

文章编号:1004-4736(2008)01-0117-03

# 实验数据的线性拟合及计算机处理

何菊明,王 芙

(武汉工程大学理学院,湖北 武汉 430074)

摘 要:介绍了实验数据处理的线性拟合法,给出了一元线性拟合的算法和C语言程序,对其它复杂情况下的处理方法也进行了简单的分析。

关键词:数据处理;线性拟合;计算机应用

中图分类号:O 4.39 文献标识码:A

根据物理实验的观测数据,寻求两个物理量之间近似的解析函数关系式或曲线方程,就是数据拟合或曲线拟合。数值拟合是物理实验数据处理的重要方法之一,本文简述了线性拟合的原理,并结合实例给出了一元线性拟合的C语言程序,对其它复杂情况下如何运用数据拟合法也进行了讨论。

## 1 一元线性拟合

### 1.1 最小二乘法

最小二乘法是线性拟合的基本原理之一<sup>[1,2]</sup>。假设两物理量Y与x的关系为 $Y = a_0 + a_1x$ ,实验的目的就是要由测量数据找到此规律,即一元线性拟合方程。其中 $a_0, a_1$ 为待定系数。

对应于实验测量值 $x_i$ ,通过此关系式可得

$$Y_i = a_0 + a_1x_i \quad (1)$$

$Y_i$ 与实际测量的 $y_i$ 一般不相等,二者的差值称为偏差,将偏差与其平方和分别记为

$$\delta_i = y_i - Y_i = y_i - a_0 - a_1x_i$$

$$\varphi = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1x_i)^2 \quad (2)$$

确定 $a_0$ 和 $a_1$ 的最佳值通常采用最小二乘法,它要求偏差的平方和最小。

由 $\frac{\partial \varphi}{\partial a_0} = 0, \frac{\partial \varphi}{\partial a_1} = 0$ 可解出

$$a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}, a_0 = \bar{y} - a_1\bar{x} \quad (3)$$

将算出的 $a_0$ 和 $a_1$ 值代入式(1)就得到拟合方程。

### 1.2 计算步骤

归纳以上过程,可得出线性拟合的具体计算

步骤如下:

- 1) 计算各种平均值: $\bar{x} = \frac{1}{n} \sum x_i, \bar{y} = \frac{1}{n} \sum y_i, \overline{xy} = \frac{1}{n} \sum x_i y_i, \overline{x^2} = \frac{1}{n} \sum x_i^2$ . 其中的 $n$ 是实验次数, $i = 1, 2, \dots, n$ ;
- 2) 按式(3)计算 $a_1$ 和 $a_0$ 。

## 2 实例研究

### 2.1 铜丝电阻随温度变化的线性拟合

以实验中铜丝样品的电阻随温度变化为例(见表1)。

表1 铜丝电阻随温度变化的数据表

Table 1 R-t data sheet of Cu

| t/°C | R/Ω  | t/°C | R/Ω  |
|------|------|------|------|
| 0    | 4.38 | 110  | 6.26 |
| 10   | 4.56 | 120  | 6.44 |
| 20   | 4.70 | 130  | 6.58 |
| 30   | 4.86 | 140  | 6.74 |
| 40   | 5.08 | 150  | 6.94 |
| 50   | 5.24 | 160  | 7.12 |
| 60   | 5.40 | 170  | 7.28 |
| 70   | 5.58 | 180  | 7.42 |
| 80   | 5.74 | 190  | 7.60 |
| 90   | 5.96 | 200  | 7.78 |
| 100  | 6.06 |      |      |

根据表1数据<sup>[1]</sup>可初步判断R-t是一元线性方程

$$R = a_0 + a_1t$$

根据前述计算步骤设计一个C语言程序(图1),画出R-t数据点并绘出拟合曲线。

计算出的 $a_1$ 与 $a_0$ 值可写出一元线性拟合方程的具体形式,也可以求出铜的电阻温度系数<sup>[3]</sup>。下面是绘制拟合曲线的C语言程序:

收稿日期:2007-09-10

作者简介:何菊明(1961-),男,湖北蕲春人,副教授,硕士,研究方向:从事大学物理教学与研究以及物理学中的数值计算与计算机模拟。

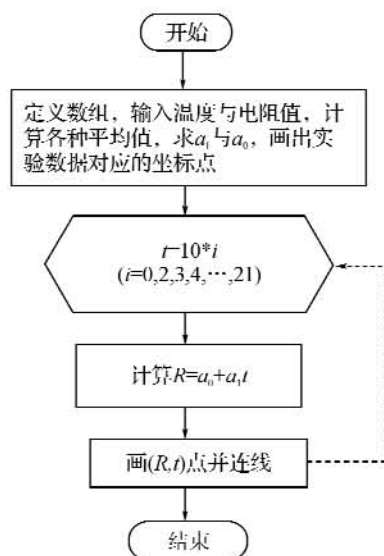


图 1 程序设计流程图

Fig. 1 The flow of programming

```

#include <graphics.h>
#include <stdlib.h>
#include <stdio.h>
main()
{
    int graphdrv=DETECT;
    int graphmode;
    int i, polypoints[] = {25, 20, 25, 460, 625, 460};
    float t, t0, R, R0, a, a0, a1, x[21], y[21];
    float mx=0, my=0, mxy=0, mx2=0;
    clrscr();
    printf("Enter y[i]:\n");
    for(i=0; i<21; i++)
    {
        x[i] = 10 * i;
        printf("x[ %d] = %d", i, (int)x[i]);
        printf(" y[ %d] = ?", i);
        scanf("%f", &y[i]);
    }
    for(i=0; i<21; i++)
    {
        mx = mx + x[i]/21;
        mx2 = mx2 + x[i] * x[i]/21;
        my = my + y[i]/21;
        mxy = mxy + x[i] * y[i]/21;
    }
    a1 = (mxy - mx * my) / (mx2 - mx * mx);
    a0 = my - a1 * mx;
    initgraph(&graphdrv, &graphmode, "");
    drawpoly(3, polypoints);
    for(i=1; i<21; i++)
  
```

```

{
    t = 25 + 30 * i;
    line(t, 460, t, 455);
}
for(i=1; i<5; i++)
{
    R = 460 - 100 * i;
    line(25, R, 30, R);
}
for(i=0; i<21; i++)
{
    line(3 * x[i] + 20, 860 - 100 * y[i], 3 * x[i] + 30,
    860 - 100 * y[i]);
    line(3 * x[i] + 25, 865 - 100 * y[i], 3 * x[i] + 25,
    855 - 100 * y[i]);
}
settextstyle(1, 0, 0);
outtextxy(180, 20, "Cu's R-t Curve");
settextstyle(2, 0, 0);
outtextxy(0, 450, "4.00");
outtextxy(0, 350, "5.00");
outtextxy(0, 250, "6.00");
outtextxy(0, 150, "7.00");
outtextxy(0, 50, "8.00");
outtextxy(25, 465, "0");
outtextxy(170, 465, "50");
outtextxy(320, 465, "100");
outtextxy(470, 465, "150");
outtextxy(620, 465, "200");
getch();
moveto(25, 860 - 100 * a0);
for(i=0; i<21; i++)
{
    t = 10 * i;
    R = a0 + a1 * t;
    lineto(3 * t + 25, 860 - 100 * R);
}
while(! kbhit());
closegraph();
}
  
```

绘出的  $R-t$  拟合曲线如图 2。

## 2.2 钢包体积与使用次数的线性拟合

炼钢厂出钢时所用的盛钢水的钢包, 在使用过程中由于钢液及炉渣对包衬耐火材料的侵蚀, 使其体积不断增大。表 2 是钢包容积与相应的使用次数的实验数据<sup>[3]</sup>。

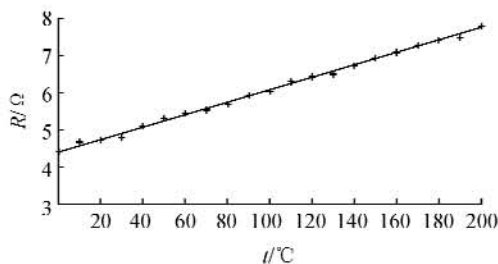


图2 R-t拟合曲线

Fig. 2 R-t Fitting curve

表2 钢包容积与使用次数的数据表

Table 2 V-N data sheet

| 次数 N | 容积 V/L | 次数 N | 容积 V/L |
|------|--------|------|--------|
| 2    | 106.42 | 11   | 110.59 |
| 3    | 108.26 | 14   | 110.60 |
| 4    | 109.58 | 15   | 110.90 |
| 5    | 109.50 | 16   | 110.76 |
| 7    | 110.00 | 18   | 111.00 |
| 8    | 109.93 | 19   | 111.20 |
| 10   | 110.49 |      |        |

对于这些数据,先在草图上描出其数据点,发现与双曲线近似,因而将次数  $N$  与容积  $V$  的关系写作

$$\frac{1}{V} = a_0 + \frac{a_1}{N}$$

令  $y = \frac{1}{V}$ ,  $x = \frac{1}{N}$ , 将以上关系改写成一元线性拟合方程式

$$y = a_0 + a_1 x$$

按上一例方法,可以绘出钢包容积  $V$  与使用次数  $N$  的拟合曲线(图3)。

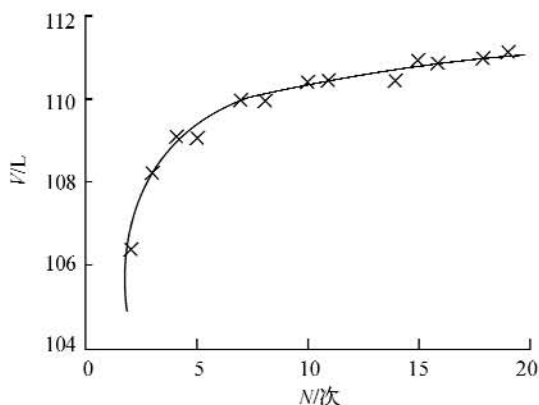


图3 V-N拟合曲线

Fig. 3 V-N Fitting curve

### 3 其它情况的处理方法

#### 3.1 特殊的一元非线性拟合

如果物理量之间并不存在线性关系,但经过适当变换后可转化为线性关系。例如半导体的电阻与温度的关系为

$$R = Ce^{\frac{E_g}{2kT}}$$

式中  $k$  是玻耳兹曼常量,  $E_g$  是禁带宽度,  $C$  在温度范围不大时可视为常量。显然,电阻  $R$  和热力学温度  $T$  之间并非线性关系。但对上式取对数后就可以得到一个线性关系

$$\ln R = \ln C + \frac{E_g}{2k} \frac{1}{T}$$

令  $Y = \ln R$ ,  $x = 1/T$ , 则转化为式(1), 其中,

$$a_0 = \ln C, a_1 = \frac{E_g}{2k}.$$

$Y$  与  $x$  的数据处理方法与一元线性拟合相同,但此时的  $R-T$  拟合曲线不是直线了<sup>[3]</sup>。

#### 3.2 任意的一元非线性拟合

如果物理规律是一元非线性函数关系,可用如下的多项式进行拟合

$$Y_i = a_0 + a_1 x + a_2 x^2 + \dots + a_k x^k$$

对这一情况,除了增加了一个求解线性代数方程组的过程以外,其它计算步骤也与一元线性拟合相同。

#### 3.3 多元线性拟合

如果物理规律具有如下形式

$$Y = a_0 + a_1 x_1 + \dots + a_k x_k$$

仍然可用  $n$  组数据 ( $n > k$ ) 进行多元线性拟合。

$$Y_i = a_0 + a_1 x_{i1} + \dots + a_k x_{ik}$$

将实验测得的  $n$  组数据代入其中,用计算机解  $k$  元线性方程组,由最小二乘法确定出  $k+1$  个系数,即得  $k$  元线性拟合方程。

## 4 结 语

本文结合实例,介绍了一元线性拟合的原理与计算步骤,并提供了计算机 C 语言绘制拟合曲线的程序,解决了求拟合方程和绘制拟合函数曲线的问题。在文章的第三部分,提出了如何将非一元或非线性问题转化为一元线性问题的方案,这对实验数值的计算机处理,具有一定的参考价值与指导意义。

参考文献:

- [1] 陈鍾贤. 计算物理学[M]. 哈尔滨: 哈尔滨工业大学出版社, 2001. 37-45.
- [2] 钟尔杰, 谢云荪. 实用数值计算方法[M]. 北京: 高等教育出版社, 2001. 51-55.
- [3] 王惠德. 物理实验[M]. 天津: 天津大学出版社, 1997. 234-235.

(下转第124页)

$$\frac{\sigma_1}{\sqrt{1 + \sum_{i=1}^2 a_{i1}^2}} = \frac{9}{\sqrt{1+3^2}} = \sqrt{\frac{81}{10}},$$

$$\frac{\sigma_2}{\sqrt{1 + \sum_{i=1}^2 a_{i2}^2}} = \frac{8}{\sqrt{1+2^2}} = \sqrt{\frac{64}{5}},$$

$$\frac{\sigma_3}{\sqrt{1 + \sum_{i=1}^2 a_{i3}^2}} = \frac{40}{\sqrt{1+10^2+2^2}} = \sqrt{\frac{1600}{105}},$$

$$\frac{\sigma_4}{\sqrt{1 + \sum_{i=1}^2 a_{i4}^2}} = \frac{19}{\sqrt{1+4^2+\left(\frac{1}{2}\right)^2}} = \sqrt{\frac{1444}{69}}.$$

因  $\frac{\sigma_4}{\sqrt{1 + \sum_{i=1}^2 a_{i4}^2}} = \sqrt{\frac{1444}{69}}$  最大, 利用定理 2,

选取第 4 列为主列, 再利用  $\theta$  法则确定  $a_{14} = 4$  为主元素. 经 1 次迭代后即得表 6.

按方法一求解共需进行 1 次迭代, 而按方法二、三求解仅需进行 1 次迭代即可求得最优解.

#### 4 结 语

从上面这个简单的例子可以看出, 按使目标函数值增加得最多的原则来确定主元素或按使目标函数值增加得最快的原则来确定主元素, 较应用“最大  $\sigma$  法则”来确定主元素, 具有迭代次数更少、收敛速度更快的特点.

参考文献:

- [1] 刁在筠, 郑汉鼎, 刘家壮, 等. 运筹学(第二版)[M]. 北京: 高等教育出版社, 2003. 28-29.
- [2] 胡运权, 郭耀煌. 运筹学教程[M]. 北京: 清华大学出版社, 2002. 23.
- [3] 唐建国. 线性规划的目标函数最速递减算法[J]. 运筹与管理, 2005, 14(4): 55-59.

## Two rules of determining the principal element in simplex method

LUO Jin, ZHANG Zhi-jun, LIU Ren-he

(School of Science, Wuhan Institute of Technology, Wuhan 430074, China)

**Abstract:** This paper gives two rules of determining the principal element in simplex method, i. e., by maximal value or rapidest way of the increasing of the objective function. An example is given to confirm that it needs less iteration degree, and gets more rate of convergence in determining the principal element comparing with the maximal  $\sigma$  rule.

**Key words:** simplex method; value of the objective function; principal element

本文编辑: 萧 宁



(上接第 119 页)

## Linear fitting and computer processing on experiment data

HE Ju-ming, WANG Fu

(School of Sciences, Wuhan Institute of Technology, Wuhan 430074, China)

**Abstract:** In this paper, a linear fitting method of data processing is introduced. Algorithm and C language computer programs of simple linear fitting are given. Then, fitting method in complicated circumstance is also simply analysed.

**Key words:** data processing; linear fitting; computer application

本文编辑: 萧 宁