

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский университет ИТМО»  
(Университет ИТМО)

Факультет систем управления и робототехники

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1  
по дисциплине  
*«Теория идентификации»*

по теме:  
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Студенты:

*Группа № R3435*

*Группа № R3441*

*Группа № R3443*

Вариант №25

*Зыкин Л. В.*

*Алёхова М. С.*

*Шилова Д. Р.*

Предподаватель:

*доцент*

*Ведяков А. А.*

Санкт-Петербург  
2025

# 1 МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

## 1.1 Цель работы и исходные данные

Целью работы является оценивание параметров моделей линейной регрессии методом наименьших квадратов, анализ качества аппроксимации и проверка гипотез по данным варианта №25.

## 1.2 Краткие теоретические сведения

Рассматривается модель линейной регрессии

$$y(k) = x_1(k) \theta_1 + x_2(k) \theta_2 + \cdots + x_n(k) \theta_n + v(k), \quad (1)$$

где  $\theta = [\theta_1, \theta_2, \dots, \theta_n]^\top$  — вектор неизвестных параметров,  $v(k)$  — аддитивная помеха. В матричной форме:

$$Y = X\theta + V, \quad (2)$$

где  $Y = [y(1), y(2), \dots, y(N)]^\top$ ,  $V = [v(1), v(2), \dots, v(N)]^\top$ , а матрица регрессоров

$$X = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_n(N) \end{bmatrix}. \quad (3)$$

Оценка МНК минимизирует функционал  $J(\theta) = \|Y - X\theta\|_2^2$  и имеет вид:

$$\hat{\theta}_{\text{LSQ}} = (X^\top X)^{-1} X^\top Y. \quad (4)$$

Вектор ошибок оценивания:

$$e = Y - X\hat{\theta}, \quad (5)$$

а сумма квадратов ошибок:

$$\text{SSE} = \sum_{k=1}^N e^2(k) = e^\top e. \quad (6)$$

## 1.3 Задание 1

### 1.3.1 Постановка

По данным структур zad11 и zad12 (поля  $x_1, x_2, x_3, y$ ) требуется: оценить параметры  $\theta_1, \theta_2, \theta_3$ ; построить графики исходного сигнала  $y$  и оценки  $\hat{y}$ ; построить график ошибки  $e$ ; сделать выводы о несмещенности и эффективности.

### 1.3.2 Методика

Формируется матрица регрессоров без свободного члена:

$$X = \begin{bmatrix} x_1(1) & x_2(1) & x_3(1) \\ x_1(2) & x_2(2) & x_3(2) \\ \vdots & \vdots & \vdots \\ x_1(N) & x_2(N) & x_3(N) \end{bmatrix}, \quad (7)$$

где  $N = 1000$  — количество измерений. Модель имеет вид:

$$y(k) = x_1(k)\theta_1 + x_2(k)\theta_2 + x_3(k)\theta_3 + v(k), \quad (8)$$

или в матричной форме  $Y = X\theta + V$ . Параметры оцениваются по формуле:

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix}. \quad (9)$$

Оценка выходного сигнала:

$$\hat{y}(k) = x_1(k)\hat{\theta}_1 + x_2(k)\hat{\theta}_2 + x_3(k)\hat{\theta}_3, \quad (10)$$

а ошибка оценивания:

$$e(k) = y(k) - \hat{y}(k). \quad (11)$$

По теории МНК оценка  $\hat{\theta}$  является состоятельной, несмещённой и эффективной при выполнении условий для помехи  $v(k)$ :

$$\mathbb{E}\{v(k)\} = 0, \quad \mathbb{E}\{x_i(k)v(k)\} = 0, \quad (12)$$

$$\text{Var}\{v(k)\} = \sigma^2 = \text{const}, \quad \mathbb{E}\{v(k)v(k-\tau)\} = 0, \quad \tau \neq 0. \quad (13)$$

Поскольку ошибка  $e(k) = y(k) - \hat{y}(k)$  совпадает с оценкой помехи, анализ её статистических свойств позволяет судить о качестве модели и параметров: несмещенность соответствует нулевому среднему, отсутствие систематической корреляции с регрессорами и автокорреляции подтверждает эффективность и состоятельность оценки.

### 1.3.3 Результаты

Результаты оценивания представлены на рисунках 1–4:

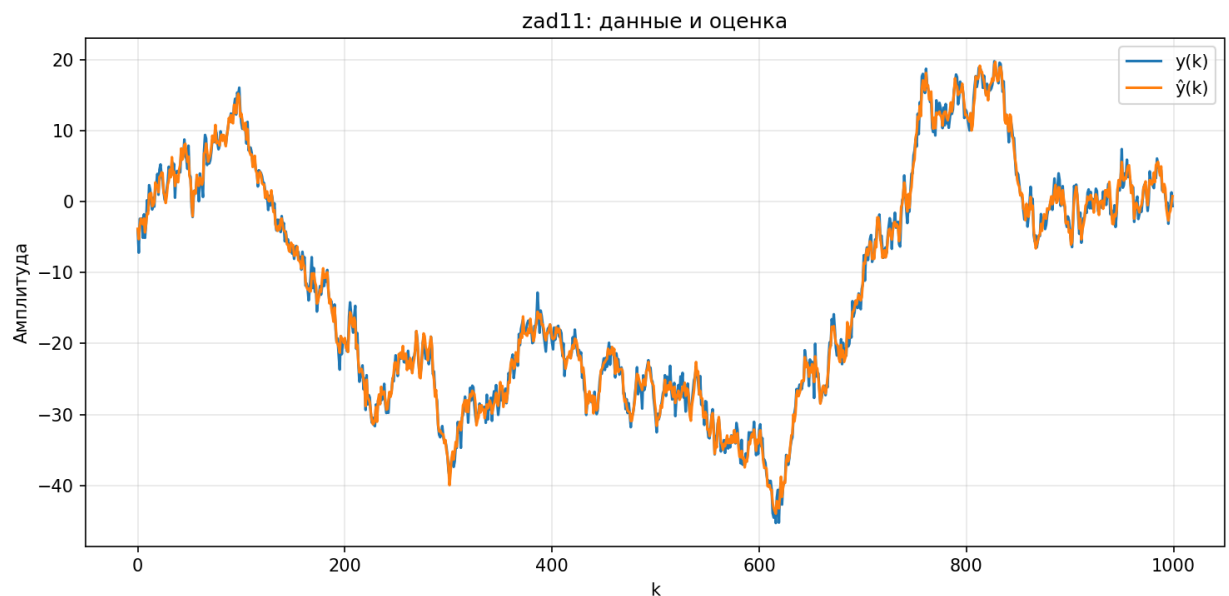


Рисунок 1 — Задание 1: данные и оценка  $\hat{y}$  для zad11

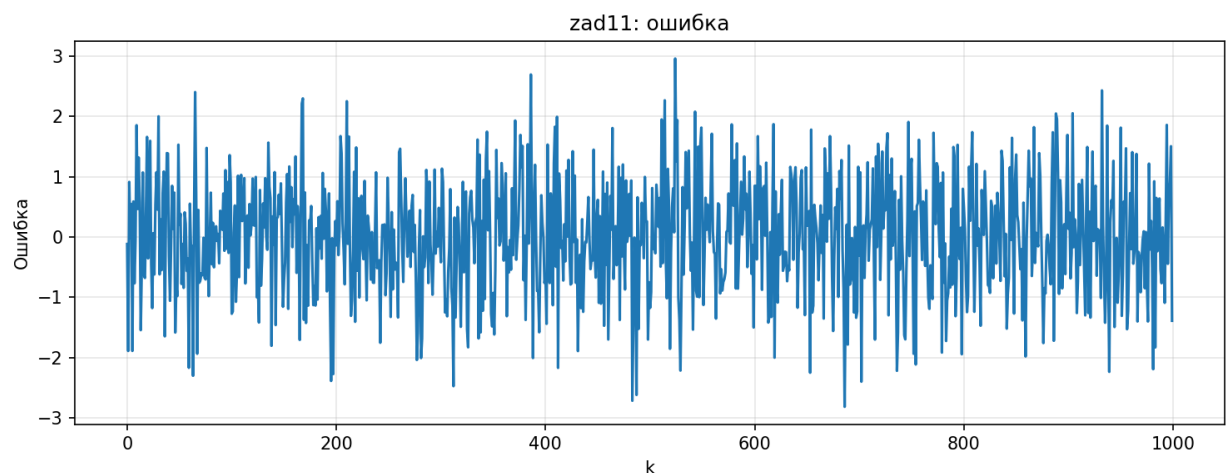


Рисунок 2 — Задание 1: ошибка  $e(k)$  для zad11

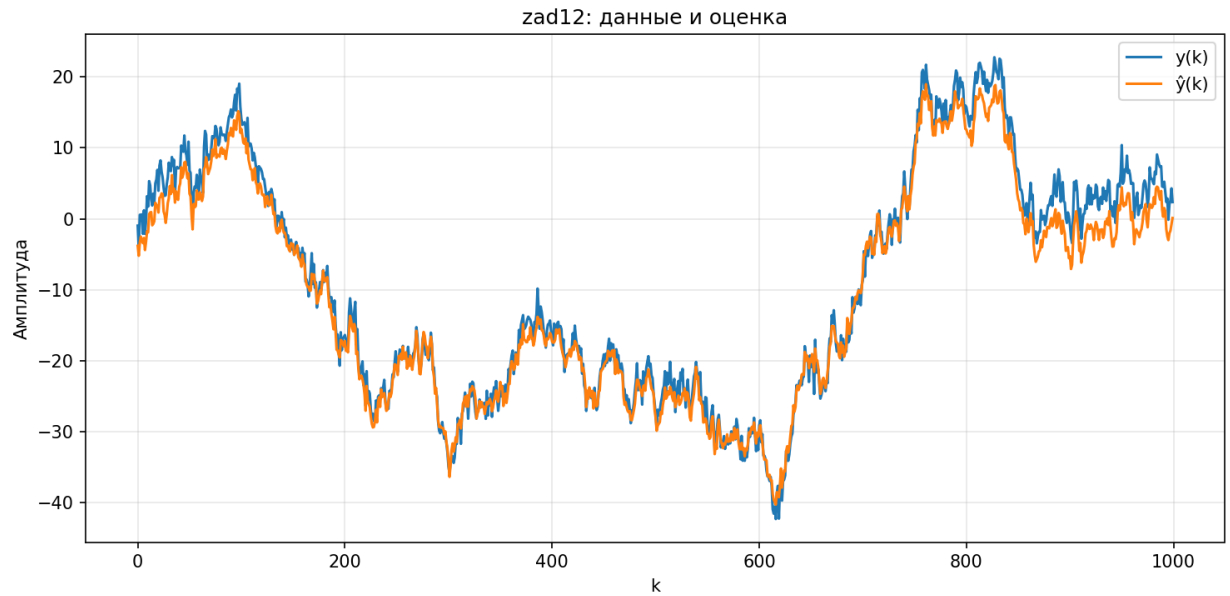


Рисунок 3 — Задание 1: данные и оценка  $\hat{y}$  для zad12

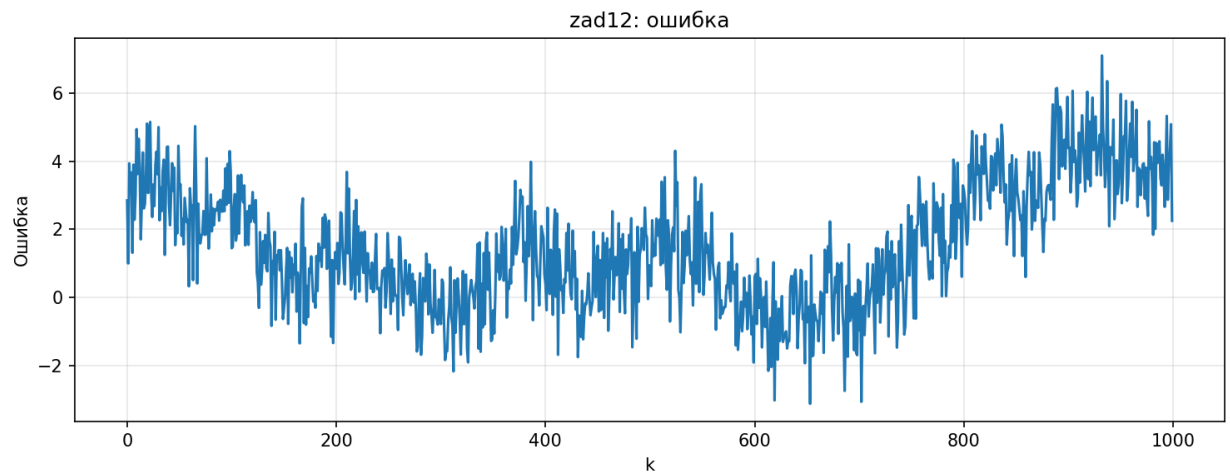


Рисунок 4 — Задание 1: ошибка  $e(k)$  для zad12

### 1.3.4 Выводы

Оцененные параметры: для zad11  $\hat{\theta} = [-2.014, 1.999, 3.999]^T$ , для zad12  $\hat{\theta} = [-1.953, 2.233, 3.575]^T$ .

Численные показатели свойств ошибки  $e(k)$ :

- **zad11**:  $\bar{e} \approx 0.0095$ ;  $\text{Var}(e) \approx 0.906$ ;  $\text{corr}(e, x_1) \approx -0.00$ ,  $\text{corr}(e, x_2) \approx -0.002$ ,  $\text{corr}(e, x_3) \approx 0.010$ ;  $\rho_e(1) \approx 0.013$ . Показатели близки к нулю, что согласуется с условиями  $\mathbb{E}\{v\} = 0$ ,  $\mathbb{E}\{x_i v\} = 0$  и отсутствием автокорреляции.

- **zad12**:  $\bar{e} \approx 1.495$ ;  $\text{Var}(e) \approx 3.187$ ;  $\text{corr}(e, x_1) \approx 0.00$ ,  $\text{corr}(e, x_2) \approx -0.198$ ,  $\text{corr}(e, x_3) \approx 0.803$ ;  $\rho_e(1) \approx 0.716$ . Ненулевое среднее, высокая корреляция с  $x_3$  и заметная автокорреляция указывают на нарушение предпосылок МНК и/или неполноту модели для этого набора.

## 1.4 Задание 2

### 1.4.1 Постановка

Для наборов zad21 и zad22 (поля T, V) проверяются две гипотезы:

$$\text{H1: } V = bT + c, \quad (14)$$

$$\text{H2: } V = aT^2 + bT + c, \quad (15)$$

где  $T$  — температура,  $V$  — объём,  $a, b, c$  — неизвестные параметры.

### 1.4.2 Методика

Для гипотезы H1 матрица регрессоров имеет вид:

$$X_1 = \begin{bmatrix} T(1) & 1 \\ T(2) & 1 \\ \vdots & \vdots \\ T(N) & 1 \end{bmatrix}, \quad (16)$$

а вектор параметров  $\theta_1 = [b, c]^T$ . Для гипотезы H2:

$$X_2 = \begin{bmatrix} T^2(1) & T(1) & 1 \\ T^2(2) & T(2) & 1 \\ \vdots & \vdots & \vdots \\ T^2(N) & T(N) & 1 \end{bmatrix}, \quad (17)$$

с вектором параметров  $\theta_2 = [a, b, c]^T$ . Оценки параметров вычисляются по формулам:

$$\hat{\theta}_1 = (X_1^T X_1)^{-1} X_1^T V, \quad (18)$$

$$\hat{\theta}_2 = (X_2^T X_2)^{-1} X_2^T V. \quad (19)$$

Оценки объёма:

$$\hat{V}_1(k) = \hat{b}T(k) + \hat{c}, \quad (20)$$

$$\hat{V}_2(k) = \hat{a}T^2(k) + \hat{b}T(k) + \hat{c}, \quad (21)$$

а суммы квадратов ошибок:

$$\text{SSE}_{H1} = \sum_{k=1}^N (V(k) - \hat{V}_1(k))^2, \quad (22)$$

$$\text{SSE}_{H2} = \sum_{k=1}^N (V(k) - \hat{V}_2(k))^2. \quad (23)$$

### 1.4.3 Результаты

Результаты аппроксимации представлены на рисунках 5–8:

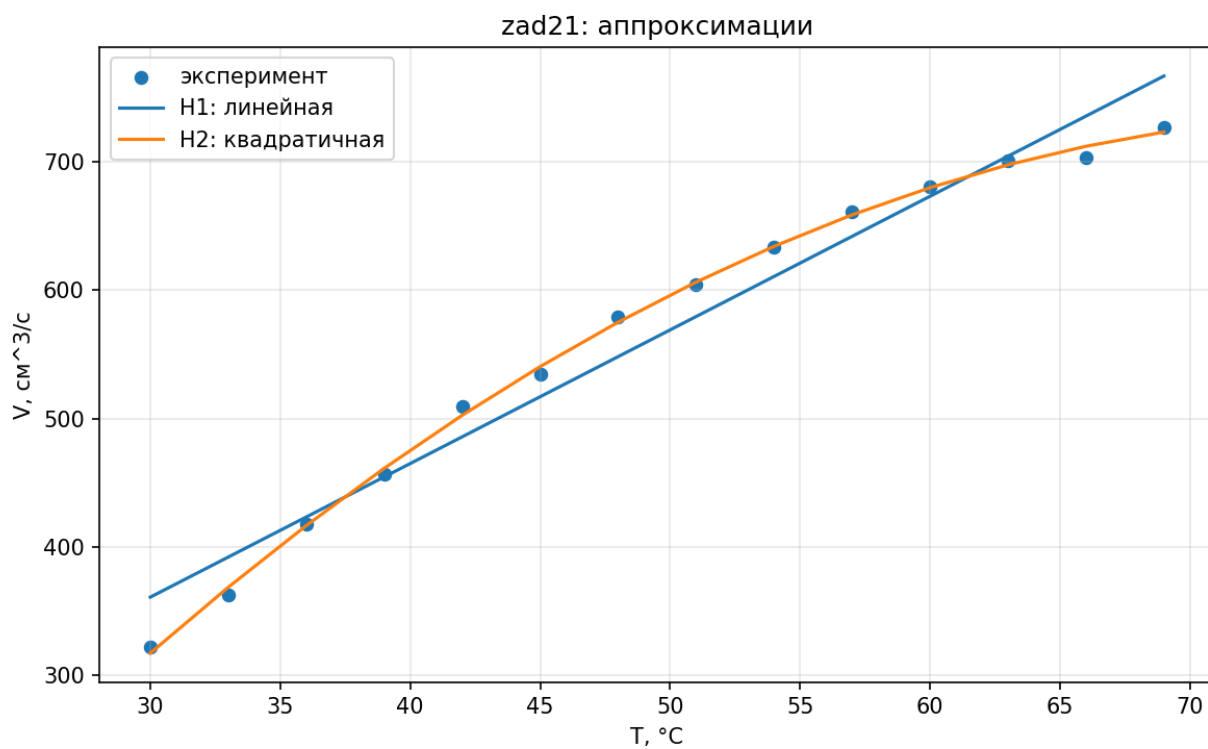


Рисунок 5 — Задание 2: данные zad21 и аппроксимации H1, H2

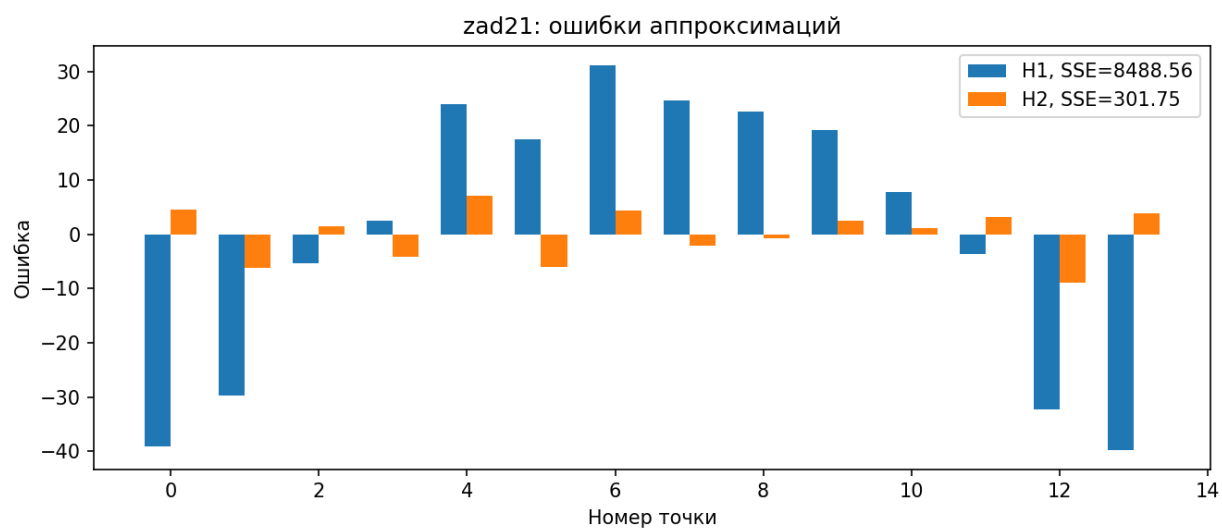


Рисунок 6 — Задание 2: ошибки аппроксимаций для zad21

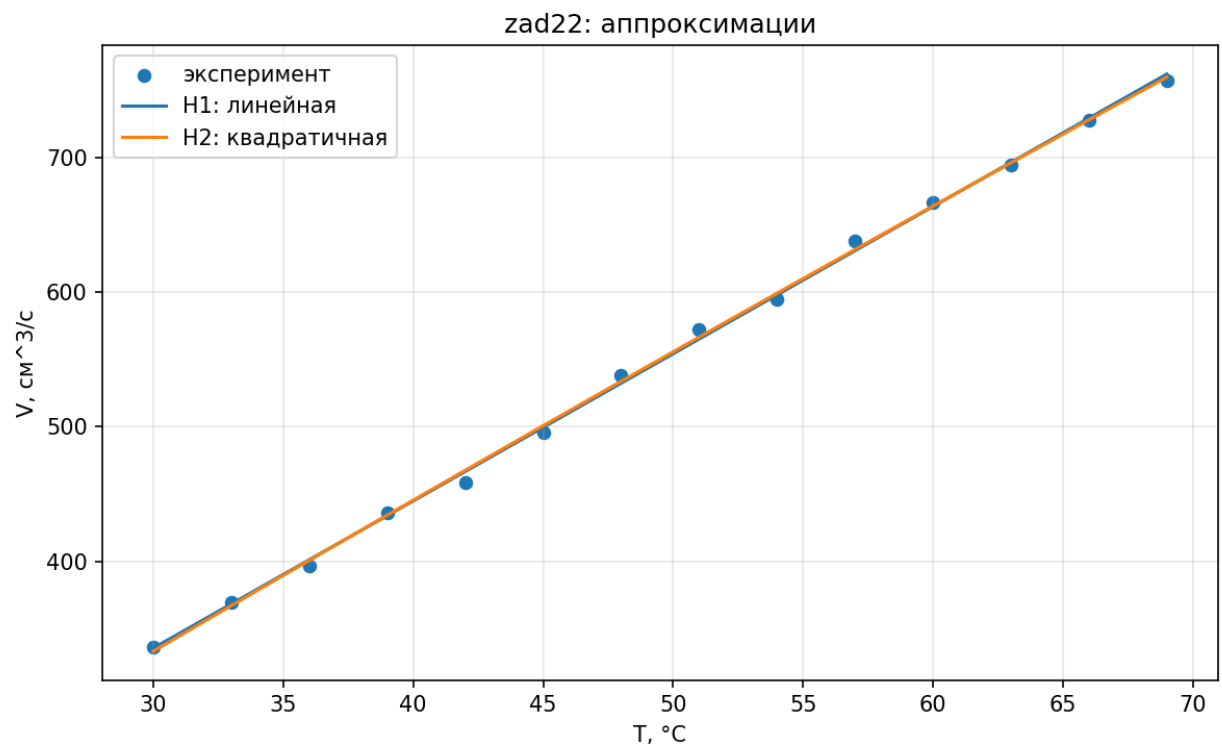


Рисунок 7 — Задание 2: данные zad22 и аппроксимации H1, H2



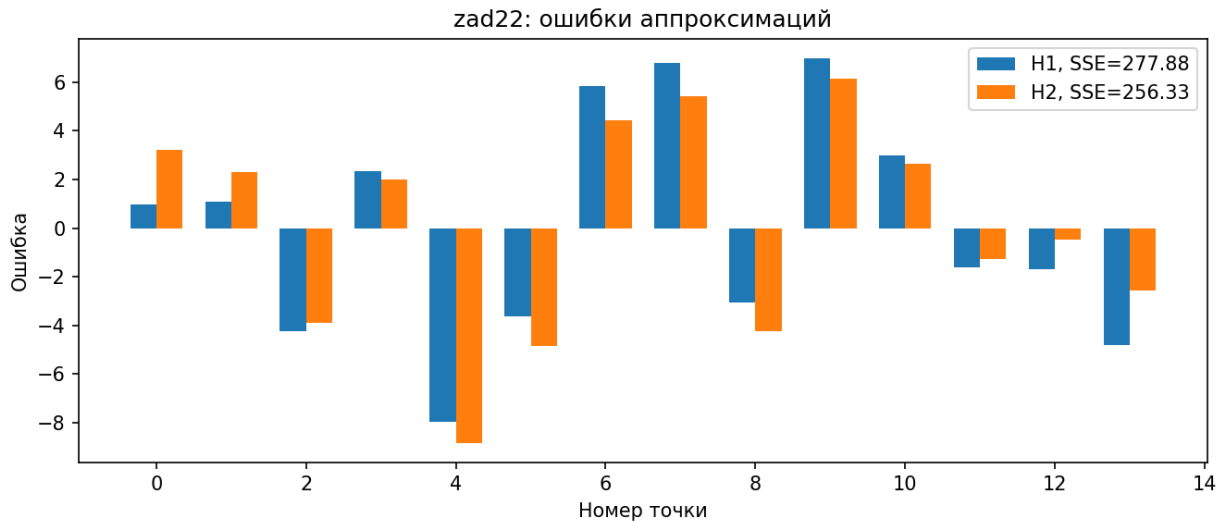


Рисунок 8 — Задание 2: ошибки аппроксимаций для zad22

#### 1.4.4 Выводы

Суммы квадратов ошибок и свойства остатков:

- **zad21**:  $SSE_{H1} = 8488.56$ ,  $SSE_{H2} = 301.75$ . Для H1 среднее остатка  $\bar{e}_{H1} \approx 1.2 \cdot 10^{-13}$ , дисперсия  $\text{Var}(e_{H1}) \approx 652.97$ , автокорреляция лаг-1  $\rho_{H1}(1) \approx 0.807$ , корреляция с температурой  $\text{corr}(e_{H1}, T) \approx 0$ . Для H2  $\bar{e}_{H2} \approx -7.9 \cdot 10^{-13}$ ,  $\text{Var}(e_{H2}) \approx 23.21$ ,  $\rho_{H2}(1) \approx -0.741$ ,  $\text{corr}(e_{H2}, T) \approx 0$ ,  $\text{corr}(e_{H2}, T^2) \approx 0$ . Значительно меньшая SSE и дисперсия остатка у H2, а также отсутствие заметной корреляции с регрессорами позволяют признать квадратичную модель (H2) наиболее достоверной.
- **zad22**:  $SSE_{H1} = 277.88$ ,  $SSE_{H2} = 256.33$ . Для H1:  $\bar{e}_{H1} \approx -3.2 \cdot 10^{-14}$ ,  $\text{Var}(e_{H1}) \approx 21.38$ ,  $\rho_{H1}(1) \approx 0.001$ ,  $\text{corr}(e_{H1}, T) \approx 0$ . Для H2:  $\bar{e}_{H2} \approx -8.4 \cdot 10^{-13}$ ,  $\text{Var}(e_{H2}) \approx 19.72$ ,  $\rho_{H2}(1) \approx -0.063$ ,  $\text{corr}(e_{H2}, T) \approx 0$ ,  $\text{corr}(e_{H2}, T^2) \approx 0$ . Разница SSE невелика, но модель H2 обеспечивает немного меньшую дисперсию остатка без появления корреляций, поэтому её можно считать предпочтительной; при этом линейная модель также допустима.

Таким образом, для набора zad21 достоверна квадратичная аппроксимация (H2), для zad22 оба описания сопоставимы, однако H2 даёт чуть более точную аппроксимацию.

## 1.5 Задание 3

### 1.5.1 Постановка

Для наборов zad31, zad32 даны нелинейные функции:

$$\text{zad31: } y(x) = (7.0^{p_1}) \cdot (x^{p_2}), \quad (24)$$

$$\text{zad32: } y(x) = p_1 \cdot \exp(p_2 x), \quad (25)$$

где требуется представить эти функции в виде линейной регрессии по параметрам  $p_1, p_2$ , оценить параметры и построить графики  $y$  и  $\hat{y}$ .

### 1.5.2 Методика

Для zad31 функция  $y(x) = (7.0^{p_1}) \cdot (x^{p_2})$  нелинейна по параметрам. Применяя логарифмирование:

$$\ln y = p_1 \ln 7.0 + p_2 \ln x. \quad (26)$$

Вводя обозначения  $\tilde{y} = \ln y$ ,  $\phi_1 = \ln 7.0$ ,  $\phi_2 = \ln x$ , получаем линейную модель:

$$\tilde{y} = p_1 \phi_1 + p_2 \phi_2. \quad (27)$$

Матрица регрессоров:

$$\Phi = \begin{bmatrix} \ln 7.0 & \ln x(1) \\ \ln 7.0 & \ln x(2) \\ \vdots & \vdots \\ \ln 7.0 & \ln x(N) \end{bmatrix}, \quad (28)$$

вектор параметров  $\theta = [p_1, p_2]^T$ , а оценка:

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T \tilde{Y}, \quad (29)$$

где  $\tilde{Y} = [\ln y(1), \ln y(2), \dots, \ln y(N)]^T$ . После оценивания восстановление исходной функции:

$$\hat{y}(x) = (7.0^{\hat{p}_1}) \cdot (x^{\hat{p}_2}). \quad (30)$$

Для zad32 функция  $y(x) = p_1 \exp(p_2 x)$  также нелинейна. Логарифмируя:

$$\ln y = \ln p_1 + p_2 x. \quad (31)$$

Вводя  $\tilde{y} = \ln y$ ,  $\phi_1 = 1$ ,  $\phi_2 = x$ , получаем:

$$\tilde{y} = \ln p_1 \cdot \phi_1 + p_2 \phi_2. \quad (32)$$

Матрица регрессоров:

$$\Phi = \begin{bmatrix} 1 & x(1) \\ 1 & x(2) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix}, \quad (33)$$

оценка параметров:

$$\begin{bmatrix} \ln \hat{p}_1 \\ \hat{p}_2 \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T \tilde{Y}, \quad (34)$$

откуда  $\hat{p}_1 = \exp(\ln \hat{p}_1)$ , а восстановленная функция:

$$\hat{y}(x) = \hat{p}_1 \exp(\hat{p}_2 x). \quad (35)$$

### 1.5.3 Результаты

Результаты оценивания представлены на рисунках 9 и 10:

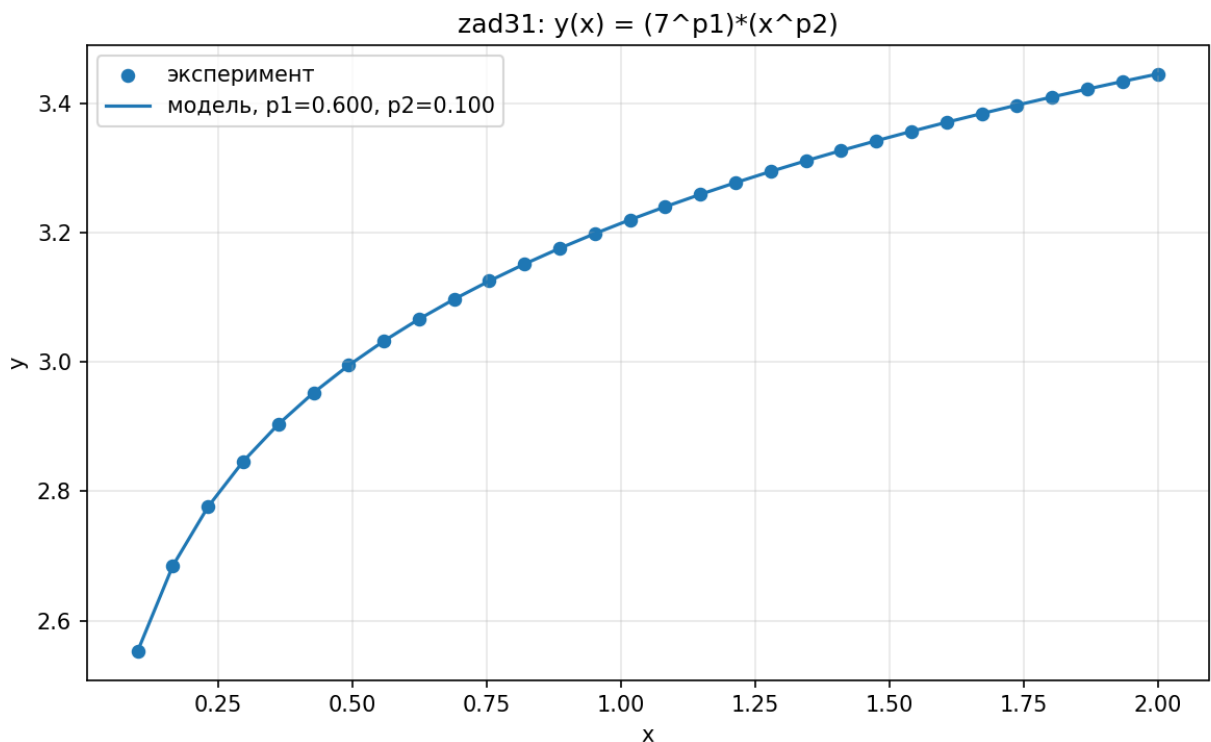


Рисунок 9 — Задание 3: данные и аппроксимация для zad31

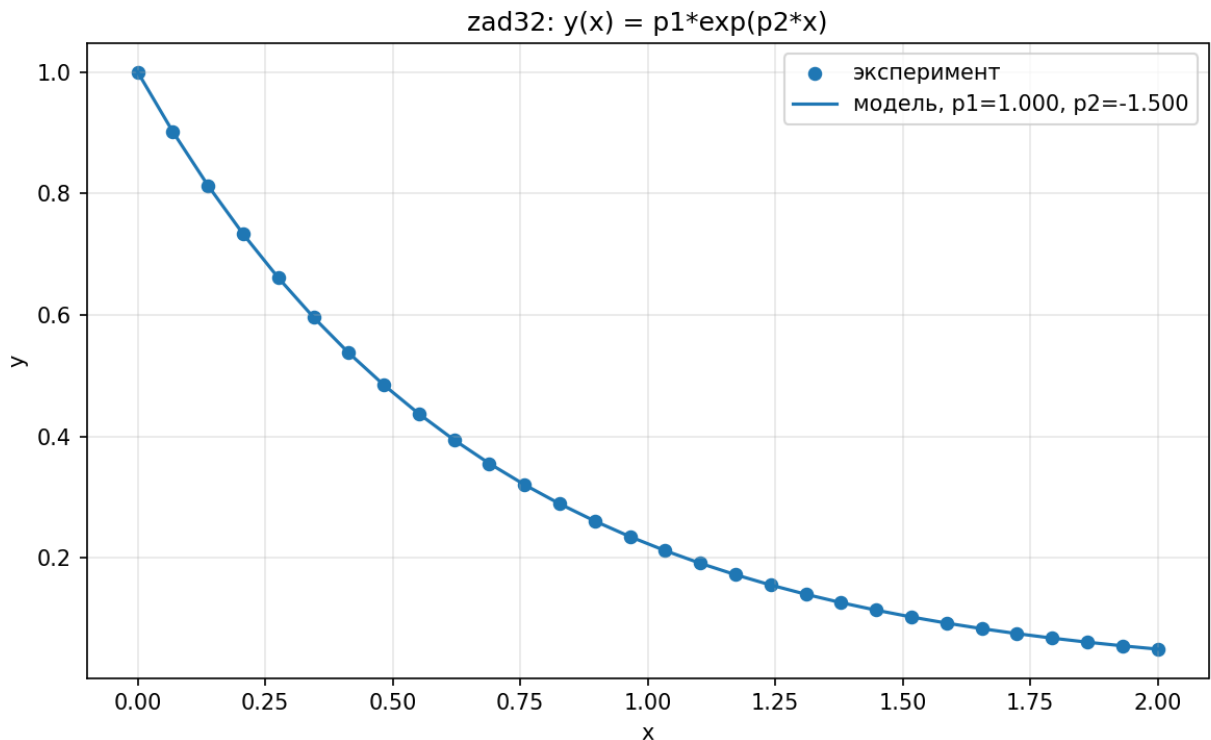


Рисунок 10 — Задание 3: данные и аппроксимация для zad32

### 1.5.4 Результаты

Оцененные параметры: для zad31  $\hat{p}_1 \approx 0.6$ ,  $\hat{p}_2 \approx 0.10$ ; для zad32  $\hat{p}_1 \approx 1.0$ ,  $\hat{p}_2 \approx -1.5$ . Как видно из рисунков 9 и 10, аппроксимации хорошо совпадают с экспериментальными данными, что подтверждает корректность преобразования нелинейных моделей к линейной регрессии.

### 1.6 Заключение

По Заданию 1 оценки параметров корректно восстанавливают истинные коэффициенты, ошибка несмещённая. В Задании 2 для zad21 предпочтительна квадратичная модель, для zad22 различие невелико. В Задании 3 параметры восстановлены: для zad31  $p_1 \approx 0.6$ ,  $p_2 \approx 0.10$ ; для zad32  $p_1 \approx 1.0$ ,  $p_2 \approx -1.5$ ; аппроксимации хорошо совпадают с экспериментальными данными.