Поэтому (14.3) может рассматриваться как задача об отыскании полинома степени $n$, наименее уклоняющегося от нуля на отрезке $[-1;1]$, со старшим коэффициентом 1:

$$\max_{x\in[-1;1]} \left| P_n(x) \right| \to \min \qquad (14.6)$$

Поиск минимального значения функционала (14.6) ведется по всем возможным полиномам степени $n$, имеющим при старшей степени коэффициент $1$.

Известно, что решением (14.6) является полином Чебышёва $T_n(x)$ со старшим коэффициентом 1:

$$P_n(x) = T_n(x). \qquad (14.7)$$

Максимальное по модулю значение $T_n(x)$ на отрезке $[-1;1]$ составит

$$\max_{x\in[-1;1]} \left| T_n(x) \right| = \frac{1}{2^{n-1}} \qquad (14.8)$$

Вернемся к (14.3). Так как решением (14.6) является $T_n(x)$, решением (14.3) станет такой $Q_{n-1}(x)$, для которого

$$x^n - Q_{n-1}(x) = T_n(x) \qquad (14.9)$$

Следовательно, решение (14.3) записывается в виде (14.4)

$$Q_{n-1}(x) = x^n - T_n(x)$$

При подстановке в формулу полинома Чебышёва слагаемые степени $n$ сокращаются, и $Q_{n-1}(x)$ не будет содержать слагаемых степени выше $n-1$.

Погрешность замены полинома $x^n$ полиномом $Q_{n-1}(x)$ в точке $x$ составит

$$x^n - Q_{n-1}(x) \qquad (14.10)$$

С учетом (14.8) и (14.9) для нормы погрешности верно

$$\max_{x\in[-1;1]} \left| x^n - Q_{n-1}(x) \right| = \max_{x\in[-1;1]} \left| T_n(x) \right| = \frac{1}{2^{n-1}}$$

что доказывает (14.5).

**Полином Чебышёва $T_n(x)$, наименее уклоняющийся от нуля на отрезке $[-1;1]$ в классе полиномов степени $n$ со старшим коэффициентом 1, имеет на отрезке $[-1;1]$ $n$ различных корней:**

$$x_s = \cos\left( \frac{\pi}{2n}(1+2s) \right), \quad s = 0,...n-1$$

**и записывается в виде**

$$T_n(x) = (x - x_0)(x - x_1)...(x - x_{n-1}).$$

**Комментарии к названиям задач**

1) Функционал задачи (14.3), а именно

$$\max_{x\in[-1;1]} \left| x^n - Q_{n-1}(x) \right|$$

имеет следующий смысл:

$$\max_{x\in[-1;1]} \Big|\underbrace{x^n - Q_{n-1}(x)}_{\text{Это разность полиномов}}\Big|$$

$\underbrace{\qquad\qquad}_{\text{Это модуль разности полиномов}}$

$\underbrace{\qquad\qquad}_{\text{Это максимальное на отрезке } [-1;1] \text{ значение модуля разности полиномов}}$

Так записывается **задача об отыскании для $x^n$ полинома наилучшего равномерного приближения в классе полиномов меньших степеней на отрезке** $[-1;1]$.

2) Функционал задачи (14.3) можно рассматривать иначе:

$$\max_{x\in[-1;1]} \Big|\underbrace{x^n - Q_{n-1}(x)}_{\text{Это полином степени n со старшим коэффициентом равным 1}}\Big|$$

$\underbrace{\qquad\qquad}_{\text{Это модуль значения полинома, то есть уклонение полинома от нуля в точке } x}$

$\underbrace{\qquad\qquad}_{\text{Этомаксимальное на отрезке } [-1;1] \text{ уклонение полинома от нуля}}$

Так записывается **задача об отыскании полинома степени $n$, наименее уклоняющегося от нуля на отрезке** $[-1;1]$ **в классе полиномов степени $n$ со старшим коэффициентом 1**.

3) Решением (14.6) является полином Чебышёва со старшим коэффициентом 1. Поэтому решение (14.3) находят из условия

$$x^n - Q_{n-1}(x) = T_n(x)$$

4) В названии задачи оптимизации использовано следующее обстоятельство:

$$\max_{x\in[-1;1]}\left|\,x^n-Q_{n-1}(x)\,\right| = \max_{x\in[-1;1]}\Bigg|\underbrace{\underbrace{x^n-Q_{n-1}(x)}_{\text{полином}}-\underbrace{\underset{\sim}{0}}_{\substack{\text{функция}\\ \text{"ноль"}}}}_{\substack{\text{уклонение полинома}\\ \text{от функции "ноль" в точке } x}}\Bigg|$$

### *Экономизация полиномов для вычисления экспоненты (пример)*

Для функции $e^x$ запишем формулу Тейлора

$$e^x=1+x+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!}+\frac{x^{n+1}}{(n+1)!}e^{\xi}. \qquad (14.11)$$

Остаток представлен в форме Лагранжа.

**С целью приближенного вычисления $e^x$ используем полином $S_n(x)$ степени $n$, полученный усечением формулы**: $e^x\approx S_n(x)$, где

$$S_n(x)=1+x+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!} \qquad (14.12)$$

**Погрешность применения $S_n(x)$ в точке $x$ (то есть погрешность усечения) составит**

$$E(x)=e^x-S_n(x)=\frac{x^{n+1}}{(n+1)!}e^{\xi},\ \xi\in[0;x] \qquad (14.13)$$

**При $x\in[-1;1]$ погрешность оценивается неравенством**

$$\max_{x\in[-1;1]}\left|\,E(x)\,\right|\le\frac{e}{(n+1)!} \qquad (14.14)$$

**Используя наилучшее равномерное приближение полинома $x^n$ на отрезке $[-1;1]$ полиномом меньших степеней, понизим степень полинома $S_n(x)$.**

Для этого в (14.12) заменим $x^n$ (старшую степень) на полином $Q_{n-1}(x)$:

$$Q_{n-1}(x)=x^n-T_n(x) \qquad (14.15)$$

Получим на основе $S_n(x)$ полином меньшей степени, обозначим его $S^*_{n-1}(x)$:

$$S^*_{n-1}(x)=1+x+\frac{x^2}{2!}+\ldots+\frac{x^{n-1}}{(n-1)!}+\frac{x^n-T_n(x)}{n!} \qquad (14.16)$$

**Для приближенного вычисления $e^x$ вместо $S_n(x)$ используем $S^*_{n-1}(x)$:**

$$e^x \approx S^*_{n-1}(x).$$

**Погрешность применения $S^*_{n-1}(x)$ в точке $x$ для вычисления $e^x$ составит**

$$E^*(x) = e^x - S^*_{n-1}(x) \qquad (14.17)$$

Исследуем эту погрешность при $x \in [-1;1]$.

**Утверждение 3.** Погрешность вычисления $e^x$ с помощью экономизированного полинома $S^*_{n-1}(x)$ при $x \in [-1;1]$ оценивается неравенством

$$\max_{x\in[-1;1]} \left| E^*(x) \right| \le \frac{e}{(n+1)!} + \frac{1}{2^{n-1}n!} \qquad (14.18)$$

Здесь $\dfrac{e}{(n+1)!}$

– оценка погрешности усечения, то есть замены $e^x$ усеченной формулой Тейлора;

$\dfrac{1}{2^{n-1}n!}$

– погрешность экономизации, то есть замены $S_n(x)$ полиномом $S^*_{n-1}(x)$.

**Доказательство**

Запишем по определению погрешность применения $S^*_{n-1}(x)$ для вычисления $e^x$, добавим и вычтем из полученного выражения полином $S_n(x)$:

$$E^*(x) = e^x - S^*_{n-1}(x) = \underbrace{e^x - S_n(x)}_{\substack{\text{погрешность} \\ \text{усечения}}} + \underbrace{S_n(x) - S^*_{n-1}(x)}_{\substack{\text{погрешность} \\ \text{экономизации}}} \qquad (14.19)$$

Очевидно, что $E^*(x)$ складывается из двух компонент: погрешности усечения и погрешности замены полинома $S_n(x)$ полиномом $S^*_{n-1}(x)$.

Как следует из (14.12) и (14.16), полиномы $S_n(x)$ и $S^*_{n-1}(x)$ отличаются только последним слагаемым, поэтому

$$S_n(x) - S^*_{n-1}(x) = \frac{x^n}{n!} - \frac{(x^n - T_n(x))}{n!} = \frac{T_n(x)}{n!} \qquad (14.20)$$

Для значений $x \in [-1;1]$ построим оценку:

$$\max_{x\in[-1;1]} \left| E^*(x) \right| \le \max_{x\in[-1;1]} \left| e^x - S_n(x) \right| + \max_{x\in[-1;1]} |S_n(x) - S^*_{n-1}(x)|$$

Из (14.14) для первого слагаемого получим

$$\max_{x\in[-1;1]}\left| e^x - S_n(x)\right| \le \frac{e}{(n+1)!}$$

Из (14.20) и свойств полинома Чебышёва следует оценка

$$\max_{x\in[-1;1]} |S_n(x) - S^*_{n-1}(x)| \le \frac{1}{n!}\cdot \max_{x\in[-1;1]} |T_n(x)| = \frac{1}{2^{n-1}n!}$$

Тогда для погрешности замены экспоненты полиномом $S^*_{n-1}(x)$ верно

$$\max_{x\in[-1;1]}\left| E^*(x)\right| \le \frac{e}{(n+1)!} + \frac{1}{2^{n-1}n!}$$

что и требовалось доказать.

**Комментарии**

1) **Если полином служит для приближенного вычисления $f(x)$, используют разные критерии целесообразности понижения степени полинома.**

В случае $f(x) = e^x$ эти критерии можно записать следующим образом:

$$\frac{e}{(n+1)!} >> \frac{1}{2^{n-1}n!} \qquad \text{(I)}$$

то есть оценка погрешности усечения формулы Тейлора до полинома степени $n$ много больше погрешности экономизации указанного полинома;

$$\frac{e}{n!} >> \frac{e}{(n+1)!} + \frac{1}{2^{n-1}n!} \qquad \text{(II)}$$

то есть оценка погрешности усечения формулы Тейлора до полинома степени $n-1$ много больше погрешности применения полинома, полученного усечением формулы Тейлора до степени $n$ и прошедшего экономизацию до степени $n-1$.

И тот, и другой критерий выполняются при достаточно больших $n$.

2) **Экономизацию проводят поэтапно, иногда много раз подряд, снижая степень полинома, например с 10 до 4.**

3) В этом примере рассмотрена погрешность вычисления экспоненты с помощью усеченной формулы Тейлора при $x\in[-1;1]$.

Вместе с тем ряд Тейлора для $f(x) = e^x$, а именно

$$e^x = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots \qquad \text{(III)}$$

сходится к значению $e^x$ в любой точке действительной оси.

Ряд (III) и полиномы, полученные при усечении ряда, можно использовать для вычисления $e^x$ при любом $x \in (-\infty;+\infty)$.

**Поскольку при больших значениях аргумента $x$ ряд (III) сходится медленно, значение $e^x$ вычисляют, выделяя целую и дробную часть числа $x$.**

**Например, если $x = 5.3$, то $e^{5.3} = e^5 \cdot e^{0.3}$.**

**Первый множитель** вычисляется возведением числа $e$ в степень $x = 5$.

**Второй множитель** можно вычислить с помощью усеченной формулы Тейлора для аргумента $x = 0.3$.

Такой аргумент попадает на отрезок $x \in [-1;1]$.

**Вместо формулы Тейлора, усеченной до полинома высокой степени $n$, можно использовать многократно экономизированный вариант, представляющий собой (без особой потери точности) полином меньшей степени.**

4) В этом разделе рассмотрены полиномы наилучшего равномерного приближения и прием экономизации полинома на отрезке $x \in [-1;1]$.

Аналогичным образом (с помощью других полиномов Чебышёва) решается проблема на других отрезках.

## 14.2.3. Метод наименьших квадратов (для обработки данных)

### *Задача о построении МНК-полинома*

Предположим, что некоторые процессы или объекты описываются вещественными переменными $X$ и $Y$, и по результатам наблюдений нужно выявить **функциональную** связь этих переменных.

Обычно одну из переменных рассматривают как причину (**фактор**, или объясняющую переменную), другую – как следствие (**отклик**, или объясняемую переменную).

Пусть $X$ – фактор, $Y$ – отклик. Рассмотрим метод построения функциональной зависимости $Y$ от $X$ в виде полинома степени не выше $K$.

Результаты наблюдений над процессами (объектами) запишем в виде пар значений

$$(X_i, Y_i),\ i = 1,...n \qquad (14.1)$$

где $n$ есть количество наблюдений, $i$ – номер наблюдения.

Точки с координатами (14.1) отметим на плоскости $X - Y$.

Если количество точек ( $n$ ) больше, чем количество неизвестных коэффициентов полинома ( $K+1$) , полином нужной степени может через эти точки не пройти.

**Метод наименьших квадратов (МНК) позволяет строить полином нужной степени, который не всегда пройдет через заданные точки, но приблизится к ним оптимальным образом**.

Чтобы сформулировать критерий оптимальности, нужно отличать значения отклика, полученные при сборе данных, от значений, вычисленных с помощью полинома.

**Определение 1. Истинными значениями отклика** $Y$ называют значения $Y = Y_i,\ i = 1,...n,$ которые измерены при $X = X_i,\ i = 1,...n$ и указаны в наборе данных (14.1).

Значения $Y$, вычисленные с помощью полинома, обозначим через $\hat{Y}$.
**Искомый полином запишем в виде**

$$\hat{Y} = b_0 + b_1 X + ... + b_K X^K \qquad (14.2)$$

**Определение 2.** Значения отклика, вычисленные при $X = X_i,\ i = 1,...n$ , обозначим через $\hat{Y}_i,\ i = 1,...n$:

$$\hat{Y}_i = b_0 + b_1 X_i + ... + b_K X_i^K,\ \ i = 1,...n \qquad (14.3)$$

**Величины** $\hat{Y}_i,\ i = 1,...n$ **называют оценочными значениями отклика** $Y$.

**Определение 3. Остатками** $\hat{\varepsilon}_i,\ i = 1,...n$ **называют разности истинных и оценочных значений отклика**

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, \quad i = 1,...n \qquad (14.4)$$

**Определение 4.** Согласно методу наименьших квадратов (МНК), среди всех полиномов вида (14.2) **наилучшим** считается тот, которому соответствует **минимальная сумма квадратов остатков**.

Такой полином называют **МНК-полиномом**.

**Пример 1**

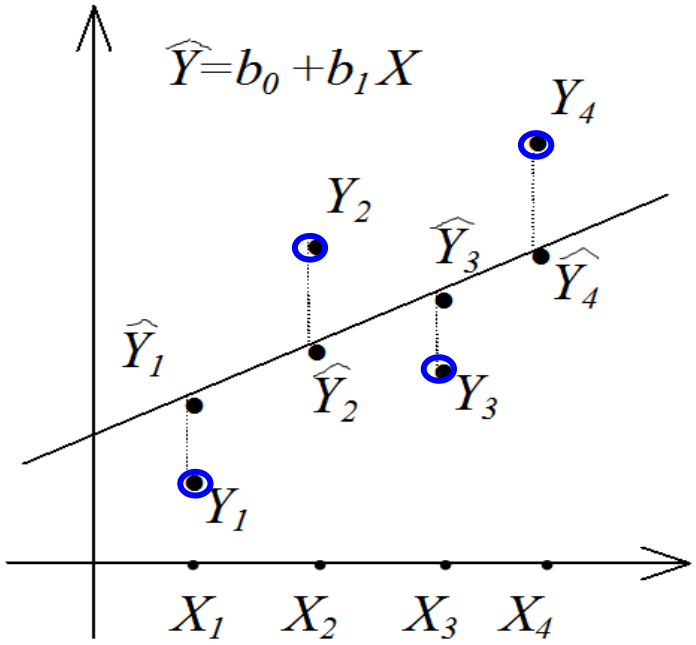


**Рисунок 1**

На рисунке показаны точками результаты 4-х наблюдений: $(X_i, Y_i)$, $i = 1,..4$, синий маркер. Через указанные точки нельзя провести полином степени $K = 1$ (нельзя провести прямую).

Поэтому показна МНК-прямая $\hat{Y} = b_0 + b_1 X$ и оценочные значения $\hat{Y}_i$, $i = 1,...4$ (они расположены на МНК-прямой).

Пунктиром показаны абсолютные значения остатков $\hat{\varepsilon}_i$, $i = 1,...4$.
Остатки с номерами 1, 3 отрицательны, с номерами 2, 4 – положительны.

МНК-прямая, показанная на рисунке, обеспечивает минимальную сумму квадратов остатков:

значение $S = \sum_{i=1}^{4} [\hat{\varepsilon}_i]^2$ для данной прямой минимально.

### *Способ построения МНК-полинома*

**Сумму квадратов остатков обозначим $S$ и выясним, чему она равна:**

$$S = \sum_{i=1}^{n} [\hat{\varepsilon}_i]^2 = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 = \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K)]^2 \qquad (14.5)$$

**Для построения МНК-полинома необходимо найти такие коэффициенты $b_j$, $j = 0,...K$, для которых $S$ принимает минимальное значение.**

**То есть нужно решить задачу**

$$S(b_0, b_1, ...b_K) = \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K)]^2 \to \min \qquad (14.6)$$

**при** $(b_0, b_1, ...b_K) \in R^{K+1}$

При решении задачи оптимизации (14.6) $S$ рассматривается как функция $K+1$ переменной $b_j$, $j = 0,...K$, а значения $X_i, Y_i$, $i = 1,...n$ есть числа: они уже получены в результате замеров (14.1).

Чтобы сформулировать результат о существовании, единственности и способе отыскания МНК-полинома, введем дополнительные обозначения.

Значения фактора $X$ и функции фактора $X$ запишем в матрицу $\boldsymbol{X}$, которую называют **матрицей регрессоров**.

Матрица $\boldsymbol{X}$ имеет размерность $n \times (K+1)$. Ее строки соответствуют наблюдениям, а столбцы – регрессорам.

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_1 & X_1^2 & ... & X_1^K \\ 1 & X_2 & X_2^2 & ... & X_2^K \\ ... & ... & ... & ... & ... \\ 1 & X_n & X_n^2 & ... & X_n^K \end{bmatrix} \qquad (14.7)$$

Первый столбец матрицы $\boldsymbol{X}$ состоит из единиц. Во втором столбце матрицы $\boldsymbol{X}$ указаны значения фактора $X$ по всем замерам: $X_i$, $i = 1,...n$. В третьем и следующих столбцах – результат возведения фактора $X$ в соответствующую степень – от $2$ до $K$:

$X_i^2$, $i = 1,...n$,

… … …

$X_i^K$, $i = 1,...n$

Дополнительно к (14.7) нужны **обозначения для векторов**: истинные значения отклика $Y$ запишем как вектор $\overline{\boldsymbol{Y}}$ размерности $n$; столбцы матрицы $\boldsymbol{X}$ – как векторы $\overline{\boldsymbol{X}}^{(j)}, j=0,...K$ размерности $n$:

$$\overline{\boldsymbol{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_n \end{bmatrix}, \quad \overline{\boldsymbol{X}}^{(0)}=\begin{bmatrix} 1 \\ 1 \\ ... \\ 1 \end{bmatrix}, \overline{\boldsymbol{X}}^{(1)}=\begin{bmatrix} X_1 \\ X_2 \\ ... \\ X_n \end{bmatrix}, \quad \dots\dots\dots \quad \overline{\boldsymbol{X}}^{(K)}=\begin{bmatrix} X_1^K \\ X_2^K \\ ... \\ X_n^K \end{bmatrix}$$

**Утверждение 1.** Для любого набора данных (14.1), такого, что ранг матрицы $\boldsymbol{X}$ равен $K+1,$ МНК-полином (14.2) **существует** и **является единственным**, а его коэффициенты $b_j, j=1,...K$ **являются решением нормальной системы уравнений**:

$$\frac{\partial S}{\partial b_j}=0,\ j=0,...K \tag{14.8}$$

Система (14.8) представляет собой СЛАУ с неизвестными $b_j, j=1,...K$:

$$\begin{bmatrix} (\overline{X}^{(0)},\overline{X}^{(0)}) & (\overline{X}^{(0)},\overline{X}^{(1)}) & ... & (\overline{X}^{(0)},\overline{X}^{(K)}) \\ (\overline{X}^{(1)},\overline{X}^{(0)}) & (\overline{X}^{(1)},\overline{X}^{(1)}) & ... & (\overline{X}^{(1)},\overline{X}^{(K)}) \\ ... & ... & ... & ... \\ (\overline{X}^{(K)},\overline{X}^{(0)}) & (\overline{X}^{(K)},\overline{X}^{(1)}) & ... & (\overline{X}^{(K)},\overline{X}^{(K)}) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ ... \\ b_K \end{bmatrix} = \begin{bmatrix} (\overline{Y},\overline{X}^{(0)}) \\ (\overline{Y},\overline{X}^{(1)}) \\ ... \\ (\overline{Y},\overline{X}^{(K)}) \end{bmatrix} \tag{14.9}$$

Символом $(*\cdot,*)$ обозначено скалярное произведение в пространстве $R^n$.

**Доказательство**

**Рассмотрим задачу оптимизации (14.6):**

$$S(b_0,b_1,...b_K) \underset{b\in R^{K+1}}{\to} \min.$$

**Точки, подозрительные на экстремум, находим из условий**

$$\frac{\partial S}{\partial b_j}=0,\ j=0,...K$$

**Шаг за шагом покажем:**

1) линейная независимость столбцов матрицы $\boldsymbol{X}$ обеспечивает существование и единственность решения СЛАУ (14.9);

2) в силу линейной независимости указанных столбцов единственное решение СЛАУ (14.9) является точкой локального минимума;

3) в силу свойств функционала $S(b_0,b_1,...b_K)$ локальный минимум является глобальным, и это означает, что задача оптимизации (14.6) решена.

**Шаг I**

**Покажем, что систему (14.8) можно записывать в виде (14.9).**

Используя (14.5), запишем частные производные функционала $S(b_0, b_1, ...b_K)$:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K)) \\ \dfrac{\partial S}{\partial b_1} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K))\ X_i \\ ... \\ \dfrac{\partial S}{\partial b_K} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K))\ X_i^K \end{cases} \quad (14.10)$$

В соответствии уравнениями (14.8) приравниваем каждую из частных производных к нулю, полученные выражения делим на 2:

$$\begin{cases} \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_0} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K)) = 0 \\ \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_1} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K))\ X_i = 0 \\ ... \\ \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_K} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + ... + b_K X_i^K))\ X_i^K = 0 \end{cases} \quad (14.11)$$

Слагаемые, не зависящие от коэффициентов $b_j, j = 0,...K$, переносим в правую часть каждого уравнения:

$$\begin{cases} \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + ... + b_K X_i^K) = \sum\limits_{i=1}^{n} Y_i \\ \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + ... + b_K X_i^K)\ X_i = \sum\limits_{i=1}^{n} Y_i X_i \\ ... \\ \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + ... + b_K X_i^K)\ X_i^K = \sum\limits_{i=1}^{n} Y_i X_i^K \end{cases} \quad (14.12)$$

Таким образом, на основе (14.8) получена СЛАУ с неизвестными $b_j, j = 0,...K$.

Запишем СЛАУ в векторном виде:

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \dots & \sum_{i=1}^{n} X_i^K \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \dots & \sum_{i=1}^{n} X_i^{K+1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^{n} X_i^K & \sum_{i=1}^{n} X_i^{K+1} & \sum_{i=1}^{n} X_i^{K+2} & \dots & \sum_{i=1}^{n} X_i^{2K} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} Y_i X_i \\ \dots \\ \sum_{i=1}^{n} Y_i X_i^K \end{bmatrix}$$

Несложно проверить, что элемент матрицы СЛАУ, расположенный в строке с номером $l$ и столбце с номером $m$, имеет вид

$$\sum_{i=1}^{n} X_i^l \cdot X_i^m = \sum_{i=1}^{n} X_i^{l+m} \qquad (14.13)$$

Данный элемент совпадает со скалярным произведением столбцов матрицы $\boldsymbol{X}$ с номерами $l$ и $m$:

$$(\overline{X}^{(l)}, \overline{X}^{(m)}) = \sum_{i=1}^{n} X_i^{l+m} \qquad (14.14)$$

Аналогично для вектора, указанного в правой части СЛАУ: элемент, расположенный в строке с номером $l$ имеет вид

$$\sum_{i=1}^{n} Y_i \, X_i^l \qquad (14.15)$$

Он совпадает со скалярным произведением вектора $\overline{\boldsymbol{Y}}$ и столбца матрицы $\boldsymbol{X}$ с номером $l$:

$$(\overline{Y}, \overline{X}^{(l)}) = \sum_{i=1}^{n} Y_i \, X_i^l \qquad (14.16)$$

Учитывая соответствие (14.13)-(14.16), используя ранее введенные обозначения для столбцов матрицы $\boldsymbol{X}$, запишем СЛАУ (14.12) в виде (14.9)

$$\begin{bmatrix} (\overline{X}^{(0)}, \overline{X}^{(0)}) & (\overline{X}^{(0)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(0)}, \overline{X}^{(K)}) \\ (\overline{X}^{(1)}, \overline{X}^{(0)}) & (\overline{X}^{(1)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(1)}, \overline{X}^{(K)}) \\ \dots & \dots & \dots & \dots \\ (\overline{X}^{(K)}, \overline{X}^{(0)}) & (\overline{X}^{(K)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(K)}, \overline{X}^{(K)}) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} (\overline{Y}, \overline{X}^{(0)}) \\ (\overline{Y}, \overline{X}^{(1)}) \\ \dots \\ (\overline{Y}, \overline{X}^{(K)}) \end{bmatrix}$$

Таким образом, представление нормальной системы уравнений (14.8) в виде СЛАУ (14.9) доказано.

**Шаг II**

**Исследуем возможность решения СЛАУ.**

Матрица СЛАУ (14.9) является матрицей Грама для векторов, являющихся столбцами матрицы $\boldsymbol{X}$:

$$Gr(\overline{X}^{(0)},...\overline{X}^{(K)}) = \begin{bmatrix} (\overline{X}^{(0)},\overline{X}^{(0)}) & (\overline{X}^{(0)},\overline{X}^{(1)}) & ... & (\overline{X}^{(0)},\overline{X}^{(K)}) \\ (\overline{X}^{(1)},\overline{X}^{(0)}) & (\overline{X}^{(1)},\overline{X}^{(1)}) & ... & (\overline{X}^{(1)},\overline{X}^{(K)}) \\ ... & ... & ... & ... \\ (\overline{X}^{(K)},\overline{X}^{(0)}) & (\overline{X}^{(K)},\overline{X}^{(1)}) & ... & (\overline{X}^{(K)},\overline{X}^{(K)}) \end{bmatrix}$$

По условию Утверждения 1, ранг $\boldsymbol{X}$ равен $K+1$. Это означает, что столбцы матрицы $\boldsymbol{X}$ линейно независимы и поэтому матрица Грама не вырождена и положительно определена.

$$\det\ Gr\ (\overline{X}^{(0)}, \overline{X}^{(1)}...\overline{X}^{(K)}) \neq 0 \qquad (14.17)$$

$$Gr\ (\overline{X}^{(0)}, \overline{X}^{(1)}...\overline{X}^{(K)}) > 0 \qquad (14.18)$$

Те же самые свойства имеет матрица СЛАУ.

**Поэтому при любой правой части СЛАУ (14.9) ее решение существует и единственно, а функционал $S(b_0,b_1,...b_K)$ имеет единственную точку, подозрительную на экстремум.**

**Шаг III**

**Исследуем критическую точку с помощью матрицы вторых производных.**

$$S''(b_0, b_1...b_K) = \begin{bmatrix} \frac{\partial^2 S}{\partial b_0^2} & ... & \frac{\partial^2 S}{\partial b_0\ \partial b_K} \\ & ... & \\ ... & ... & ... \\ \frac{\partial^2 S}{\partial b_K\ \partial b_0} & ... & \frac{\partial^2 S}{\partial b_K^2} \end{bmatrix} \qquad (14.19)$$

Вычисляя вторые частные производные, убеждаемся в том,

что для функционала $S(b_0,b_1,...b_K)$

матрица $S''(b_0,b_1,...b_K)$

с точностью до множителя 2

совпадает с матрицей Грама

линейно независимых столбцов матрицы $\boldsymbol{X}$:

$$S''(b_0, b_1...b_K) = 2 \cdot Gr(\overline{X}^{(0)}, \overline{X}^{(1)}...\overline{X}^{(K)}) \qquad (14.20)$$

Приведем некоторые примеры совпадения их элементов:

$$\begin{cases} \dfrac{\partial^2 S}{\partial b_0^2} = 2\,n & \dfrac{\partial^2 S}{\partial b_0 \partial b_1} = 2\sum_{i=1}^{n} X_i \quad \ldots\ldots\ldots \dfrac{\partial^2 S}{\partial b_0 \partial b_K} = 2\sum_{i=1}^{n} X_i^K \\ \ldots & & \\ \ldots & \ldots & \ldots\ldots\ldots \dfrac{\partial^2 S}{\partial b_K^2} = 2\sum_{i=1}^{n} X_i^{2K} \end{cases}$$

Из (14.20) следует, что $S''(b_0, b_1 \ldots b_K)$, во-первых, не зависит от коэффициентов $b_j, j = 0,\ldots K$; во-вторых, положительно определена:

$$S''(b_0, b_1 \ldots b_K) = S'' > 0 \qquad (14.21)$$

**Таким образом, точка, подозрительная на экстремум, является точкой локального минимума.**

**Решение СЛАУ (14.9) является точкой локального минимума функционала $S(b_0, b_1, \ldots b_K)$.**

**Аналогично Утверждению 4 Модуля 14.1 доказывается:**

**Решение нормальной системы уравнений (14.9), являясь точкой локального минимума функционала $S(b_0, b_1, \ldots b_K)$, является решением задачи минимизации (14.6), то есть глобальным минимумом $S(b_0, b_1, \ldots b_K)$.**

Считаем, что Утверждение 1 доказано.

**Следствие**. Если в наборе данных (14.1) истинные значения отклика $Y$ представляют собой значения некоторого полинома степени не выше $K$, а именно

$$Y_i = a_0 + a_1 X_i + \ldots + a_K X_i^K, \quad i = 1,\ldots n \qquad (14.22)$$

методом наименьших квадратов будет построен именно этот полином:

$$\hat{Y} = b_0 + b_1 X + \ldots + b_K X^K,$$

где $b_j = a_j$, $j = 0,\ldots K$

**Доказательство**

Если в наборе данных (14.1) истинные значения отклика $Y$ представляют собой значения полинома (14.22) степени не выше $K$,

для данного полинома истинные и оценочные значения отклика совпадают.

Тогда остатки равны нулю

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - Y_i = 0, \quad i = 1,...n..$$

и полиному (14.22) соответствует нулевая сумма квадратов остатков.

Значение функционала $S(b_0, b_1,...b_K)$ меньше нуля не бывает.

Единственным решением (14.6) будет именно (14.22).

**Пример 2**

В таблице приведены результаты 4-х замеров:

| $X_i$ | 0 | 1 | 3 | 6 |
|---|---|---|---|---|
| $Y_i$ | 0.1 | 0.8 | 3.05 | 5.95 |

Нужно построить линейную зависимость $Y$ от $X$ методом наименьших квадратов и затем приблизить данные полиномом 2-й степени. В каждом из случаев указать, какой функционал должен быть минимизирован.

**Решение**

1) Строим полином степени $K = 1$, то есть МНК-прямую.

Количество наблюдений $n = 4$. Полином запишем в виде $\hat{Y} = b_0 + b_1 X$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0, b_1) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i)\,]^{\,2} \to \min$$

В данной задаче

$$S(b_0, b_1) = [0.1 - (b_0 + b_1 \cdot 0)]^2 + [0.8 - (b_0 + b_1 \cdot 1)]^2 +$$
$$+ [3.05 - (b_0 + b_1 \cdot 3)]^2 + [5.95 - (b_0 + b_1 \cdot 6)]^2$$

Если формулы (14.9) были «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0.$$

В этом примере о способе записи (14.9) не забываем.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{X}$ размерности 4×2 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$$

Столбцы матрицы регрессоров определяют векторы

$$\bar{X}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \bar{X}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_j$, $j = 0,1$ имеет вид:

$$\begin{bmatrix} 1+1+1+1 & 1\cdot 0+1\cdot 1+3\cdot 1+6\cdot 1 \\ 1\cdot 0+1\cdot 1+1\cdot 3+1\cdot 6 & 0\cdot 0+1\cdot 1+3\cdot 3+6\cdot 6 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.1\cdot 1+0.8\cdot 1+3.05\cdot 1+5.95\cdot 1 \\ 0.1\cdot 0+0.8\cdot 1+3.05\cdot 3+5.95\cdot 6 \end{bmatrix}$$

то есть

$$\begin{bmatrix} 4 & 10 \\ 10 & 46 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 9.9 \\ 45.65 \end{bmatrix}$$

**Решение СЛАУ и МНК-прямая показаны на Рисунке 2.**

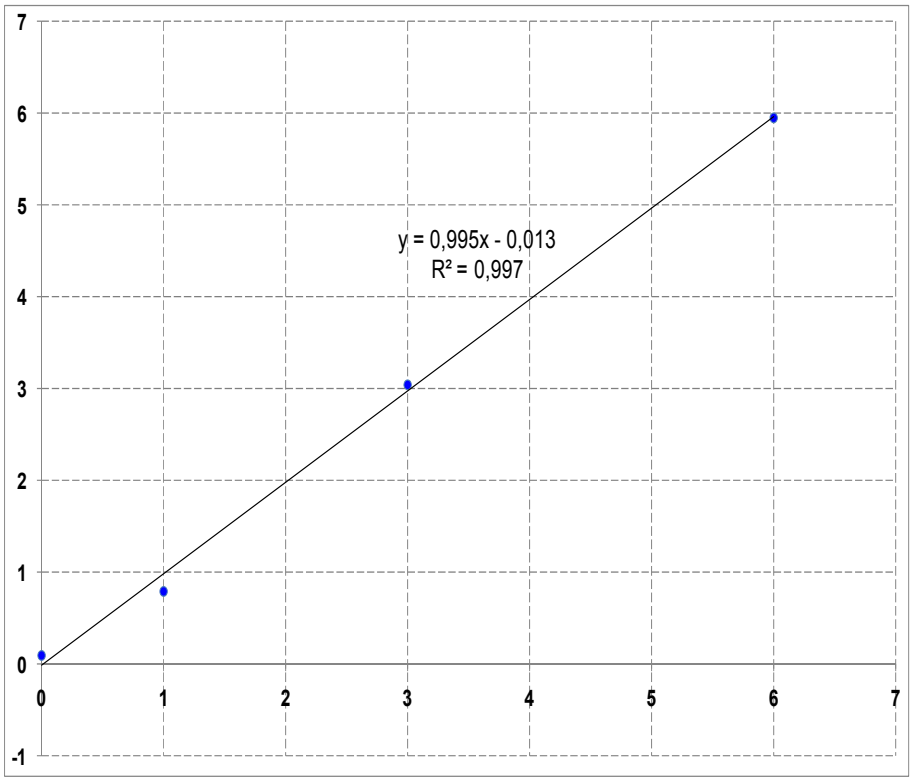


**Рисунок 2**

2) Строим полином степени $K = 2$, то есть МНК-параболу.

Количество наблюдений $n = 4$. Полином запишем в виде

$\hat{Y} = b_0 + b_1 X + b_2 X^2$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0, b_1, b_2) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i + b_2 X_i^2)]^{\,2} \to \min$$

В данной задаче

$$S(b_0, b_1, b_2) = [0.1 - b_0]^2 + [0.8 - (b_0 + b_1 + b_2)]^2 +$$

$$+ [3.05 - (b_0 + b_1 \cdot 3 + b_2 \cdot 9)]^2 + [5.95 - (b_0 + b_1 \cdot 6 + b_2 \cdot 36)]^2$$

Как и в предыдущем случае, если формулы (14.9) «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0, \quad \frac{\partial S}{\partial b_2} = 0.$$

Как и в предыдущем случае, о формулах (14.9) не забываем.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{X}$ размерности 4×3 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 6 & 36 \end{bmatrix}$$

Столбцы матрицы регрессоров определяют векторы

$$\overline{\boldsymbol{X}}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \overline{\boldsymbol{X}}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix} \qquad \overline{\boldsymbol{X}}^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 9 \\ 36 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_j$, $j = 0, 1, 2$ имеет вид:

$$\begin{bmatrix} 4 & 10 & 46 \\ 10 & 46 & 244 \\ 46 & 244 & 1378 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 9.9 \\ 45.65 \\ 242.45 \end{bmatrix}$$

**Решение СЛАУ и МНК-парабола показаны на Рисунке 3.**

3) Докажите, что попытка построить МНК полином степени $K = 3$

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 + b_3 X^3$$

приведет к тому, что в результате минимизации функционала

$$S(b_0, b_1, b_2, b_3) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3)]^{\,2} \to \min$$

будет построен **интерполяционный полином** степени $3$, см. Рисунок 4.

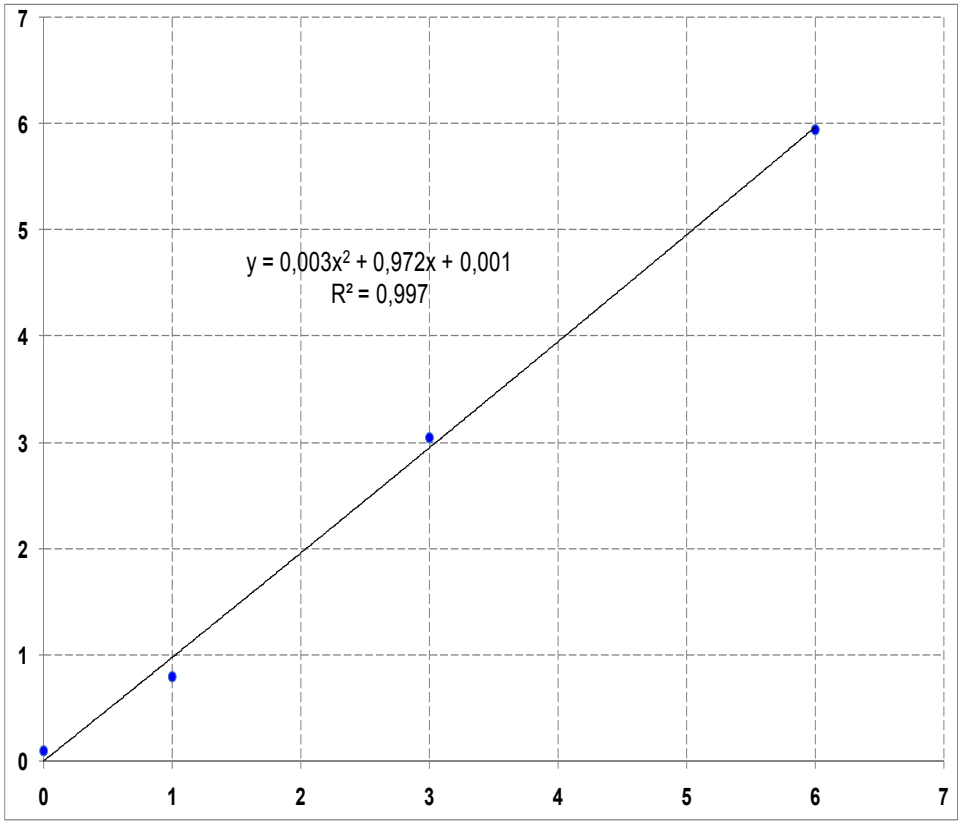


**Рисунок 3**

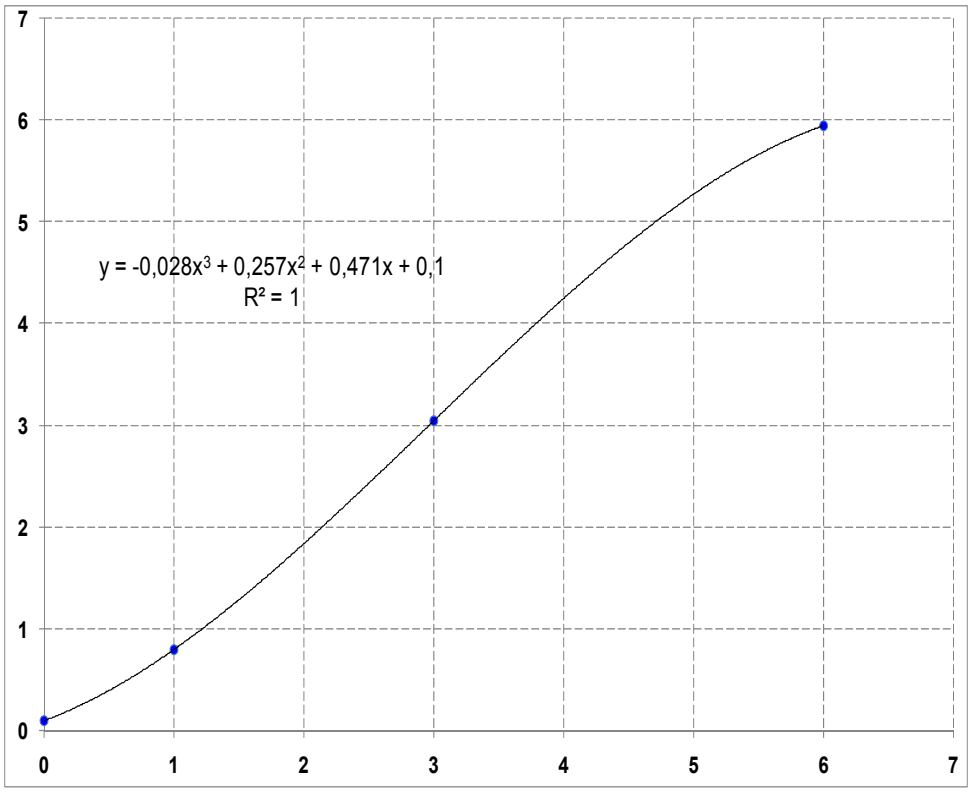


**Рисунок 4**

### *Задача о построении обобщенного МНК-полинома*

**Рассмотрим обобщение метода наименьших квадратов**
**на случай нескольких объясняющих переменных (нескольких факторов) и (или)**
**случай более сложных функциональных зависимостей**
**между откликом и факторами.**

> В этом разделе важно, чтобы функциональная зависимость отклика от факторов оставалась линейной относительно неизвестных параметров.

**Пусть количество объясняющих переменных (факторов) равно** $m$.

Фактор с номером $j$ обозначим $X^{(j)}$, $j = 1,...m$.

Через $X$ обозначим весь набор факторов:

$$X = (X^{(1)}, X^{(2)}, ...X^{(m)}), \ X \in R^m.$$

**Отклик обозначим** $Y$ .

Результаты наблюдений над процессами (объектами) запишем в виде наборов значений

$$X_i = (X_i^{(1)}, X_i^{(2)}, ...X_i^{(m)}), \ Y_i, \ i = 1,...n \qquad (14.23)$$

где $n$ есть количество наблюдений, $i$ – номер наблюдения,

$X_i \in R^m, i = 1,...n$ – набор значений факторов для наблюдения (объекта) с номером $i$,

$Y_i, \ i = 1,...n$ – значение отклика $Y$ для наблюдения (объекта) с номером $i$.

**Для описания функциональной зависимости отклика** $Y$ **от факторов** $X^{(j)}$, $j = 1,...m$ **будут использованы функции**

$$\varphi_0(X), \ \varphi_1(X), ... \varphi_K(X),$$

**а сама зависимость найдена в виде линейной комбинации этих функций:**

$$Y = b_0 \cdot \varphi_0(X) + b_1 \cdot \varphi_1(X) + ... + b_K \cdot \varphi_K(X) \qquad (14.24)$$

**Выражение (14.24) называют обобщенным полиномом**.

**Примеры обобщенных полиномов**

1) Если фактор только один, то есть $X \in R, \ m = 1,$ и для описания зависимости $Y$ от $X$ выбраны функции

$$\varphi_0(X) = 1, \ \varphi_1(X) = X, ... \varphi_K(X) = X^K$$

обобщенный полином (14.24) представляет собой полином степени $K$:

$$Y = b_0 + b_1 X + ... + b_K X^K$$

2) Если $X \in R,\ m = 1,$ и для описания зависимости $Y$ от $X$ выбраны функции

$$\varphi_0(X) = X,\ \varphi_1(X) = X^2$$

обобщенный полином является полиномом степени $2$ относительно $X$ без константы:

$$Y = b_0 X + b_1 X^2$$

3) Если $X \in R,\ m = 1,$ и для описания зависимости $Y$ от $X$ выбраны функции

$$\varphi_0(X) = X,\ \varphi_1(X) = e^X$$

обобщенный полином есть функция вида

$$Y = b_0 X + b_1 e^X$$

4) Если есть два фактора, то есть $X \in R^2,\ m = 2,$ и выбраны функции

$$\varphi_0(X) = 1,\ \varphi_1(X) = X^{(1)}, \ldots \varphi_2(X) = X^{(1)} \sin(X^{(2)})$$

обобщенный полином есть функция вида

$$Y = b_0 + b_1 X^{(1)} + b_2 X^{(1)} \sin(X^{(2)})$$

5) Если есть два фактора, то есть $X \in R^2,\ m = 2,$ и выбраны функции

$$\varphi_0(X) = 1,\ \varphi_1(X) = X^{(1)}, \ldots \varphi_2(X) = X^{(2)}$$

обобщенный полином есть линейная функция вида

$$Y = b_0 + b_1 X^{(1)} + b_2 X^{(2)}$$

**Функции $\varphi_0(X),\ \varphi_1(X), \ldots \varphi_K(X)$ выбирают, руководствуясь характером функциональной зависимости отклика от факторов, а неизвестные параметры $b_j,\ j = 0,\ldots K$ можно подобрать методом наименьших квадратов (МНК).**

**При этом в классе функций вида (14.24) будет выбрана такая, которая приблизит результаты наблюдений (14.23) оптимальным образом.**

Как и в предыдущем разделе, чтобы сформулировать критерий оптимальности, нужно отличать значения отклика, полученные при сборе данных, от вычисляемых значений.

**Определение 5. Истинными значениями отклика** $Y$ называют значения

$$Y = Y_i, \quad i = 1,...n,$$

измеренные при $X = X_i, \quad i = 1,...n$ и указанные в наборе значений (14.23).

Значения $Y$, вычисляемые с помощью (14.24), обозначим через $\hat{Y}$.
**Искомую зависимость запишем в виде**

$$\hat{Y} = b_0 \cdot \varphi_0(X) + b_1 \cdot \varphi_1(X) + ... + b_K \cdot \varphi_K(X) \qquad (14.25)$$

**Определение 6.** Значения отклика, **вычисленные** при $X = X_i, \quad i = 1,...n$, обозначим через $\hat{Y}_i, \quad i = 1,...n$:

$$\hat{Y} = b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i), \quad i = 1,...n \qquad (14.26)$$

**Величины** $\hat{Y}_i, \quad i = 1,...n$ **называют оценочными значениями отклика** $Y$.

**Определение 7. Остатками** $\hat{\varepsilon}_i, \quad i = 1,...n$ **называют разности истинных и оценочных значений отклика**

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, \quad i = 1,...n \qquad (14.27)$$

**Определение 8.** Согласно методу наименьших квадратов (МНК), среди всех обобщенных полиномов вида (14.25) **наилучшим** считается тот, которому соответствует **минимальная сумма квадратов остатков**.

Такой полином называют обобщенным **МНК-полиномом**.

### *Способ построения обобщенного МНК-полинома*

**Запишем, чему равна** $S$ **– сумма квадратов остатков:**

$$S = \sum_{i=1}^{n} [\hat{\varepsilon}_i]^2 = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 =$$

$$= \sum_{i=1}^{n} [Y_i - (b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i))]^2$$

(14.28)

**Для построения обобщенного МНК-полинома необходимо найти такие коэффициенты** $b_j, \ j = 0,...K$, **для которых** $S$ **принимает минимальное значение, то есть решить задачу**

$$S(b_0, b_1 ... b_K) = \sum_{i=1}^{n} [\, Y_i - (b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i))]^{\,2} \to \min \tag{14.29}$$

**при** $(b_0, b_1, ... b_K) \in R^{K+1}$

Здесь $S$ рассматривается как функция $K+1$ переменной $b_j$, $j = 0,...K$, а значения

$X_i = (X_i^{(1)}, X_i^{(2)}, ... X_i^{(m)})$, $Y_i$, $i = 1,...n$

указаны в наборе данных (14.23) и являются числами.

Как и в случае «обычных» полиномов, введем дополнительные обозначения.

Значения факторов $X$ и функции $\varphi_0(X)$, $\varphi_1(X), ... \varphi_K(X)$ запишем в матрицу $\boldsymbol{\Phi}$, которую называют **матрицей регрессоров**.

Матрица $\boldsymbol{\Phi}$ имеет размерность $n \times (K+1)$. Ее строки соответствуют наблюдениям, а столбцы – регрессорам.

$$\boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(X_1) & \varphi_1(X_1) & \varphi_2(X_1) & ... & \varphi_K(X_1) \\ \varphi_0(X_2) & \varphi_1(X_2) & \varphi_2(X_2) & ... & \varphi_K(X_2) \\ ... & ... & ... & ... & ... \\ \varphi_0(X_n) & \varphi_1(X_n) & \varphi_2(X_n) & ... & \varphi_K(X_n) \end{bmatrix} \tag{14.30}$$

Первый столбец матрицы $\boldsymbol{\Phi}$ состоит из значений функции $\varphi_0(X)$ по всем наблюдениям $i = 1,...n$, то есть по всем точкам $X_i = (X_i^{(1)}, X_i^{(2)}, ... X_i^{(m)})$, $i = 1,...n$.

Во втором столбце матрицы $\boldsymbol{\Phi}$ указаны значения функции $\varphi_1(X)$ по всем наблюдениям $i = 1,...n$.

В третьем и следующих столбцах – значения функций $\varphi_j(X)$, $j = 2,...K$ по всем наблюдениям $i = 1,...n$.

Дополнительно к (14.30) нужны **обозначения для векторов**: истинные значения отклика $Y$ запишем как вектор $\overline{\boldsymbol{Y}}$ размерности $n$; столбцы матрицы $\boldsymbol{\Phi}$ – как векторы $\overline{\boldsymbol{\Phi}}^{(j)}$, $j = 0,..K$ размерности $n$:

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}, \quad \overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} \varphi_0(X_1) \\ \varphi_0(X_2) \\ \dots \\ \varphi_0(X_n) \end{bmatrix}, \quad \overline{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} \varphi_1(X_1) \\ \varphi_1(X_2) \\ \dots \\ \varphi_2(X_n) \end{bmatrix}, \quad \dots\dots\dots \quad \overline{\boldsymbol{\Phi}}^{(K)} = \begin{bmatrix} \varphi_K(X_1) \\ \varphi_K(X_2) \\ \dots \\ \varphi_K(X_n) \end{bmatrix}$$

**Утверждение 2.** Для любого набора данных (14.23), такого, что ранг матрицы $\boldsymbol{\Phi}$ равен $K+1$, обощенный МНК-полином (14.24) **существует** и **является единственным**, а его коэффициенты $b_j, j=1,...K$ **являются решением нормальной системы уравнений**:

$$\frac{\partial S}{\partial b_j} = 0,\ j = 0,...K \tag{14.31}$$

Система (14.31) представляет собой СЛАУ с неизвестными $b_j, j=1,...K$:

$$\begin{bmatrix} (\overline{\boldsymbol{\Phi}}^{(0)}, \overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(0)}, \overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(0)}, \overline{\boldsymbol{\Phi}}^{(K)}) \\ (\overline{\boldsymbol{\Phi}}^{(1)}, \overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(1)}, \overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(1)}, \overline{\boldsymbol{\Phi}}^{(K)}) \\ \dots & \dots & \dots & \dots \\ (\overline{\boldsymbol{\Phi}}^{(K)}, \overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(K)}, \overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(K)}, \overline{\boldsymbol{\Phi}}^{(K)}) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} (\overline{\boldsymbol{Y}}, \overline{\boldsymbol{\Phi}}^{(0)}) \\ (\overline{\boldsymbol{Y}}, \overline{\boldsymbol{\Phi}}^{(1)}) \\ \dots \\ (\overline{\boldsymbol{Y}}, \overline{\boldsymbol{\Phi}}^{(K)}) \end{bmatrix} \tag{14.32}$$

Символом $(*\cdot,*)$ обозначено скалярное произведение в пространстве $R^n$.

Утверждение 2 доказывается аналогично Утверждению 1.

**Пример 3**

В таблице приведены результаты 4-х замеров:

| $X_i$ | 0 | 1 | 3 | 6 |
|---|---|---|---|---|
| $Y_i$ | 0.1 | 0.8 | 3.05 | 5.95 |

Нужно построить линейную зависимость $Y$ от $X$ (без константы) методом наименьших квадратов. Указать, какой функционал будет минимизирован.

**Решение**

1) Количество наблюдений $n=4$. Обобщенный полином запишем в виде $\hat{Y} = b_0 X$. Считаем, что $\varphi_0(X) = X$ и $K=0$.

Параметр полинома должен быть решением задачи оптимизации

$$S(b_0) = \sum_{i=1}^{4} [\, Y_i - b_0 X_i \,]^2 \to \min$$

В данной задаче

$$S(b_0) = [0.1 - b_0 \cdot 0]^2 + [0.8 - b_0 \cdot 1]^2 +$$

$$+ [3.05 - b_0 \cdot 3]^2 + [5.95 - b_0 \cdot 6)]^2$$

Если формулы (14.32) были «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений. В данном случае она состоит из одного уравнения

$$\frac{\partial S}{\partial b_0} = 0.$$

Все-таки о способе записи (14.32) удобнее не забывать.

Вектор истинных значений отклика $\overline{Y}$ и матрица регрессоров $\boldsymbol{\Phi}$ размерности 4×1 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(0) \\ \varphi_0(1) \\ \varphi_0(3) \\ \varphi_0(6) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Единственный столбец матрицы регрессоров соответствует вектору

$$\overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_0$ имеет вид

$$(\overline{\boldsymbol{\Phi}}^{(0)}, \overline{\boldsymbol{\Phi}}^{(0)}) \cdot b_0 = (\overline{\boldsymbol{Y}}, \overline{\boldsymbol{\Phi}}^{(0)})$$

то есть

$$(0 \cdot 0 + 1 \cdot 1 + 3 \cdot 3 + 6 \cdot 6) \cdot b_0 = (0.1 \cdot 0 + 0.8 \cdot 1 + 3.05 \cdot 3 + 5.95 \cdot 6)$$

$$46 \cdot b_0 = 45.65$$

$$b_0 = \frac{45.65}{46} = 0.9923913$$

Следовательно,

$$\hat{Y} = 0.9923913 \cdot X$$

**МНК-прямая (без константы) показана на Рисунке 5.**

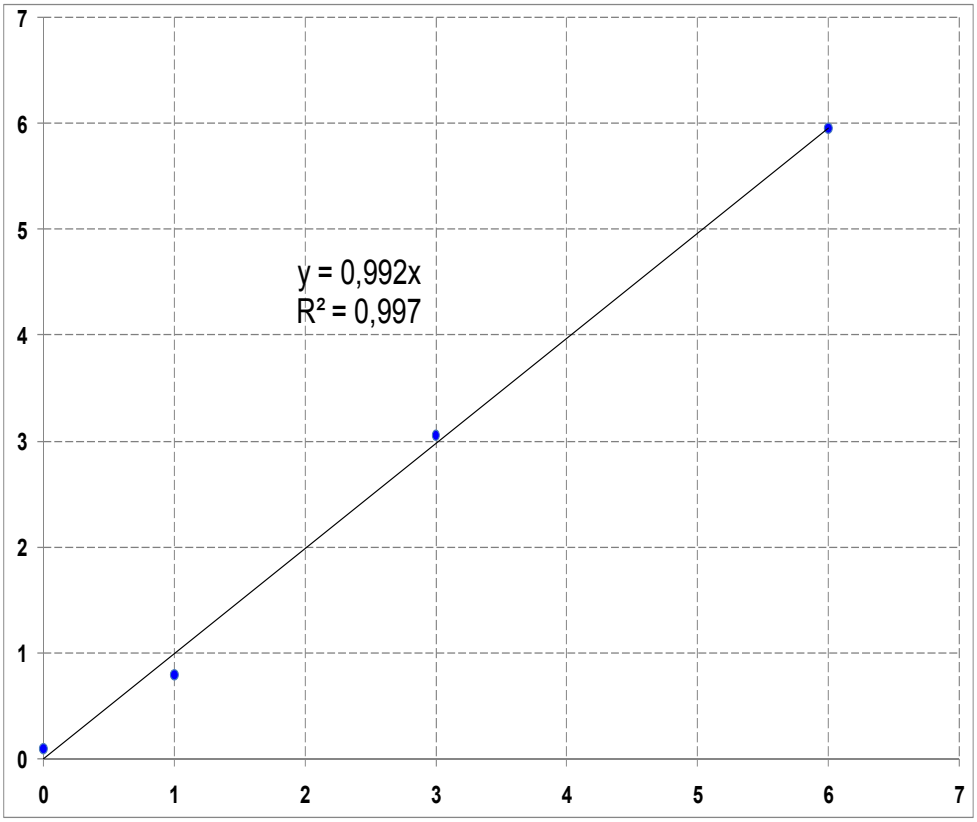


**Рисунок 5**

**Пример 4**

В таблице приведены результаты замеров:

| | | | | |
|---|---|---|---|---|
| $X_i^{(1)}$ | 0 | 1 | 3 | 6 |
| $X_i^{(2)}$ | 2 | 5 | 7 | 8 |
| $Y_i$ | 2.1 | 6.03 | 9.98 | 14.02 |

Нужно построить линейную зависимость $Y$ от $X=(X^{(1)},X^{(2)})$ (без константы) методом наименьших квадратов. Указать, какой функционал будет минимизирован.

**Решение**

1) Количество наблюдений $n=4$. Обобщенный полином запишем в виде $\hat{Y}=b_0X^{(1)}+b_1X^{(2)}$. Считаем, что $\varphi_0(X)=X^{(1)}$, $\varphi_1(X)=X^{(2)}$ и $K=1$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0)=\sum_{i=1}^{4}[\,Y_i-(b_0X_i^{(1)}+b_1X_i^{(2)})]^{\,2}\to\min$$

В данной задаче

$$S(b_0, b_1) = [2.1 - (b_0 \cdot 0 + b_1 \cdot 2)]^2 + [6.03 - (b_0 \cdot 1 + b_1 \cdot 5)]^2 +$$
$$+ [9.98 - (b_0 \cdot 3 + b_1 \cdot 7)]^2 + [14.02 - (b_0 \cdot 6 + b_1 \cdot 8)]^2$$

Решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0.$$

Используя (14.30), нормальную систему уравнений можно выписать быстрее.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{\Phi}$ размерности 4×2 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 2.1 \\ 6.03 \\ 9.98 \\ 14.02 \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(X_1) & \varphi_1(X_1) \\ \varphi_0(X_2) & \varphi_1(X_2) \\ \varphi_0(X_3) & \varphi_1(X_3) \\ \varphi_0(X_4) & \varphi_1(X_4) \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 1 & 5 \\ 3 & 7 \\ 6 & 8 \end{bmatrix}$$

Здесь $X_1 = (0, 2), X_2 = (1, 5), X_3 = (3, 7), X_4 = (6, 8)$.

Столбцы матрицы регрессоров соответствуют векторам

$$\overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix} \qquad \overline{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} 2 \\ 5 \\ 7 \\ 8 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_j$, $j = 0, 1$ имеет вид:

$$\begin{bmatrix} 0+1+9+36 & 2\cdot 0+5\cdot 1+7\cdot 3+8\cdot 6 \\ 2\cdot 0+5\cdot 1+7\cdot 3+8\cdot 6 & 4+25+49+64 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} =$$
$$= \begin{bmatrix} 2.1\cdot 0+6.03\cdot 1+9.98\cdot 3+14.02\cdot 6 \\ 2.1\cdot 2+6.03\cdot 5+9.98\cdot 7+14.02\cdot 8 \end{bmatrix}$$

то есть

$$\begin{bmatrix} 46 & 74 \\ 74 & 142 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \ldots$$

**СЛАУ нужно решить самостоятельно.**

### *Критерии качества МНК-приближения*

**Основными критериями для проверки качества МНК-полинома являются:**

1) **величина функционала** $S$, то есть сумма квадратов остатков

$$S = \sum_{i=1}^{n} [\hat{\varepsilon}_i]^2 = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 \qquad (14.33)$$

(чем она меньше, тем лучше).

2) **коэффициент детерминации** $R^2$, его определением служит формула

$$R^2 = \frac{\sum_{i=1}^{n} [\hat{Y}_i - Y_{сред.}]^2}{\sum_{i=1}^{n} [Y_i - Y_{сред.}]^2} \qquad (14.34)$$

где

$$Y_{сред.} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{\sum_{i=1}^{n} \hat{Y}_i}{n}$$

Областью значений коэффициента $R^2$ является отрезок $[0;1]$. Чем ближе $R^2$ к единице, тем лучше.

3) величина $s$, ее называют **«стандартная ошибка оценки»**, определяют формулой

$$s = \sqrt{\frac{\sum_{i=1}^{n} [\hat{\varepsilon}_i]^2}{n-(K+1)}} \qquad (14.35)$$

Здесь через $K+1$ обозначено число параметров, которые должны быть найдены при построении МНК-полинома. Чем меньше $s$, тем лучше.

4) визуальный и статистический анализ остатков $\hat{\varepsilon}_i$, $i = 1,...n$. Например, графики зависимости остатков от фактора $X$ или отклика $Y$, а также нормальный вероятностный график остатков.

Если графики остатков хаотичны, значит, функциональная зависимость отклика $Y$ от фактора $X$ описывается МНК-полиномом достаточно полно (не учтено и не может быть учтено только случайное поведение $Y$).

Если графики остатков показывают какую-либо закономерность, желательно пересмотреть аппроксимирующее уравнение (полином, обобщенный полином) и включить в него недостающие компоненты.

**Анализ качества построенного МНК-полинома (обощенного полинома) начинают с анализа остатков.**

## Модуль 14.2 – Практикум по теме «Методы приближения функций и обработки экспериментальных данных, основанные на решении задач оптимизации»

### Пример 1 – наилучшие приближения в конечномерных классах гильбертовых пространств (наилучшие среднеквадратичные приближения)

1) Постройте функцию $\varphi(x)$ – наилучшее приближение функции $f(x)=x^{12}$

в классе **полиномов степени не выше** $2$

в метрике гильбертова пространства

интегрируемых с квадратом на отрезке $[0;1]$ функций

со скалярным произведением

$$(f,g) = \int_0^1 f(x)g(x)dx\,.$$

2) Укажите погрешность найденного наилучшего приближения.

3) Используя on-line сервис или математический пакет, постройте на отрезке $x\in[0;1]$, указанном в условии задачи, график $f(x)=x^{12}$ и ее наилучшего приближения $\varphi(x)$

(две функции на одном графике).

**Решение**

**О терминах данной задачи**

В данной задаче нужно приблизить функцию $f(x)=x^{12}$

выражением вида $\alpha_0+\alpha_1\cdot x+\alpha_2\cdot x^2$

(полиномом степени не выше $2$).

Функция $f(x)=x^{12}$ является элементом гильбертова пространства $L_2[0;\ 1]$

интегрируемых с квадратом на отрезке $[0;1]$ функций

со скалярным произведением

$$(f,g)_{L_2[0;1]} = \int_0^1 f(x)g(x)dx\,.$$

Норма любого элемента $f\in L_2[0;1]$ (из этого пространства) определяется через скалярное произведение:

$$\| f \|_{L_2[0;1]} = \sqrt{\int_0^1 f^2(x)\,dx}$$

Расстояние между элементами $f,\ g \in L_2[0;1]$ в этом пространстве определяют через норму разности элементов:

$$\rho(f,g) = \| f - g \|_{L_2[0;1]} = \sqrt{\int_0^1 (f(x)-g(x))^2\,dx}$$

Наилучшее приближение необходимо найти в подпространстве, элементами которого являются полиномы степени не выше $2$, такие элементы записываются в виде

$$\varphi(x) = \alpha_0 + \alpha_1 \cdot x + \alpha_2 \cdot x^2$$

Функции

$$\varphi_0(x) = 1,\ \varphi_1(x) = x,\ \varphi_2(x) = x^2$$

можно рассматривать как базис подпространства,

а числа $\alpha_0, \alpha_1, \alpha_2$ - как коэффициенты при элементах базиса.

Подпространство обозначим $K_3$, его можно записывать в виде

$$K_3 = \{\, \alpha_0 + \alpha_1 \cdot x + \alpha_2 \cdot x^2 \mid \alpha_i \in R,\quad i = 0,1,2 \,\}$$

его размерность равна $3$.

Так как каждая из базисных функций $\varphi_0(x) = 1,\ \varphi_1(x) = x,\ \varphi_2(x) = x^2$

является элементом гильбертова пространства $L_2[0;\ 1]$,

любой элемент $K_3$ (полином степени не выше $2$)

также является элементом $L_2[0;\ 1]$.

Подпространство $K_3$ является подпространством $L_2[0;\ 1]$.

**О постановке задачи оптимизации**

**Наилучшим приближением для $f(x) = x^{12}$ в подпространстве $K_3$ является такой элемент $\varphi \in K_3$ (такой полином степени не выше $2$), расстояние до которого является минимальным**:

$$\rho(f,\varphi) = \| f-\varphi \|_{L_2[0;1]} = \sqrt{\int_0^1 (f(x)-\varphi(x))^2\, dx} \to \min$$

Для отыскания такого полинома вводится функционал

$$S(\alpha_0,\alpha_1,.\alpha_2) = (f-\varphi, f-\varphi)_{L_2[0;1]} = \int_0^1 (f-(\alpha_0+\alpha_1 x+\alpha_2 x^2))^2\, dx$$

и ставится задача минимизации

$$S(\alpha_0,\alpha_1,\alpha_2) \underset{\alpha\in R^3}{\to} \min$$

**Функционал $S(\alpha_0,\alpha_1,.\alpha_2)$ имеет следующий смысл: это квадрат расстояния**

**от элемента $f(x)=x^{12}$ до элемента $\varphi(x)=\alpha_0+\alpha_1\cdot x+\alpha_2\cdot x^2$**

**по правилам гильбертова пространства $L_2[0;\ 1]$.**

Величина $S(\alpha_0,\alpha_1,.\alpha_2)$ **за счет подбора чисел** $\alpha_0,\alpha_1,\alpha_2$ должна стать **минимальной**.

## О решении задачи оптимизации

Для решения задачи оптимизации записывают нормальную систему уравнений

$$\frac{\partial S}{\partial \alpha_i}=0,\ \ i=0,1,2$$

**В соответствии с Утверждением 1 (Модуль 14.2, часть II) система записывается в виде СЛАУ**

$$\begin{bmatrix} \int_0^1 dx & \int_0^1 xdx & \int_0^1 x^2dx \\ \int_0^1 xdx & \int_0^1 x^2dx & \int_0^1 x^3dx \\ \int_0^1 x^2dx & \int_0^1 x^3dx & \int_0^1 x^4dx \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \int_0^1 x^{12}\cdot 1\cdot dx \\ \int_0^1 x^{12}\cdot x\cdot dx \\ \int_0^1 x^{12}\cdot x^2\cdot dx \end{bmatrix}$$

потому что в соответствии с правилами вычисления скалярных произведений в $L_2[0;\ 1]$

$$(\varphi_0, \varphi_0)_{L_2[0;1]} = \int_0^1 1 \cdot 1 \cdot dx = 1$$

$$(\varphi_0, \varphi_1)_{L_2[0;1]} = \int_0^1 1 \cdot x \cdot dx = \frac{1}{2}$$

$$(\varphi_0, \varphi_2)_{L_2[0;1]} = \int_0^1 1 \cdot x^2 \cdot dx = \frac{1}{3}$$

$$(\varphi_1, \varphi_1)_{L_2[0;1]} = \int_0^1 x \cdot x \cdot dx = \frac{1}{3}$$

$$(\varphi_1, \varphi_2)_{L_2[0;1]} = \int_0^1 x \cdot x^2 \cdot dx = \frac{1}{4}$$

$$(\varphi_2, \varphi_2)_{L_2[0;1]} = \int_0^1 x^2 \cdot x^2 \cdot dx = \frac{1}{5}$$

$$(f, \varphi_0)_{L_2[0;1]} = \int_0^1 x^{12} \cdot 1 \cdot dx = \frac{1}{13}$$

$$(f, \varphi_1)_{L_2[0;1]} = \int_0^1 x^{12} \cdot x \cdot dx = \frac{1}{14}$$

$$(f, \varphi_2)_{L_2[0;1]} = \int_0^1 x^{12} \cdot x^2 \cdot dx = \frac{1}{15}$$

В данном случае СЛАУ размерности 3*3 принимает вид

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{13} \\ \frac{1}{14} \\ \frac{1}{15} \end{bmatrix}$$

Нужно найти решение $\alpha_0, \alpha_1, \alpha_2$, и эти числа определят для $f(x) = x^{12}$ элемент наилучшего приближения $\varphi(x) = \alpha_0 + \alpha_1 \cdot x + \alpha_2 \cdot x^2$

в классе полиномов степени не выше $2$

в метрике гильбертова пространства $L_2[0;\ 1]$

**О погрешности наилучшего приближения**

**Погрешностью** приближения элемента $f$ гильбертова пространства $L_2[0;1]$ элементом $\varphi$ конечномерного подпространства $K_3 \subset L_2[0;1]$ является элемент $z \in L_2[0;1]$, определяемый как

$$z = f - \varphi$$

**Качество приближения** характеризуется **нормой погрешности**, то есть значением

$$\| z \|_{L_2[0;1]}$$

которое в данном случае является **корнем квадратным из минимального значения функционала** $S(\alpha_0, \alpha_1, \alpha_2)$:

$$\| z \|_{L_2[0;1]} = \| f - \varphi \|_{L_2[0;1]} = \sqrt{S(\alpha_0, \alpha_1, \alpha_2)}.$$

Если решение СЛАУ найдено, значение функционала $S(\alpha_0, \alpha_1, \alpha_2)$ можно найти по формуле

$$S(\alpha_0, \alpha_1, \alpha_2) =$$

$$= (f, f)_{L_2[0;1]} - 2\sum_{i=0}^{2} \alpha_i (f, \varphi_i)_{L_2[0;1]} + \sum_{i=0}^{2} \sum_{j=0}^{2} \alpha_i \cdot \alpha_j \ (\varphi_i, \varphi_j)_{L_2[0;1]}$$

и узнать, велика или мала погрешность.

Смысл **нормы погрешности** в метрике пространства $L_2[0;1]$ таков: **норма погрешности** «отвечает» за абсолютную (без учета знака) **величину площади, заключенной между графиками функции** $f$ **и наилучшего приближения** $\varphi$ **на участке** $x \in [0;1]$.

Если площадь, заключенная между графиками, велика, норма погрешности в метрике пространства $L_2[0;1]$ велика. Если площадь, заключенная между графиками, мала, норма погрешности в метрике пространства $L_2[0;1]$ мала.

«Локальные всплески» (существенные отличия графиков $f$ и $\varphi$ на небольших участках отрезка $x \in [0;1]$, если такие отличия будут, или в отдельных точках отрезка) в метрике пространства $L_2[0;1]$ «во внимание не принимаются».

**При построении графиков** $f$ **и** $\varphi$ **обратите внимание на абсолютную величину площади, заключенной между графиками**.

**Пример 2 – наилучшие равномерные приближения, экономизация степенных рядов**

1) Проведите экономизацию полинома, построенного для вычисления функции $f(x)=e^x$ на отрезке $x\in[-1;1]$ на основе формулы Тейлора по степеням $x$, усеченной до степени $n=4$ включительно (то есть степень остатка не менее $n+1=5$).

2) Оцените погрешность применения на отрезке $x\in[-1;1]$

экономизированного полинома.

3) Сравните погрешность применения экономизированного полинома

с погрешностью применения на отрезке $x\in[-1;1]$ «другой» формулы Тейлора,

изначально усеченной до той степени,

которую имеет экономизированный полином.

4) Проведите повторную экономизацию (то есть еще одно понижение степени уже экономизированного полинома) и анализ погрешности применения нового полинома.

**Решение**

**Шаг 1**

Для функции $e^x$ запишем формулу Тейлора

$$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}e^{\xi},\ \xi\in[0;x]$$

Остаток представлен в форме Лагранжа.

**Шаг 2**

С целью приближенного вычисления $e^x$ используем полином $S_4(x)$ степени $n=4$, полученный **усечением формулы**, то есть

$e^x\approx S_4(x)$, где

$$S_4(x)=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}$$

**Шаг 3**

Погрешность применения $S_4(x)$ в точке $x$ (то есть погрешность усечения) составит

$$E(x) = e^x - S_4(x) = \frac{x^5}{5!} e^{\xi},\ \xi \in [0; x]$$

При $x \in [-1; 1]$ верна оценка

$$\max_{\xi \in [-1;1]} \left| e^{\xi} \right| \le e$$

поэтому для погрешности усечения формулы Тейлора верна оценка

$$\max_{x \in [-1;1]} \left| E(x) \right| \le \frac{e}{5!} = \frac{e}{120}$$

то есть

$$\max_{x \in [-1;1]} \left| E(x) \right| \le 0.02265$$

**Шаг 4**

Чтобы провести **экономизацию** полинома $S_4(x)$ при $x \in [-1; 1]$,

нужен **полином Чебышёва степени** $n = 4$, **наименее уклоняющийся от нуля на отрезке** $[-1; 1]$ **в классе полиномов степени** $n = 4$ **со старшим коэффициентом, равным единице** *(название у него такое, короче нельзя).*

Такой полином обозначим $T_4(x)$.

**Комментарий**

То, что выше сказано о $T_4(x)$ текстом, записывают так:

$T_4(x)$ есть решение задачи

$$\max_{x \in [-1;1]} \left| P_4(x) \right| \to \min$$

когда в качестве $P_4(x)$ рассматривают все полиномы степени $n = 4$, у которых коэффициент при $x^4$ равен единице.

**Шаг 5**

Чтобы записать $T_4(x)$, используем сведения о его корнях:

$$x_s = \cos\left( \frac{\pi}{2 \cdot 4} (1 + 2s) \right), \quad s = 0,...3$$

и учтем, что старший коэффициент полинома равен единице.

Поэтому

$$T_4(x) = (x - \cos\frac{\pi}{8})(x - \cos\frac{3\pi}{8})(x - \cos\frac{5\pi}{8})(x - \cos\frac{7\pi}{8})$$

После преобразований получим

$$T_4(x) = \left( x^2 - \cos^2\left(\frac{\pi}{8}\right)\right) \cdot \left( x^2 - \cos^2\left(\frac{3\pi}{8}\right)\right) = x^4 - x^2 + \frac{1}{8}$$

**Шаг 6**

Запишем максимальное по модулю значение, которое принимает на отрезке $[-1;1]$ полином Чебышёва $T_4(x)$:

$$\max_{x\in[-1;1]} \left| T_4(x) \right| = \frac{1}{2^{4-1}} = \frac{1}{2^3} = \frac{1}{8}$$

**Шаг 7**

Проведем **экономизацию** полинома $S_4(x)$ при $x \in [-1;1]$.

Для этого в формуле полинома $S_4(x)$ заменим $x^4$ полиномом $\{ x^4 - T_4(x) \}$.

Полином, полученный после замены, обозначим $S_3^*(x)$:

$$S_3^*(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{\{ x^4 - T_4(x)\}}{4!}$$

Полином $S_3^*(x)$ можно записывать разными способами, перечислим их:

$$S_3^*(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{\{ x^4 - \left( x^4 - x^2 + \frac{1}{8}\right)\}}{4!}$$

(здесь видно, как получен $S_3^*(x)$)

$$S_3^*(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{\{ x^2 - \frac{1}{8}\}}{4!}$$

(здесь видно, что $S_3^*(x)$ не содержит степени $4$ и отличается от $S_4(x)$ только последним слагаемым)

$$S_3^*(x) = \left(1 - \frac{1}{8 \cdot 24}\right) + x + \left(\frac{1}{2} + \frac{1}{24}\right) \cdot x^2 + \frac{x^3}{6}$$, то есть

$$S_3^*(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{x^3}{6}$$

(здесь видно, как нужно вычислять (программировать) $S_3^*(x)$)

**Шаг 8**

**Погрешность применения $S_3^*(x)$ в точке $x$ для вычисления $e^x$ (по определению) составит**

$$E^*(x) = e^x - S_3^*(x)$$

По Утверждению 3, для погрешности при $x \in [-1;1]$ верна оценка

$$\max_{x \in [-1;1]} \left| E^*(x) \right| \le \frac{e}{5!} + \frac{1}{2^3 \cdot 4!},$$

то есть

$$\max_{x \in [-1;1]} \left| E^*(x) \right| \le \frac{e}{120} + \frac{1}{192} = 0.02786$$

Здесь $\frac{e}{120}$ – оценка погрешности усечения, то есть замены $e^x$ полиномом $S_4(x)$;

$\frac{1}{2^3 \cdot 4!}$ – погрешность экономизации, то есть замены $S_4(x)$ полиномом $S_3^*(x)$.

**Комментарии и выводы**

Как и следовало ожидать, применение экономизированного $S_3^*(x)$ имеет несколько большую погрешность, чем применение усеченной формулы Тейлора $S_4(x)$

$$\max_{x \in [-1;1]} \left| e^x - S_4(x) \right| \le 0.02265$$

$$\max_{x\in[-1;1]}\left|e^x - S_3^*(x)\right| \le 0.02786$$

Напомним: экономизация не уменьшает, а **увеличивает** погрешность применения формулы, но снижает **вычислительную погрешность** (которая здесь не показана), потому что вычисляются полиномы меньших степеней.

**Шаг 9**

**Сравним, что в данном случае полезнее:**

использовать экономизированный полином степени $3$

$$S_3^*(x) = \frac{191}{192} + x + \frac{13}{24}\cdot x^2 + \frac{x^3}{6}$$

или записать формулу Тейлора в виде

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!}e^{\xi}, \qquad \xi\in[0;x]$$

и применять ее усечение до степени $3$.

Усечение до степени $3$ имеет вид

$$S_3(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

Погрешность применения полинома $S_3(x)$ в точке $x$ составит

$$E_3(x) = e^x - S_3(x) = \frac{x^4}{4!}e^{\xi},\ \xi\in[0;x]$$

При $x\in[-1;1]$ верна оценка

$$\max_{x\in[-1;1]}\left|E_3(x)\right| \le \frac{e}{4!} = \frac{e}{24}$$

то есть

$$\max_{x\in[-1;1]}\left|E_3(x)\right| \le 0.11326$$

Очевидно, что на отрезке $[-1;1]$ экономизированный полином третьей степени $S_3^*(x)$ даст погрешность в три (почти в четыре) раза меньше , чем усеченная до третьей степени формула Тейлора $S_3(x)$:

$$\max_{x\in[-1;1]}\left|e^x - S_3^*(x)\right| \le 0.02786$$

$$\max_{x\in[-1;1]}\left|e^x - S_3(x)\right| \le 0.11326$$

**Выводы**

$S_3^*(x)$ лучше, чем $S_3(x)$.

**Шаг 10**

Проведем **повторную экономизацию,** то есть понизим степень уже экономизированного полинома

$$S_3^*(x) = \frac{191}{192} + x + \frac{13}{24}\cdot x^2 + \frac{x^3}{6}$$

используя **наилучшее равномерное приближение** полинома $x^3$ на отрезке $[-1;1]$ полиномом меньшей степени.

Потребуется **полином Чебышёва степени** $n=3$, **наименее уклоняющийся от нуля на отрезке** $[-1;1]$ **в классе полиномов степени** $n=3$ **со старшим коэффициентом, равным единице** *(как раньше: название такое, сократить нельзя)*.

Такой полином обозначим $T_3(x)$. Он имеет вид

$$T_3(x) = (x - \cos\frac{\pi}{6})(x - \cos\frac{3\pi}{6})(x - \cos\frac{5\pi}{6})$$

то есть

$$T_3(x) = \left(x^2 - \cos^2\left(\frac{\pi}{6}\right)\right)\cdot x = x^3 - \frac{3}{4}\cdot x$$

Запишем максимальное по модулю значение, которое принимает на отрезке $[-1;1]$ полином $T_3(x)$:

$$\max_{x\in[-1;1]}\left|T_3(x)\right| = \frac{1}{2^{3-1}} = \frac{1}{2^2} = \frac{1}{4}$$

Проведем **экономизацию** $S_3^*(x)$ при $x\in[-1;1]$.

Для этого в формуле полинома $S_3^*(x)$ заменим $x^3$ полиномом $\{x^3 - T_3(x)\}$.

Полином, полученный после замены, обозначим $S_2^{**}(x)$:

$$S_2^{**}(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{\{x^3 - T_3(x)\}}{6}$$

то есть

$$S_2^{**}(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{\{x^3 - \left(x^3 - \frac{3}{4} \cdot x\right)\}}{6}$$

или

$$S_2^{**}(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{3}{4 \cdot 6} \cdot x$$

$$S_2^{**}(x) = \frac{191}{192} + \frac{9}{8} \cdot x + \frac{13}{24} \cdot x^2$$

**Шаг 11**

**Исследуем погрешность применения новой формулы.**

**Погрешность применения $S_2^{**}(x)$ в точке $x$ для вычисления $e^x$ (по определению) составит**

$$E^{**}(x) = e^x - S_2^{**}(x)$$

Запишем эту погрешность, вычитая и добавляя удобные для анализа величины:

$$E^{**}(x) = e^x - S_2^{**}(x) = \underbrace{e^x - S_4(x)}_{\substack{\textit{погрешность} \\ \textit{усечения}}} + \underbrace{S_4(x) - S_3^*(x)}_{\substack{\textit{погрешность} \\ \textit{экономизации}}} + \underbrace{S_3^*(x) - S_2^{**}(x)}_{\substack{\textit{погрешность} \\ \textit{повторной} \\ \textit{экономизации}}}$$

Для первых двух компонент погрешности оценки уже получены (см. Утверждение 3).

Напомним, что $S_3^*(x)$ и $S_2^{**}(x)$ отличаются только последним слагаемым:

$$S_3^*(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{x^3}{6}$$

$$S_2^{**}(x) = \frac{191}{192} + x + \frac{13}{24} \cdot x^2 + \frac{\{x^3 - T_3(x)\}}{6}$$

Поэтому

$$S_3^*(x) - S_2^{**}(x) = T_3(x) \cdot \frac{1}{6}$$

и на отрезке $[-1;1]$ справедлива оценка

$$\max_{x\in[-1;1]} \left| S_3^*(x) - S_2^{**}(x) \right| = \frac{1}{6} \cdot \max_{x\in[-1;1]} \left| T_3(x) \right| = \frac{1}{6 \cdot 4}$$

**Таким образом, для погрешности применения $S_2^{**}(x)$ в точке $x$ для вычисления $e^x$ при $x \in [-1;1]$** доказана оценка

$$\max_{x\in[-1;1]} \left| E^{**}(x) \right| \le \frac{e}{5!} + \frac{1}{2^3 \cdot 4!} + \frac{1}{6 \cdot 4}.$$

то есть

$$\max_{x\in[-1;1]} \left| e^x - S_2^{**}(x) \right| \le 0.06953$$

**Комментарии и выводы**

Повторная экономизация (степень полинома 2) приводит к заметному росту начальной погрешности усечения:

$$\max_{x\in[-1;1]} \left| e^x - S_2^{**}(x) \right| \le 0.06953$$

что намного хуже, чем усеченная до степени 4 формула Тейлора

$$\max_{x\in[-1;1]} \left| e^x - S_4(x) \right| \le 0.02265$$

намного хуже, чем экономизированный на ее основе полином степени 3

$$\max_{x\in[-1;1]} \left| e^x - S_3^*(x) \right| \le 0.02786$$

но не хуже, а лучше формулы Тейлора, усеченной до степени 3:

$$\max_{x\in[-1;1]} \left| e^x - S_3(x) \right| \le 0.11326$$

**Выбор формулы для вычисления экспоненты остается на усмотрение исследователя.**

## ЛИТЕРАТУРА

### а) литература по тематическому блоку


1. Бабенко К.И. Основы численного анализа. Москва – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2002. – 848 с.
2. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – 7-е изд. М.: БИНОМ. Лаборатория знаний, 2011. – 636 с.
3. Вержбицкий В.М. Основы численных методов. – М.: Высшая школа, 2002. – 840 с.
4. Гутер Р.С., Овчинский Б.В. Элементы численного анализа и математической обработки результатов опыта. М.: Наука, 1970. – 432 с.
5. Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа. М.: Наука, 1967. – 368 с.
6. Мак-Кракен Д., Дорн У. Численные методы и программирование на ФОРТРАНе. М.: Мир, 1977.
7. Марчук Г.И. Методы вычислительной математики. М.: Наука, 1980. – 536 с.
8. Разностные схемы в задачах газовой динамики на неструктурированных сетках / Под ред. проф. В.Н. Емельянова, д.ф.-м.н. К.Н. Волкова. М.: ФИЗМАТЛИТ, 2015. – 416 с.
9. Стронгина Н.Р., Баркалов К.А. Численные методы. Семестр 8. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Ид.н. 831Е.14.08.
10. Перов А.А., Протогенов А.П. Численные методы в физических исследованиях. – Н. Новгород: Нижегородский университет, 2019. – 69 с.


### б) литература об организации учебного процесса по дисциплине


11. Садовничий В.А. Международный форум «Университеты, общество и будущее человечества». Доклад ректора МГУ имени М.В. Ломоносова академика В.А. Садовничего на Международном форуме «Университеты, общество и будущее человечества» 25 марта 2019 года. М.: Издательство Московского университета, 2019. – 36 с.
12. Высокопроизводительные параллельные вычисления. 100 заданий для расширенного лабораторного практикума. М.: ФИЗМАТЛИТ, 2018. – 248 с.
13. Программы дисциплин по направлению «Прикладная математика и информатика». Учебно-методическое объединение Университетов. Учебно-методический совет по прикладной математике и информатике. М.: Изд-во факультета ВМиК МГУ, 2002. С. 59 – 62.
14. Стронгина Н.Р. Цифровизация и качество обучения на примере фундаментальной дисциплины «Численные методы» // Научные вести. 2021. №2 (31). – С. 85 – 103.

Наталья Романовна **Стронгина**

# КУРС «ЧИСЛЕННЫЕ МЕТОДЫ»

## Методы приближения функций и обработки экспериментальных данных, основанные на решении задач оптимизации

**(Модуль 14.2)**

***Учебно-методическое пособие***

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского».
603950, Нижний Новгород, пр. Гагарина, 23.