

11 Обработка экспериментальных данных

11.1 Метод наименьших квадратов

Метод наименьших квадратов позволяет по экспериментальным данным подобрать такую аналитическую функцию, которая проходит настолько близко к экспериментальным точкам, насколько это возможно.

Пусть в результате эксперимента были получены некоторые данные, отображенные в виде таблицы (табл. 11.1). Требуется построить аналитическую зависимость, наиболее точно описывающую результаты эксперимента.

Таблица 11.1 Экспериментальные данные

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	...	x_n
y_i	y_1	y_2	y_3	y_4	y_5	y_6	y_7	...	y_n

Идея метода наименьших квадратов заключается в том, что функцию $Y = f(x, a_0, a_1, \dots, a_k)$ необходимо подобрать таким образом, чтобы сумма квадратов отклонений измеренных значений y_i от расчетных Y_i была наименьшей:

$$S = \sum_{i=1}^n (y_i - f(x_i, a_0, a_1, \dots, a_k))^2 \rightarrow \min \quad (11.1)$$

Задача сводится к определению коэффициентов a_i из условия (11.1). Для реализации этой задачи в Scilab предусмотрена функция

$$[a, S] = \text{datafit}(F, z, c)$$

где F - функция параметры, которой необходимо подобрать; z - матрица исходных данных; c - вектор начальных приближений; a - вектор коэффициентов; S - сумма квадратов отклонений измеренных значений от расчетных.

Рассмотрим использование функции `datafit` на примере.

ЗАДАЧА 11.1.

В результате опыта холостого хода определена зависимость потребляемой из сети мощности (P , Вт) от входного напряжения (U , В) для асинхронного двигателя (табл. 11.2).

Таблица 11.2. Зависимость потребляемой из сети мощности от входного напряжения.

U, В	132	140	150	162	170	180	190	200	211	220	232	240	251
P, Вт	330	350	385	425	450	485	540	600	660	730	920	1020	1350

Методом наименьших квадратов подобрать зависимость вида $P = a_1 + a_2 U + a_3 U^2 + a_4 U^3$.

Далее приведено решение задачи с комментариями.

//Функция, вычисляющая разность между экспериментальными

//и теоретическими значениями,

//перед использованием необходимо определить

//z=[x;y] - матрицу исходных данных и

//c - вектор начальных значений коэффициентов,

//размерность вектора должна совпадать

//с количеством искомым коэффициентов

```

function [zr]=G(c,z)
zr=z(2)-c(1)-c(2)*z(1)-c(3)*z(1)^2-c(4)*z(1)^3
endfunction
//Исходные данные
x=[1.32 1.40 1.50 1.62 1.70 1.80 1.90...
   2.00,2.11,2.20,2.32,2.40,2.51];
y=[3.30 3.50 3.85 4.25 4.50 4.85 5.40...
   6.00 6.60 7.30 9.20 10.20 13.50];
//Формирование матрицы исходных данных
z=[x;y];
//Вектор начальных приближений
c=[0;0;0;0];
//Решение задачи
[a,err]=datafit(G,z,c)
S =
    0.5287901
a =
    - 51.576664
     95.594671
    - 55.695312
     11.111453

```

Листинг 11.1

Итак, в результате работы функции `datafit` была подобрана аналитическая зависимость вида $P = -51.577 + 95.595U - 55.695U^2 + 11.111U^3$, а сумма квадратов отклонений измеренных значений от расчетных составила 0.529.

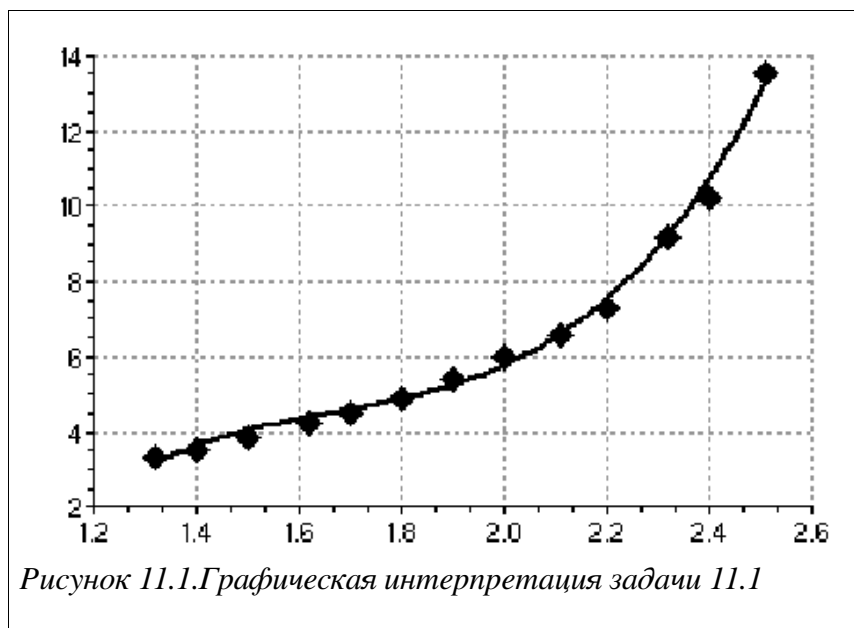
Геометрическая интерпретация задачи (рис. 11.1):

```

//Построение графика экспериментальных данных
plot2d(x,y,-4);
//Построение графика подобранной функции
t=1.32:0.01:2.51;
Ptc=a(1)+a(2)*t+a(3)*t^2+a(4)*t^3;
plot2d(t,Ptc);

```

Листинг 11.2



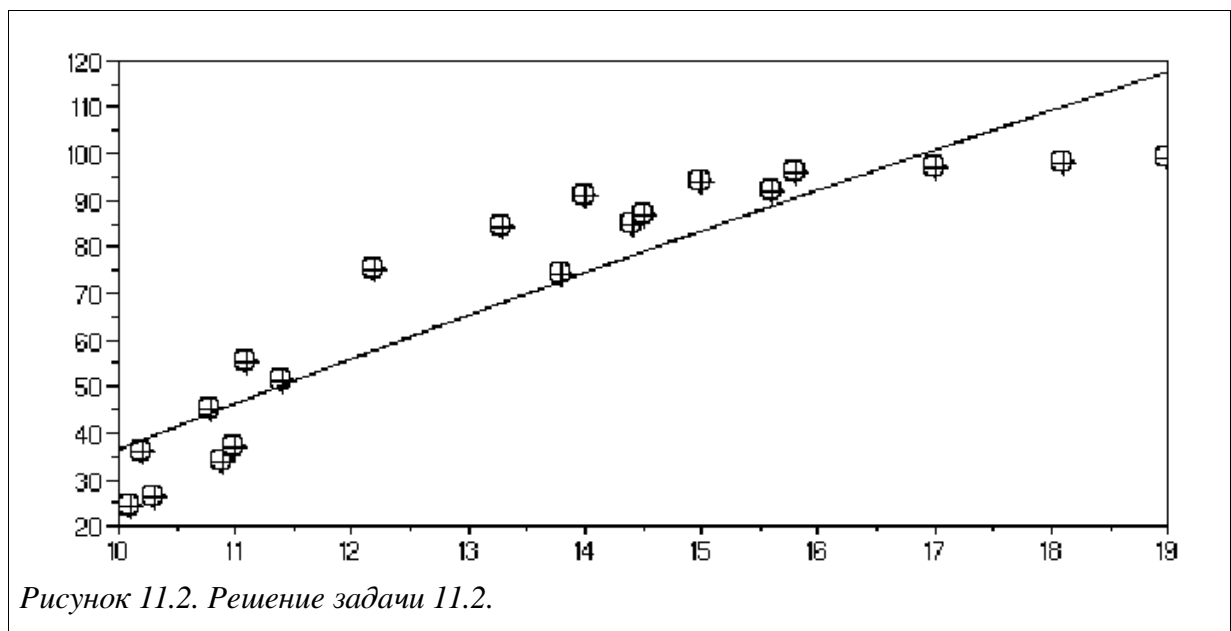
ЗАДАЧА 11.2.

К некоторым экспериментальным данным подобрать зависимость вида $Y = a_1 \cdot x^{a_2} + a_3$.

Далее представлено решение задачи и ее графическая интерпретация (рис.11.2).

```
function [zr]=F(c,z)
zr=z(2)-c(1)*z(1)^c(2)-c(3);
endfunction
x=[10.1,10.2,10.3,10.8,10.9,11,11.1,11.4,12.2,13.3,13.8,...
    14,14.4,14.5,15,15.6,15.8,17,18.1,19];
y=[24,36,26,45,34,37,55,51,75,84,74,91,85,87,...
    94,92,96,97,98,99];
z=[x;y];
c=[0;0;0];
[a,S]=datafit(F,z,c);
t=10:0.01:19;
Yt=a(1)*t^a(2)+c(3);
plot2d(x,y,-3);
plot2d(t,Yt);
```

Листинг 11.3



Одной из наиболее часто используемых в методе наименьших квадратов функций является прямая, описываемая уравнением вида $y=a_1+a_2x$, которая называется *линией регрессии* y на x . Параметры a_1 и a_2 являются *коэффициентами регрессии*. Показатель, характеризующий тесноту линейной связи между x и y и называемый *коэффициентом корреляции* и рассчитывается по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - Mx) \cdot (y_i - My)}{\sqrt{\sum_{i=1}^n (x_i - Mx)^2 \sum_{i=1}^n (y_i - My)^2}}, \quad Mx = \frac{\sum_{i=1}^n x_i}{n}, \quad My = \frac{\sum_{i=1}^n y_i}{n} \quad (11.2)$$

Значение коэффициента корреляции удовлетворяет соотношению $-1 \leq r \leq 1$. Чем меньше отличается абсолютная величина r от единицы, тем ближе к линии регрессии располагаются экспериментальные точки. Если коэффициент корреляции близок к нулю, то это означает, что между x и y отсутствует линейная связь, но может существовать другая, нелинейная, зависимость.

Аналогом коэффициента корреляции r для нелинейных зависимостей является *индекс корреляции*, рассчитываемый по формуле:

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - Y_i)^2}{\sum_{i=1}^n (y_i - My)^2}} \quad (11.3)$$

где y - экспериментальные значения, Y - значения, найденные методом наименьших квадратов, My - среднее значение y . Индекс корреляции по своему абсолютному значению колеблется в пределах от 0 до 1. При функциональной зависимости индекс корреляции равен 1. В случае отсутствия связи $R=0$. Если коэффициент корреляции r является мерой тесноты связи только для линейной формы, то индекс корреляции R - и для линейной, и для криволинейной. При прямолинейной связи коэффициент корреляции по своей абсолютной величине равен индексу корреляции.

Для расчета *коэффициентов регрессии* в Scilab предназначена функция

$$a = \text{regress}(x, y)$$

где x и y - экспериментальные данные, a - вектор коэффициентов линии регрессии a_1 и a_2 .
Рассмотрим работу этой функции на примере.

ЗАДАЧА 11.3.

В «Основах химии» Д.И. Менделеева приводятся данные о растворимости азотнокислого натрия NaNO_3 в зависимости от температуры воды. Число условных частей NaNO_3 растворяющихся в 100 частях воды при соответствующих температурах представлено в табл. 11.3. Требуется определить растворимость азотнокислого натрия при температуре 32 градуса в случае линейной зависимости и найти коэффициент и индекс корреляции.

Таблица 11.3. Данные о растворимости NaNO_3 в зависимости от температуры воды.

0	4	10	15	21	29	36	51	68
66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

Решение задачи:

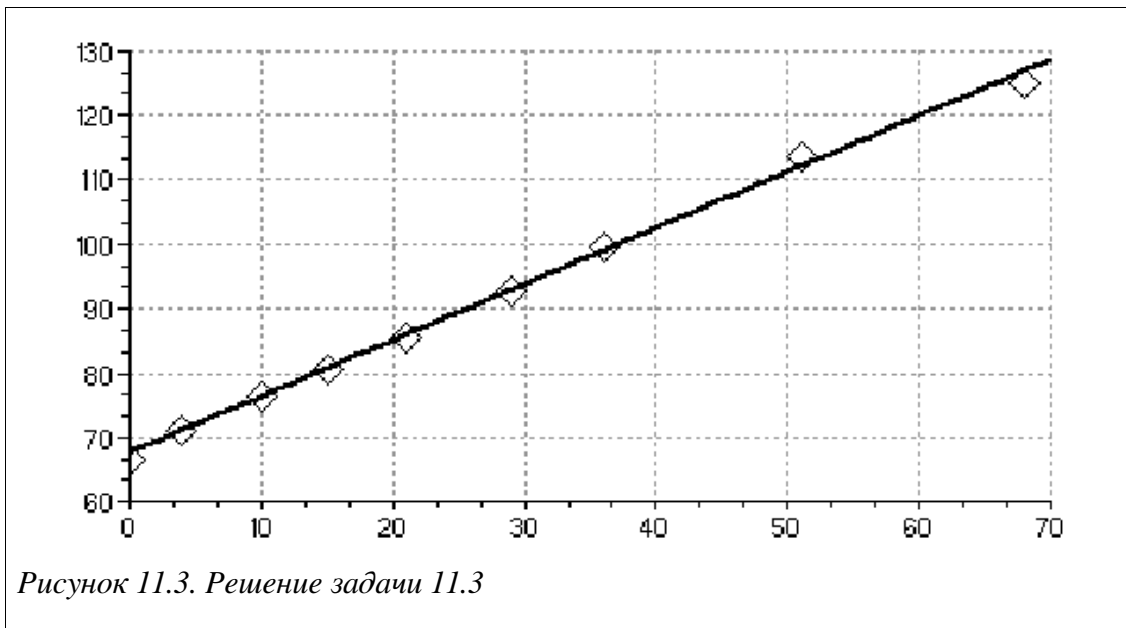
```
//Экспериментальные данные
x=[0 4 10 15 21 29 36 51 68];
y=[66.7 71 76.3 80.6 85.7 92.9 99.4 113.6 125.1];
//Расчет коэффициентов регрессии
a=regress(x,y)
a =
    67.507794
    0.8706404
//Растворимость азотного натрия при температуре 32 градуса
-->t=32;a(1)+a(2)*t
ans = 95.368287
//Коэффициент корреляции (11.2)
r=sum((x-mean(x)).*(y-mean(y)))/...
sqrt(sum((x-mean(x))^2)*sum((y-mean(y))^2))
r = 0.9989549
//Индекс корреляции (11.3)
R=sqrt(1-sum((y-(a(1)+a(2)*x))^2)/sum((y-mean(y))^2))
R = 0.9989549
```

Листинг 11.4

Построение графика экспериментальных данных и линии регрессии (рис. 11.3):

```
t=0:70; Yt=a(1)+a(2)*t;
plot2d(x,y,-5); plot2d(t,Yt);
```

Листинг 11.5



11.2 Интерполяция функций

Простейшая задача интерполирования заключается в следующем. На отрезке $[a; b]$ заданы точки $x_0, x_1, x_2, \dots, x_n$ (всего $n+1$ точка), которые называют узлами интерполяции, и значения некоторой функции $f(x)$ в этих точках:

$$f(x_0)=y_0, f(x_1)=y_1, f(x_2)=y_2, \dots, f(x_n)=y_n. \quad (11.4)$$

Требуется построить интерполирующую функцию $F(x)$, принадлежащую известному классу и принимающую в узлах интерполяции те же значения, что и $f(x)$:

$$F(x_0)=y_0, F(x_1)=y_1, F(x_2)=y_2, \dots, F(x_n)=y_n. \quad (11.5)$$

Для решения подобной задачи довольно часто используют сплайн-интерполяцию (от английского слова *spline* рейка, линейка). Один из наиболее распространенных вариантов интерполяции — интерполяция кубическими сплайнами. Кроме того существуют квадратичные и линейные сплайны.

В Scilab для построения *линейной интерполяции* служит функция

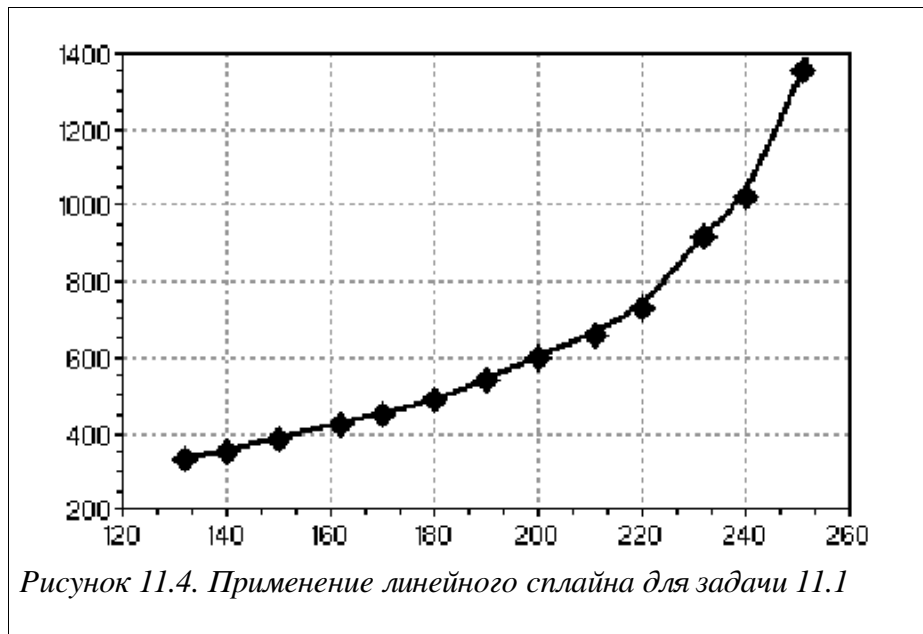
$$y=\text{interpln}(z, x)$$

где z — матрица исходных данных; x — вектор абсцисс; y — вектор значений линейного сплайна в точка x .

Далее приведен пример использования функции `interpln`. Здесь линейный сплайн применяется для решения задачи 11.1. Графическое решение задачи показано на рис. 11.4.

```
x=[132 140 150 162 170 180 190 200 211 220 232 240 251];
y=[330 350 385 425 450 485 540 600 660 730 920 1020 1350];
plot2d(x, y, -4);
z=[x; y];
t=132:5:252; ptd=interpln(z, t);
plot2d(t, ptd);
```

Листинг 11.6



Построение *кубического сплайна* в Scilab состоит из двух этапов: вначале вычисляются коэффициенты сплайна с помощью функции $d = \text{splin}(x, y)$, а затем рассчитываются значения интерполяционного полинома в точке $y = \text{interp}(t, x, y, d)$.

Функция $d = \text{splin}(x, y)$ имеет следующие параметры: x — строго возрастающий вектор, состоящий минимум из двух компонент; y — вектор того же формата, что и x ; d — результат работы функции, коэффициенты кубического сплайна.

Для функции $y = \text{interp}(t, x, y, k)$ параметры x , y и d имеют те же значения, параметр t — это вектор абсцисс, а y — вектор ординат, являющихся значениями кубического сплайна в точках x .

ЗАДАЧА 11.4.

Найти приближенное значение функции при заданном значении аргумента с помощью интерполяции кубическими сплайнами в точках $x_1 = 0,702$, $x_2 = 0,512$, $x_3 = 0,608$. Функция задана таблично (табл. 11.4).

Таблица 11.4. Исходные данные к задаче 11.5

0.43	0.48	0.55	0.62	0.7	0.75
1.63597	1.73234	1.87686	2.03345	2.22846	2.35973

Задачу можно решить так:

```
x=[0.43 0.48 0.55 0.62 0.7 0.75];
y=[1.63597 1.73234 1.87686 2.03345 2.22846 2.35973];
plot2d(x,y,-4); //График экспериментальных данных
coeff=splin(x,y);
X=[0.702 0.512 0.608];
//Значение функции в заданных точках
Y=interp(X,x,y,coeff)
Y = 2.2335678    1.7969698    2.0057073
plot2d(X,Y,-3); //Нанесение точек на график
//Построение кубического сплайна
t=0.43:0.01:0.75;
```

```
ptd=interp(t,x,y,koeff);  
plot2d(t,ptd);
```

Листинг 11.7

Графическая часть задачи представлена на рис.11.5.

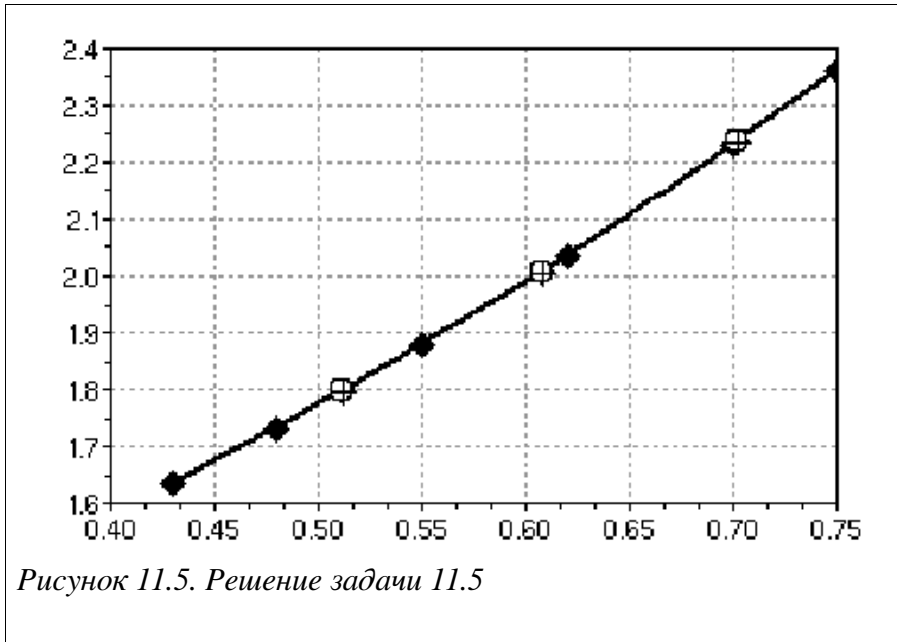


Рисунок 11.5. Решение задачи 11.5