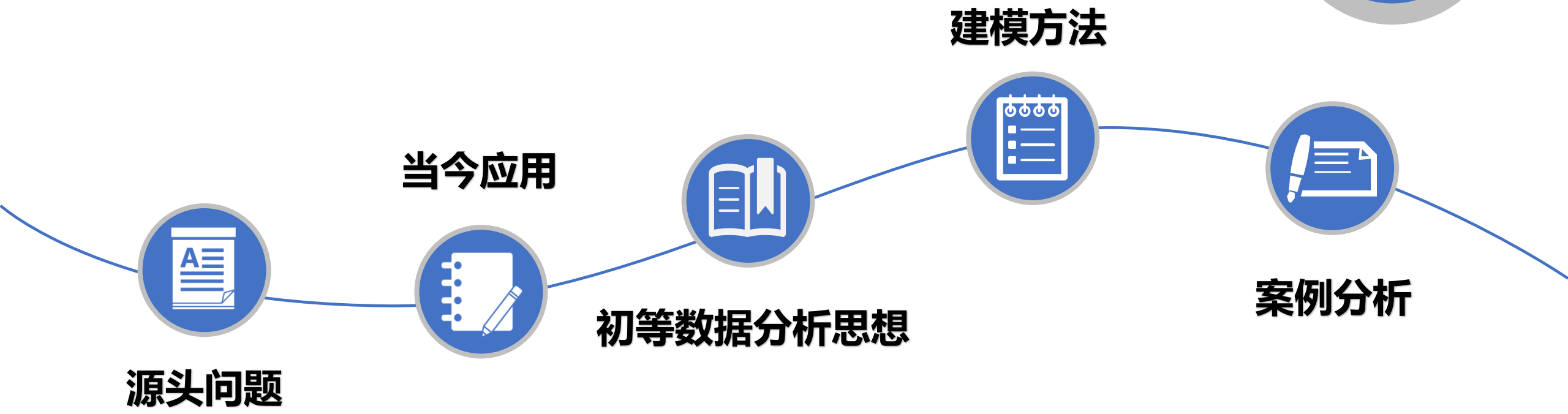
初等数据处理方法

谭忠

目录
CONTENTS



Part 1

源头问题与当今应用

函数概念的萌芽：古代对图形轨迹的研究；一个或几个量的变化会引起另一个量的变化，反映量与量之间的相互依赖关系。

（1）函数的力学来源：16-17世纪，欧洲国家争霸，船只定位、炸弹精准落点、各种运动。

（2）第一次涉及变量，引入函数思想-1637 年笛卡尔的《几何学》中，

（3）函数解析定义的开始-英国数学家格雷果里在1667 年给出

（4）公认最早提出函数概念-17 世纪德国数学家莱布尼茨。

如何采集数据、采集什么数据、怎样分析数据等问题

建立变量之间关系的三种基本方法：

观察法：利用变量之间的比例关系

拟合方法、插值方法

统称为初等数据分析方法

在现实问题或竞赛命题中，有的提供数据.

例 2.1 2004 年全国大学生数学建模竞赛题**“奥运场馆周围临时商店的建设问题”**，数据给了很多，这些数据是规划局在前三次亚运会期间采集的数据．采集人大概是“志愿者”之类的，完全不懂应该采集什么数据才能对解决这类问题有用，而将所有数据记录下来.

有的根本不给数据.

例 2.2 2010 年全国大学生数学建模竞赛题**“试就某个或某几个方面评估上海世博会的影响力”**．首先你必须确定从哪个方面评估，什么数据最能体现这个问题的本质，然后去查找相关数据。

有的问题问你需要什么数据.

例 2.3 2011 年全国大学生数学建模竞赛 A 题**《重金属的污染问题》**，最后一问“需要什么数据，我们可以确定城市土质变化的趋势?

有的问题需要数据，但问题中不仅没有给出数据，就连采集什么数据也没有说明，需要你自己判明应该采集什么数据才能说明这件事情.

例 2.4 2015 年国赛 B 题“**互联网 +”时代的出租车资源配置**有多家公司依托移动互联网建立了打车软件服务平台，实现了乘客与出租车司机之间的信息互通，同时推出了多种出租车的补贴方案．问题是搜集相关数据，建立数学模型研究如下问题:

(1) 试建立合理的指标，并分析不同时空出租车资源的“供求匹配”程度

(2) 分析各公司的出租车补贴方案是否对“缓解打车难”有帮助?

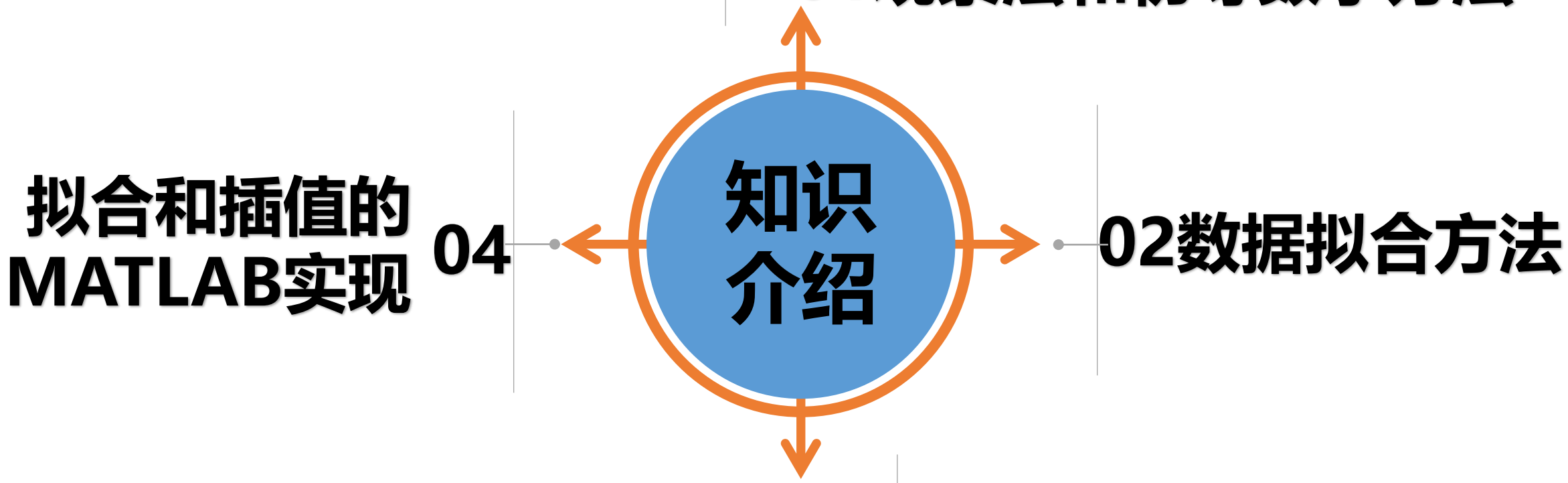
(3) 如果要创建一个新的打车软件服务平台，你们将设计什么样的补贴方案，并论证其合理性.

另外，无论我们怎样精心设计并极其细心地进行试验，我们仍需在拟合模型前评估数据的精确性．数据是如何收集的？收集过程中测量设备的精度如何?有没有疑问的点等。

Part 2

初等数据分析思想与建模方法

2.2初等数据分析思想与建模方法



2.2初等数据分析思想与建模方法

在分析一个数据集合时，可能遇到的问题是：

（1）根据收集的数据进行建模.

（2）按照选出的一个或多个模型（函数）类型对数据进行拟合.

（3）从一些已经拟合的模型（函数）类型中选取最合适的．例如，判断用指数模型是否比用最佳多项式模型要好.

2.2.1 观察法和初等数学方法

一、通过大量数据，利用变量之间的比例性质，得到自然规律

采集数据可以追溯到7000年前的尼罗河沿岸居民对潮水的数据记录.

最简单方式就是数据本身呈现比例关系.

17-18世纪应用数据建立了许多物理规律.

例 2.5 (Kepler(开普勒)第三定律)

开普勒应用第谷（Tycho Brahe)收集了 13 年有关火星的相对运动的观察资料．到 1609 年形成了他的头两条定律：a)每个行星都沿一条椭圆轨道运行，太阳在该椭圆的一个焦点处；b)对每个行星来说，在相等的时间里，该行星和太阳的连线扫过相等的面积．

开普勒花了许多年来验证并形成了第三定律，即轨道周期T（天数)与从太阳到行星的平均距离R之间的关系 $T = cR^{\frac{3}{2}}$ ．表 2.1 中的数据来自 1993 年的世界年鉴。

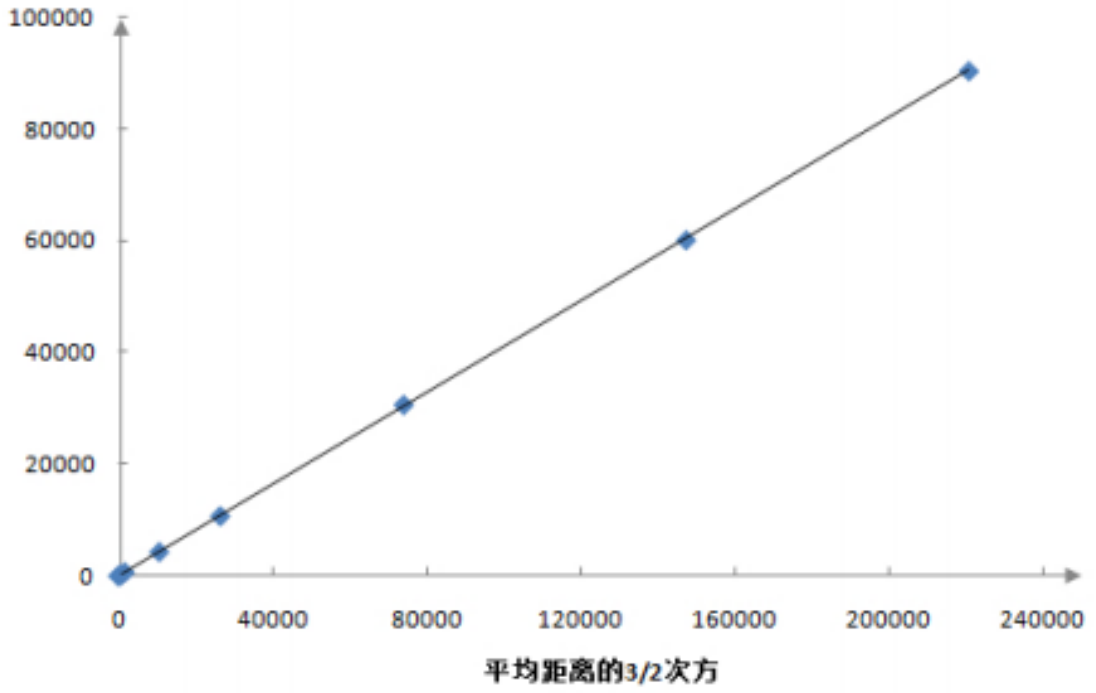
2.2初等数据分析思想与建模方法

表 2.1

行星	周期（天数）	平均距离 (百万英里)
水星	88.0	36
金星	224.7	67.25
地球	365.3	93
火星	687.0	141.75
木星	4331.8	483.80
土星	10 760.0	887.97
天王星	30 684.0	1764.50
海王星	60 1 88.3	2791.05
冥王星	90 4 66.8	3653.90

2.2初等数据分析思想与建模方法

图2.1画出了周期对平均距离的3/2次方的图形．该图形近似于一条通过原点的直线．纵坐标是周期（天数），横坐标量纲是英里×(10的负4次方)．



任取过原点的这条直线上的两点，很容易估计其斜率（比例常数）：斜率 $= \frac{90466.8-88}{220869.1-216} \approx 0.410$

$$T = 0.410R^{\frac{3}{2}}.$$

估计其模型为

17-19世纪建立大量物理定律的数学模型：

例 2.6 (波义耳定律 (Boyle’ s law) 一定量的理想气体的压强P、体积 V 和绝对温度T 之间具有定量关系

$$P = \frac{RT}{V}$$

R 是普适气体常量，即压强 P 的变化同时依赖于 V 和T

例 2.7 (虎克定律 (Hooke)1678 年) 一个线性弹簧，它的形变(x) 与弹力 (F) 之间的定量关系为

$$\boldsymbol{F = -Kx}$$

负号表示形变的方向与弹力方向相反。

例 2.8 (牛顿 (Newton)万有引力公式 1687 年) 考虑两个物体之间的相互作用时，对于它们之间的相互吸引力的数学模型

$$F = k\frac{m_1 m_2}{r^2}$$

这个数学模型及其理论是基于大量天文观测数据由牛顿在 17 世纪创立的.

例 2.9 (欧姆定律 (Ohm’s law)1826 年) 在同一电路中，通过某一导体的电流跟这段导体两端的电压成正比，跟这段导体的电阻成反比，这就是欧姆定律

$$I = \frac{U}{R}, \qquad U = IR, \qquad R = \frac{U}{I}$$

公式中物理量的单位：I：(电流）的单位是安培（A）、U：（电压）的单位是伏特（V）、R：（电阻）的单位是欧姆（Ω）

大量经济学领域的函数模型如下：

(1)生产函数-在一定技术条件下生产要素投入量与产品的最大产出量之间的定量关系。

(2)需求函数-需求量与价格直降的函数关系

(3)供给函数-供给量与价格直降的函数关系

(4)总成本函数；　　(5)总收益函数；

(6)总利润函数；　　(7)效用函数；

(8)消费函数；　　(9)储蓄函数。

二、通过观察，利用初等数学的知识建立数学模型

例2.17由于地面凹凸不平，我们很难将椅子一次放稳，由此提出如下问题：将4条腿长相同的方椅子放在不平的地上，怎样才能放平？如何才能把它抽象成数学问题？

问题分析：

假定椅子中心不动，每条腿的着地点视为几何上的点，用 A、B、C、D 表示，把 AC 和 BD连线看做坐标系中的 x 轴和 y 轴，把转动椅子看做坐标的旋转，如图 2.3.

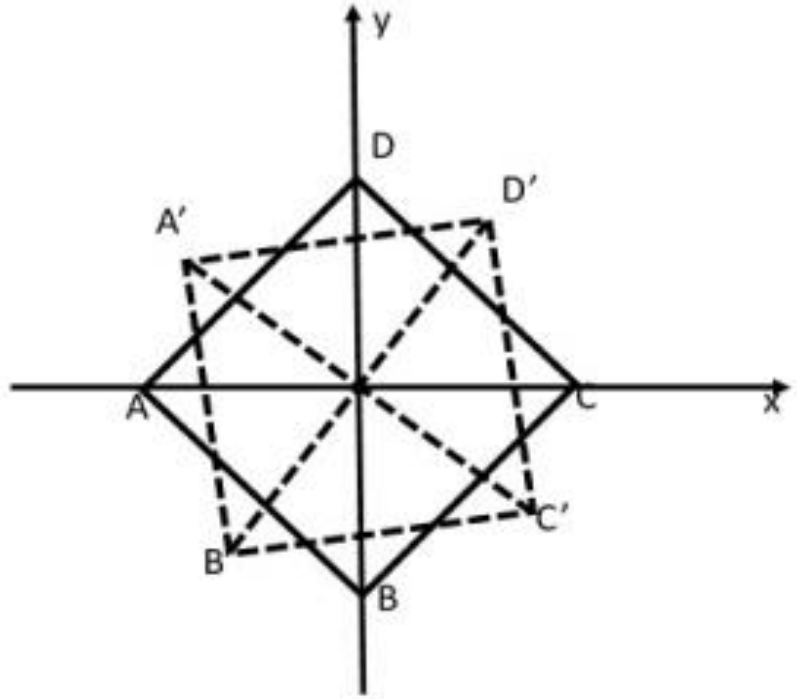


图 2.3　椅子旋转图

用 θ 表示对角线 AC 转动后与初始位置 x 轴正向的夹角．设 g(θ) 表示 A，C 两腿旋转 θ角度后与地面距离之和．f(θ) 表示 B，D 两腿旋转 θ 角度后与地面距离之和．

当地面形成的曲面为连续函数时，f(θ)，g(θ) 皆为连续函数．因为三条腿总能同时着地，即对任意 θ，总有

$$f(\theta)\cdot g(\theta)=0$$

不妨设初始位置 θ = 0 时 g(θ) = 0，f(θ) > 0，于是问题转化为：是否存在一个 θ_0，使

$$f(\theta_0) = g(\theta_0) = 0$$

这样，椅子问题就抽象成如下数学问题：已知 f(θ)，g(θ) 连续，g(0) = 0，f(0) > 0，且对任意的 θ 都有$f(\theta) \cdot g(\theta) = 0$．求证：存在$\theta_0$，使得

$$f(\theta_0) = g(\theta_0) = 0$$

数学问题的证明：

令

$$h(\theta) = g(\theta) - f(\theta)$$

则

$$h(0) = g(0) - f(0) < 0$$

将椅子转动 $\frac{\pi}{2}$**，即将AC与BD位置互换，则有** $g\left(\frac{\pi}{2}\right) > 0, f\left(\frac{\pi}{2}\right) = 0,$

所以

$$h\left(\frac{\pi}{2}\right)=g\left(\frac{\pi}{2}\right)-f\left(\frac{\pi}{2}\right)>0$$ 而h(θ)是连续函数，根据连续函数零点定理知必存在$\theta_0\in(0,\frac{\pi}{2})$ 使得 $h(\theta_0)=0$

即 $g(\theta_0)=f(\theta_0)$ 又由条件对任意θ，恒有 $f(\theta)\cdot g(\theta)=0$

所以

$$g(\theta_0)=f(\theta_0)=0$$

即存在θ_0方向，四条腿能同时着地.

所以椅子问题的答案是：如果地面为光滑曲面，

椅子中心不动最多转动角度，则四条腿一定可以同时着地。

2.2.2 数据拟合方法

一、源头问题

统计学：用部分数据反映整体或整体的趋势。

通过部分数据获得变量之间的函数关系，即能否根据一组试验观测数据(部分数据)找到变量之间的函数关系(整体或整体的趋势)。

也就是，从一组试验观测数据(x_i,y_i)，$i = 0,1,\ldots,n$找到自变量x与因变量y之间的函数关系，用某近似函数$y = f(x)$来表示。

函数$y = f(x)$的产生办法可采用数据拟合与函数插值两种办法来实现

2.2.2 数据拟合方法

例如实验测得如下一列数据

X	-3	-2	-1	0	1	2	3
Y	-8.0942	-3.0942	-0.0942	0.9058	-0.0942	-3.0942	-8.0942

作如下散点图 2.4.

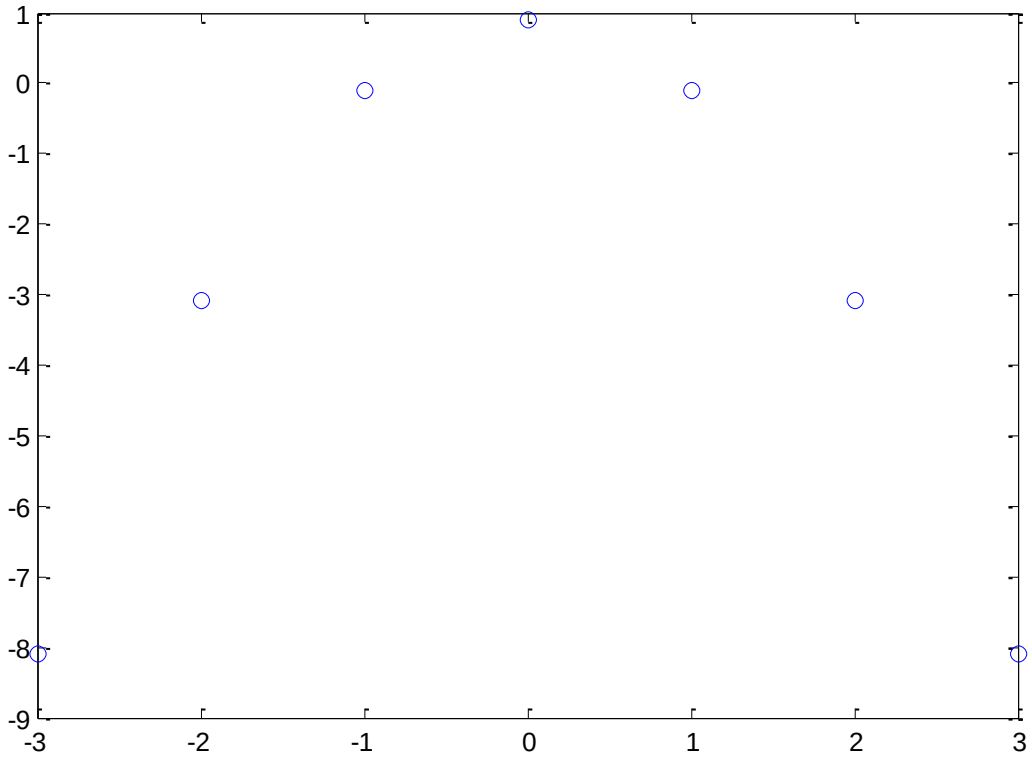


图 2.4 散点图

问题 1：请找出一个函数经过所有数据点。

问题 2：请预测当 x = 3.5 时，y 的值.

(1)上述问题分别对应拟合和插值问题.

(2)拟合和插值都是要求通过已知的观测数据去寻求某个近似函数，使得近似函数与已知数据有较高的拟合精度.

(3)拟合和插值在数学方法上完全不同 .

具体来说：

数据拟合：求近似函数，不要求过所有的已知数据点，只要求在某种意义下它在这些点上的总偏差最小．主要用来反应数据的基本趋势．

即数据拟合主要是考虑到观测数据寻求整体误差最小、能较好反映观测数据的近似函数$y = f(x)$，此时并不要求所得到的近似函数$y = f(x)$满足

$$y_i = f(x_i),\ i = 0,1,\cdots,n。$$

插值：求过已知有限个数据点的近似函数，要求所求的近似函数过已知的数据点.

即函数插值要求近似函数y = f (x)在每一个观测x_i处一定要满足

$$y_i = f(x_i),\ i = 0,1,\cdots,n。$$

在这种情况下，通常要求观测数据相对比较准确，即不考虑观测误差的影响。

二、数据拟合思想与建模方法

数据拟合有几种不同的判别准则：

使偏差的绝对值之和最小；

使偏差的最大绝对值最小；

使偏差的平方和最小(即最小二乘法)。

1、Chebyshev 近似准则

假设想要对数据点集用一条直线y = ax + b近似，应如何选择 a 和 b，使直线最好地近似数据？从图上看，当多于两个点时，它们不可能精确地处于一条直线上，一些数据点和直线间总存在一些纵向差异，我们称这些纵向差异为绝对偏差 .

定义 2.1 给定 m+1 个数据点(x_i,y_i)，$i = 0,1,\ldots,m$的集合，用直线y = ax + b 拟合该集合，确定参数 a 和b，使任一数据(x_i,y_i)和其对应的直线上的点$(x_i, ax_i+ b)$间的距离之和最小，即：极小化绝对偏差

$$|y_i- y(x_i)|$$

的和.

将直线的极小化绝对偏差之和准则推广到给定曲线情形：

给定某一函数类型y=f(x)，以及m个数据点(x_i, y_i)的集合，极小化绝对偏差$|y_i - y(x_i)|$的和，也就是确定函数y = f(x)的参数，极小化

$$\sum_{i=1}^{m} |y_i - y(x_i)|$$

再看另一种选择

定义 2.2 给定m个数据点的集合 (x_i, y_i)，用直线y=ax+b拟合该集合，确定参数 a 和 b，使任一数据点(x_i, y_i)和其对应的直线上的点$(x_i, ax_i + b)$间的距离最小，也就是对整个数据点集极小化最大绝对偏差$|y_i - y(x_i)|$.

现推广到给定曲线的情形：

给定某函数y=f(x)和m个数据点(x_i, y_i)的一个集，对整个集合极小化最大绝对偏差$|y_i - y(x_i)|$，即确定函数类型y=f(x)的参数从而极小化数量：

$$\text{Max } |y_i - y(x_i)|, i=1, 2, \ldots, m$$

这一重要的准则常称为Chebyshev近似准则

例 2.18 设我们要度量图2.5表示的线段AB，BC和AC，假定你的测量的结果为AB=13、BC=7、AC=19.

这时，AB和BC值加起来是20而不是测出的19．现在用Chebyshev准则来解决这一个单位的差异，也就是用一个方法为三个线段指定数值，使得指定的和观测的任一对应数之间的最大偏差达到极小.

图 2.5

解 假定对每一次测量有相同的信任度，这样每一测量值有相等的权值．于是，差异应均等地分配到每一线段．令x_1代表线段AB长度的真值，x_2代表BC的真值．为易于表示，令r_1、r_2、r_3表示真值和测量值间的差异．即

$$\begin{cases} x_1 - 13 = r_1(\text{线段}AB) \\ x_2 - 7 = r_2(\text{线段}BC) \\ x_1 + x_2 - 19 = r_3(\text{线段}AC) \end{cases}$$

数值r_1、r_2、r_3称为残差．

如果用 Chebyshev 近似准则，应指定r_1、r_2、r_3的值，使三个数值$|r_1|$、$|r_2|$、$|r_3|$的最大者达到最小．如果记最大的数为 r，那么我们要求最小化 r，三个约束条件为

$$\begin{cases} |r_1| \le r \text{或} -r \le r_1 \le r \\ |r_2| \le r \text{或} -r \le r_2 \le r \\ |r_3| \le r \text{或} -r \le r_3 \le r \end{cases}$$

这些条件的每一个对应两个不等式．

例如$|r_1|\leq$r能替换r$-r_1\geq$0和r$+r_1\geq$0,其他类似. 问题则叙述为经典的数学问题:

$$\min r$$

$$\text{s.t.}\begin{cases} r-x_1+13\geq 0(r-r_1\geq 0) \\ r+x_1-13\geq 0(r+r_1\geq 0) \\ r-x_2+7\geq 0(r-r_2\geq 0) \\ r+x_2-7\geq 0(r+r_2\geq 0) \\ r-x_1-x_2+19\geq 0(r-r_3\geq 0) \\ r+x_1+x_2-19\geq 0(r+r_3\geq 0) \end{cases}$$

这一问题称为线性规划问题.

推广这一过程，给定某一函数类型 y = f(x)，其参数待定，以及给定 m 个数据点 (x_i, y_i)的一个集合，并确定出残差为 $r_i = y_i - f(x_i)$. 如果 r 代表这些残差的最大绝对值，那么问题表示如下

$$\min r$$

$$\text{s.t.} \quad \begin{matrix} r-r_i \geq 0 \\ r+r_i \geq 0 \end{matrix} \quad \text{其中，i=1,2,...，m.}$$

2、最小二乘准则

问题：确定函数类型 y = f(x) 的参数，极小化和数

$$\sum_{i=1}^{m} |y_i - \mathbf{f}(x_i)|^2$$

例 2.19 下表是收集到的数据

x	1	2	3	4
z	8.1	22.1	60.1	165

画出这批数据在直角坐标上的散点图．如图 2.6看起来两者呈指数关系，因此可设 z 与温度 x 的关系为

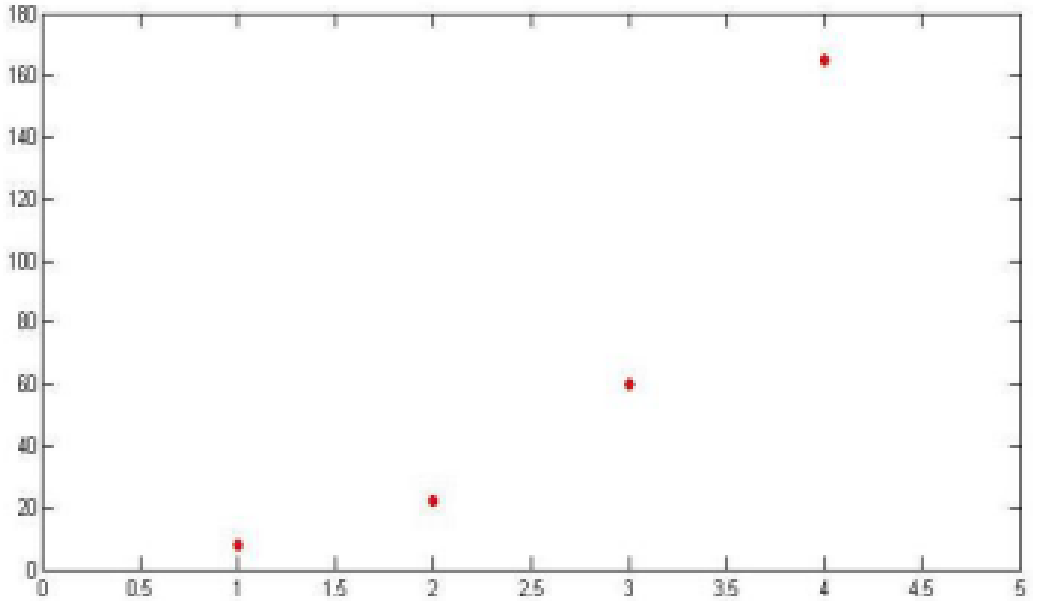


图 2.6　z 对 x 的图

$$z = \beta e^{\alpha x}$$

任务：具体确定常数α，β．上式两边取对数，令

$$y = \ln z,\ a = \alpha,\ b = \ln\beta,$$

则原式变成了线性关系

$$y = ax + b$$

表格变为：

x	1	2	3	4
$y=\ln z$	2.1	3.1	4.1	5.1

散点图变为图 2.7：

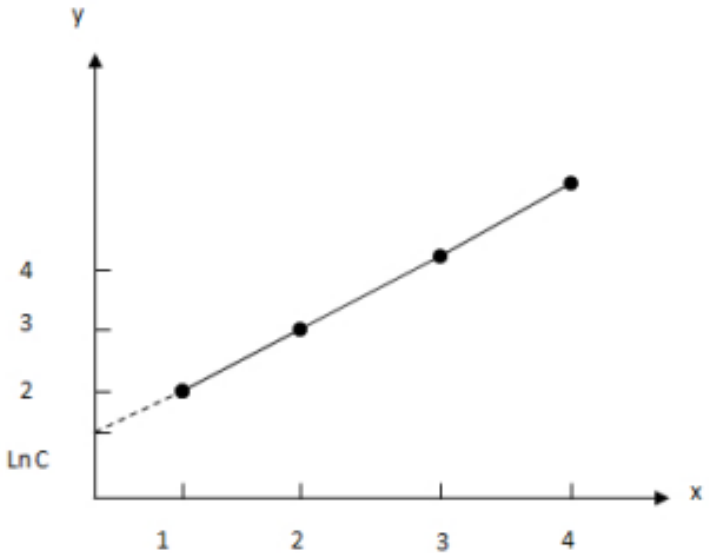


图 2.7　lnz 对 x 的图

于是，问题化为找一直线 y = ax + b，即求 a，b使得上表中的数据基本满足这个函数关系．使得所有观测值 y_i 与函数值 ax_i + b 之偏差的平方和

$$Q=\sum_{i=1}^{m}|y_i - ax_i - b|^2$$

最小.

确定常数 a，b 用的就是二元函数求极值的方法，显然 Q 是 a，b 的函数．令

$$\frac{\partial Q}{\partial a} = -2\sum_{i=1}^{n}(y_i - ax_i - b)x_i = 2a\sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}x_i y_i + 2b\sum_{i=1}^{n}x_i = 0$$

$$\frac{\partial Q}{\partial b} = -2\sum_{i=1}^{n}(y_i - ax_i - b) = 2a\sum_{i=1}^{n}x_i - 2\sum_{i=1}^{n}y_i + 2nb = 0$$

就得到线性方程组

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

解这个方程组，得到

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

由问题知，Q 在这个(a,b)点取最小值。

现在解决本段开始提出的问题.从表2.4 可得表2.5:

i	1	2	3	4
x_i	1	2	3	4
y_i	2.1	3.1	4.1	5.1

通过计算可得

$\sum_{i=1}^{4} x_i$	$\sum_{i=1}^{4} x_i^2$	$\sum_{i=1}^{4} x_i y_i$	$\sum_{i=1}^{4} y_i$	a	b	e^b
10	30	41	14.4	1	1.1	3

由表2.6可知

$$a = 1, b = 1.1$$

于是，表的拟合直线方程为

$$y = x + 1.1,$$

z 与 x 的关系为 $z = 3e^x$

2.2.3 插值方法

一、源头问题

已知某函数y = f(x)(未知)的一组观测或试验数据(x_i, y_i)(i = 0, 1, 2, • • • , n)，要寻求一个函数φ(x)，使得

$$\varphi(x_i) = y_i \ (i = 0, 1, 2, \cdots, n)$$

则

$$\varphi(x) \approx f(x).$$

具体而言，实际中在不知道函数 y = f(x) 的具体表达式的情况下，对于

x = x_i有实验测量值

$$y = y_i \ (i = 0, 1, 2, \cdots, n),$$

寻求另一函数 φ(x) 使满足

$$\varphi(x_i) = y_i = f(x_i) (i = 0, 1, 2, \cdots, n)$$

称此问题为一维插值问题.

并称函数 φ(x) 为 f(x) 的插值函数，x_i (i = 0, 1, 2, ・・・ ,n)称为插值结点，φ(x_i) = y_i (i = 0, 1, 2,・・・,n)称为插值条件，则 φ(x) ≈ f(x).

插值问题除了一维插值问题外，还有二维插值问题.

下面介绍几种基本的、常用的一维插值方法：

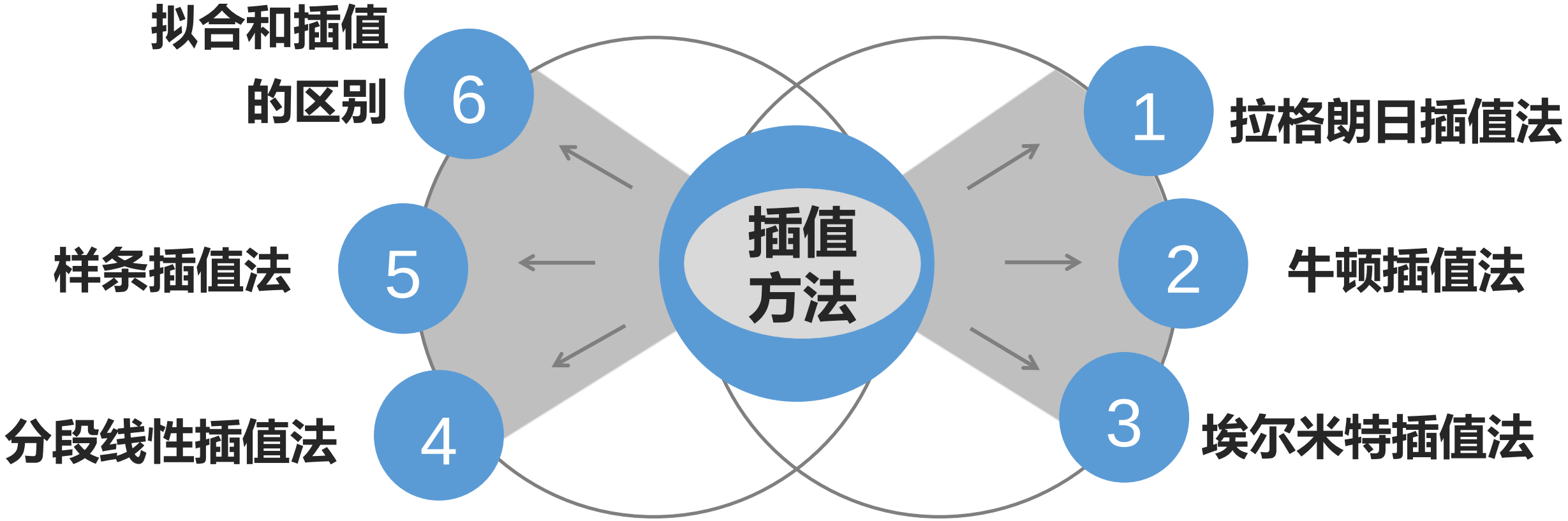
拉格朗日插值法

牛顿插值法

Hermite插值法

分段线性插值法

三次样条插值法。


插值
方法
1
拉格朗日插值法
2
牛顿插值法
3
埃尔米特插值法
4
分段线性插值法
5
样条插值法
6
拟合和插值
的区别

二、插值思想与建模方法

1.拉格朗日（Lagrange）插值:

（1）插值多项式简介

已知函数y=f(x)在n+1个相异点

x_i (i = 0, 1, 2, • • • ,n)

上的函数值为

y_i (i = 0, 1, 2, • • • ,n)

要求一个次数不超过 n 的代数多项式

$$p_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

使在结点 x_i 上有

$$p_n(x_i) = y_i \quad (i = 0, 1, 2, \cdots, n)$$

成立，称$p_n(x)$为插值多项式。

则f(x)的n+1个待定系数a_0，a_1，…，a_n满足

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = y_0 \\ a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = y_1 \\ \cdots\cdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = y_n \end{cases}$$

记此方程组的系数矩阵为 A，则

$$\det(A)=\begin{vmatrix}1 & x_0 & \cdots & x_0^n\\ 1 & x_1 & \cdots & x_1^n\\ \vdots & \vdots & & \vdots\\ 1 & x_n & \cdots & x_n^n\end{vmatrix}$$

是范德蒙德行列式．当 x_0，x_1，x_2，…，x_n 互不相同时，此行列式值不为零．因此，方程组有唯一解．这表明，只要 n + 1 个插值节点x_0，x_1，x_2，…，x_n互异，满足插值条件的插值多项式存在唯一

当 x ∈ [a, b] 且 x≠x_i (i = 0, 1, 2, • • • ,n) 时，称被插函数f(x)与插值函数多项式P_n(x)之间的差

$$R_n(x)=\mathrm{f}(x)-P_n(\mathrm{x})$$

为插值多项式P_n(x)的截断误差，或插值余项.

即用多项式函数$P_n(x)$作为插值函数时，希望通过解方程组而得到待定系数

$$a_0,\ a_1,\ \cdots,\ a_n$$

的做法，当n比较大时是不现实的。因此，我们采用拉格朗日(Lagrange)插值多项式。

（2）拉格朗日 (Lagrange)插值多项式

首先构造一组基函数

$$l_i(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j}$$

$$= \frac{(x - x_0)\cdots(x - x_{i-1})(x - x_{i+1})\cdots(x - x_n)}{(x_i - x_0)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}$$

$$(i = 0, 1, \cdots, n)$$

显然$l_i(x)$是 n 次多项式，且满足：

令
$$l_i(x_j)=\begin{cases}0 & j\neq i\\ 1 & j=i\end{cases}$$

$$p_n(x)=\sum_{i=0}^{n} y_i l_i(x)=\sum_{i=0}^{n} f(x_i)\prod_{j=0,j\neq i}^{n}\frac{x-x_j}{x_i-x_j}$$

称$P_n(x)$即为n次拉格朗日(Lagrange)插值多项式，同样由唯一性，n+1个节点的n次拉格朗日插值多项式存在且唯一.

当 f(x) 在 [a, b] 上充分光滑时，利用罗尔 (Rolle)定理可推出：对于任意

x ∈ [a, b]，插值多项式 P_n(x)的余项

$$R_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i), \xi \in (a, b).$$

例2.20　设f(x)=$\sqrt[3]{x}$ 取结点为 x = 1、1.728、2.744，求f(x)的二次拉格朗日插值多项式P_n(x)及其余项的表达式，并计算P_2(2)

($\sqrt[3]{2}$= 1.2599210・・・).

解 取 x_0= 1，x_1= 1.728，x_2= 2.744 为插值结点，则函数f(x)= $\sqrt[3]{x}$的相应的函数值为f(x_0)=1，f(x_1)=1.2，f(x_2)=1.4。于是，由拉格朗日插值公式，

$$
\begin{aligned}
f(x) &\approx \boldsymbol{p}_2(x) \\
&= 1 \cdot \frac{(\boldsymbol{x}-1.728)(\boldsymbol{x}-2.744)}{(1-1.728)(1-2.744)} \\
&+1.2 \cdot \frac{(\boldsymbol{x}-1)(\boldsymbol{x}-2.744)}{(1.728-1)(1.728-2.744)} \\
&+1.4 \cdot \frac{(\boldsymbol{x}-1)(\boldsymbol{x}-1.728)}{(2.744-1)(2.744-1.728)} \\
&\approx -0.0447x^2 + 0.3965x + 0.6481
\end{aligned}
$$

将 x = 2 代入就得到$\sqrt[3]{2}$ 的近似值

$$\sqrt[3]{2}\approx P_2(2)=1.2626$$

它与准确值的差的绝对值（称为绝对误差)约为 0.0027，而由插值余项估计公式，其误差约为

$$|R_n(2)| = \left|\frac{5}{81}\cdot\frac{(2-1)(2-1.728)(2-2.744)}{\xi^{\frac{8}{3}}}\right| \leq 0.0125$$

2. 牛顿（Newton）插值：

(1)函数的差商及其性质

设有函数f(x)，其中取x_0，x_1，x_2 ,…，x_n表示一系列互不相同的节点，可定义以下差商：

一阶差商：$f[x_i, x_j] = \frac{f(x_i)-f(x_j)}{x_i-x_j}$

二阶差商：$f[x_i, x_j, x_k] = \frac{f[x_i,x_j]-f[x_j,x_k]}{x_i-x_k}$

n 阶差商：

$$f[x_0, x_1, \dots, x_n] = \frac{f[x_0, x_1, \dots, x_{n-1}] - f[x_1, x_2, \dots, x_n]}{x_0 - x_n}$$

(I)差商的可加性

$$f[x_0, x_1, \dots, x_n] = \sum_{k=0}^{n} \frac{f(x_k)}{\prod_{j=0, j\neq k}^{n} (x_k - x_j)}$$

(II)差商的对称性：在f[x_0，x_1，x_2，…,x_n]中任意调换x_i，x_j的次序其值不变.即有：

$$f[x_0, \cdots, x_i, \cdots, x_j, \cdots, x_n] = f[x_0, \cdots, x_j, \cdots, x_i, \cdots, x_n]$$

$$f(x) = f(x_0) + (x - x_0)f[x, x_0]$$
$$f[x, x_0] = f[x_0, x_1] + (x - x_1)f[x, x_0, x_1]$$
$$f[x, x_0, x_1] = f[x_0, x_1, x_2] + (x - x_2)f[x, x_0, x_1, x_2]$$
$$\cdots$$
$$f[x, x_0, \cdots, x_{n-1}] = f[x_0, x_1, \ldots, x_n] + (x - x_n)f[x, x_0, \cdots, x_n]$$

将以上各式分别乘以 1,$(x-x_0)$,

$(x-x_0)(x-x_1)$, …,

$(x-x_0)(x-x_1)\cdots(x-x_{n-1})$，然后等式两边相加可得

$$
\begin{aligned}
&f(x)=f(x_0)+(x-x_0)f[x_0,x_1]\\
&+\cdots+\\
&+(x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_0,x_1,\cdots,x_n]\\
&+(x-x_0)(x-x_1)\cdots(x-x_n)f[x,x_0,x_1,\cdots,x_n]
\end{aligned}
$$

记

$$N_n(x) = f(x_0) + (x - x_0)f[x_0, x_1] + \cdots +$$

$$(x - x_0)(x - x_1)\cdots(x - x_{n-1})f[x_0, x_1, \cdots, x_n]$$

显然 $N_n(x)$ 是至多 n 次多项式，且满足插值条件 $N_n(x_i)=f(x_i)$ 。这种形式的插值多项式称为Newton插值多项式.

优点：每增加一个节点，插值多项式只增加一项，即

$$N_{n+1}(x) = N_n(x) +$$

$$(x - x_0)(x - x_1)\cdots(x - x_n)f[x_0, x_1, \cdots, x_{n+1}]$$

便于进行递推运算，且计算量小于Lagrange插值．余项为:

$$\mathrm{R}\ (x) = (x - x_0)(x - x_1)\cdots(x - x_n)f[x_0, x_1, \cdots, x_n, x]$$

3. 埃尔米特（Hermite）插值：

如果对插值函数，要求它在节点处

与函数同值，

有一阶、二阶或高阶导数同值

这就是埃尔米特插值问题.

本节仅考虑在节点处函数值及一阶导数同值的埃尔米特插值。

一般提法为： 设已知函数y=f(x)在 n+1 个互异节点 x_0，x_1，x_2,…，x_n上的

函数同值 $y_i = f(x_i)$ 和导数同值 $y_i' = f'(x_i)$

要求一个至多2n+1次多项式H(X)，使得

$$H(x_i) = y_i \qquad H'(x_i) = y_i'(i = 0,1,\dots,n)$$

满足上述条件的多项式H(x)称为埃尔米特插值多项式。

其具体形式如下所示，

$$H(x)=\sum_{i=0}^{n} h_i[(x_i-x)(2a_iy_i-y_i')+yi]$$

其中 $h_i=\prod_{j=0,j\neq i}^{n}\left(\frac{x-x_j}{x_i-x_j}\right)^2$

$$a_i=\sum_{j=0,j\neq i}^{n}\frac{1}{x_i-x_j}$$

高次插值多项式的龙格 (Runge)现象

1、多数情况下，$P_n(x)$ 的次数越高，逼近 f(x) 的效果越好.

2、但是往往会造成 $P_n(x)$ 的收敛性与稳定性变差，逼近效果不理想，甚至发生龙格 (Runge)现象。

龙格 (Runge)现象：

在[−1,1]上用n+1个等距节点作函数 f(x)=1/(1+25x^2)的插值多项式 P_n(x)，则随着n的增大， P_n(x)振荡越来越大．计算结果与理论证明表明，当n趋于无穷大时，P_n (x)在区间中部收敛于f(x)，但对满足条件 0.726…≤ |x| < 1 的 x， P_n(x) 并不收敛于 f(x)．

4.分段线性插值法

分段线性插值就是将每两个相邻的节点用直线连起来，形成一条折线就是分段线性插值函数，记作$I_n(x)$,它满足$I_n(x_i)=y_i$，且$I_n(x)$在每个小区间$[x_i, x_{i+1}]$上是线性函数.

具体表示如下：$I_n(x) = \sum_{i=0}^{0} y_i l_i(x)$

其中 $l_i(x) =$

$$\begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}}, x \in [x_{i-1}, x_i] \\ \frac{x-x_{i+1}}{x_i-x_{i+1}}, x \in [x_i, x_{i+1}] \\ 0, \text{ 其他} \end{cases}$$

这样构造的I_n(x) 有良好的收敛性，即对于 x ∈ [a, b] 有

$$\lim_{n\to\infty} I_n(x) = f(x).$$

5.样条插值法

设给定区间[a, b]的一个分划

Δ： $a=x_0 < x_1 < \cdots < x_n = b$

如果函数 s(x) 满足条件：

(1)在每个子区间 $[x_{i-1}, x_i](i = 1,2, \ldots, n)$ 上是 k 次多项式；

(2)s(x)及直到k−1阶的导数在[a,b]上连续;

则称 s(x) 是关于分划 △ 的一个 k 次多项式样条函数， x_0， x_1，x_2,…， x_n称为样条结点， x_1， x_2,…, x_{n-1}称为内结点， x_0， x_n称为边界结点，这类样条函数的全体记作 $S_p(\Delta,k)$ 称为k 次样条函数空间.

若$s(x) \in S_p(\Delta, k)$，则 s(x) 是关于分划 Δ 的 k 次多项式样条函数.k 次多项式样条函数的一般形式为

$$s_k(x) = \sum_{i=0}^{k} \frac{\alpha_i x^i}{i!} + \sum_{j=1}^{n-1} \frac{\beta_j}{k!} (x - x_j)_+^k$$

其中α_i(i = 0, 1, · · · , k) 和β_j(j = 1, 2, · · · , n − 1) 均为任意常数，而

$$(x - x_j)_+^k = \begin{cases} (x - x_j)^k, x \geq x_j, (j, =, 1,2 \cdots n-1) \\ 0, x < x_j \end{cases}$$

k=2和3，为二次样条函数和三次样条函数。

二次样条函数：对于 [a, b] 上的分划 $\triangle$: a = $x_0 < x_1 < x_2 <$ x_n = b，则

$$s_2(x) = \alpha_0 + \alpha_1 x + \frac{\alpha_2}{2!}x^2 + \sum_{j=1}^{n-1} \frac{\beta_j}{2!}(x - x_j)_+^2 \in S_p(\Delta, 2),$$

其中

$$(x - x_j)_+^2 = \begin{cases} (x - x_j)^k, x \geq x_j \\ 0, x < x_j \end{cases}, (j = 1,2, \cdots, n-1)$$

三次样条函数：对于 [a，b] 上的分划 △ ：a = x_0>x_1>x_2 >x_n =b，则

$$s_3(x) = \alpha_0 + \alpha_1 x + \frac{\alpha_2}{2!}x^2 + \frac{\alpha_3}{3!}x^3 + \sum_{j=1}^{n-1}\frac{\beta_j}{3!}(x-x_j)_+^3 \in S_p(\Delta,3),$$

其中

$$(x-x_j)_+^3 = \begin{cases}(x-x_j)^k, x \geq x_j \\ 0, x < x_j\end{cases}, (j = 1,2,\cdots,n-1)$$

插值与拟合的区别：

共性：都是通过已知部分数据求整体的近似函数.

区别：数据拟合不要求近似函数通过所有数据点，而是要求它能较好地反映数据整体变化趋势．插值问题要求所得的近似函数（曲线或曲面）经过所已知的所有数据点.

2.2.4 拟合与插值的 MATLAB 编程实现

一、Matlab 多项式拟合

a = polyfit(x, y, n):多项式拟合，返回降幂排列的多项式系数．其中 x，y 是数据点的值，n 为拟合的最高次数．

y = polyval(a, x):计算拟合的多项式在 x 处的值．

在处理一些无约束条件的最小二乘拟合时往往会涉及到最小二乘优化，最小二乘优化是一类比较特殊的优化问题，在处理这类问题时，Matlab 也提供了一些强大的函数．在 Matlab 优化工具箱中，用于求解最小二乘优化问题的函数有：lsqlin、lsqcurvefit、lsqnonlin、lsqnonneg

二、Matlab 数据拟合工具箱

在 Matlab 中的工作区，输入命令 cftool，便会出现如下拟合工具箱

5 个命令按钮的功能分别如下：

Data 按钮：可输出、查看和平滑数据；

Fitting 按钮：可拟合数据、比较拟合曲线和数据集；

Exclude 按钮：可以从拟合曲线中排除特殊的数据点；

Ploting 按钮：在选定区间后，单击按钮，可以显示拟合曲线和数据集；

三、Matlab 一维插值

用 Matlab 实现分段线性插值不需要编制函数程序，Matlab 中关于一维插值的函数为 interp1 .

y = interp1(x0, y0, x,' method'),

其中 method 指定插值的方法，默认为线性插值．其值可为：' nearest' 最近项插值，' linear' 线性插值，' spline' 逐段 3 次样条插值，' cubic' 保凹凸性 3 次插值.

详细情况请使用 help interp1;

四、Matlab 二维插值

当插值节点为网格节点时，命令为

z = interp2(x0, y0, z0, x, y,' method'),

其中 x0，y0 分别为 m 维和 n 维向量，表示节点，z0 为 m × n 维矩阵，表示节点值，x，y 为一维数组，表示插值点，x 与 y 应是方向不同的向量，即一个是行向量，另一个是列向量，z 为矩阵，它的行数为 x 的维数，列数为 y 的维数，表示得到的插值，' method' 的用法同上面的一维插值.

Part 3

案例分析

案例分析

- 案例一 机床加工
- 案例二 海底样貌探测
- 案例三 醉汉行进路线

案例一： 机床加工(理工类)

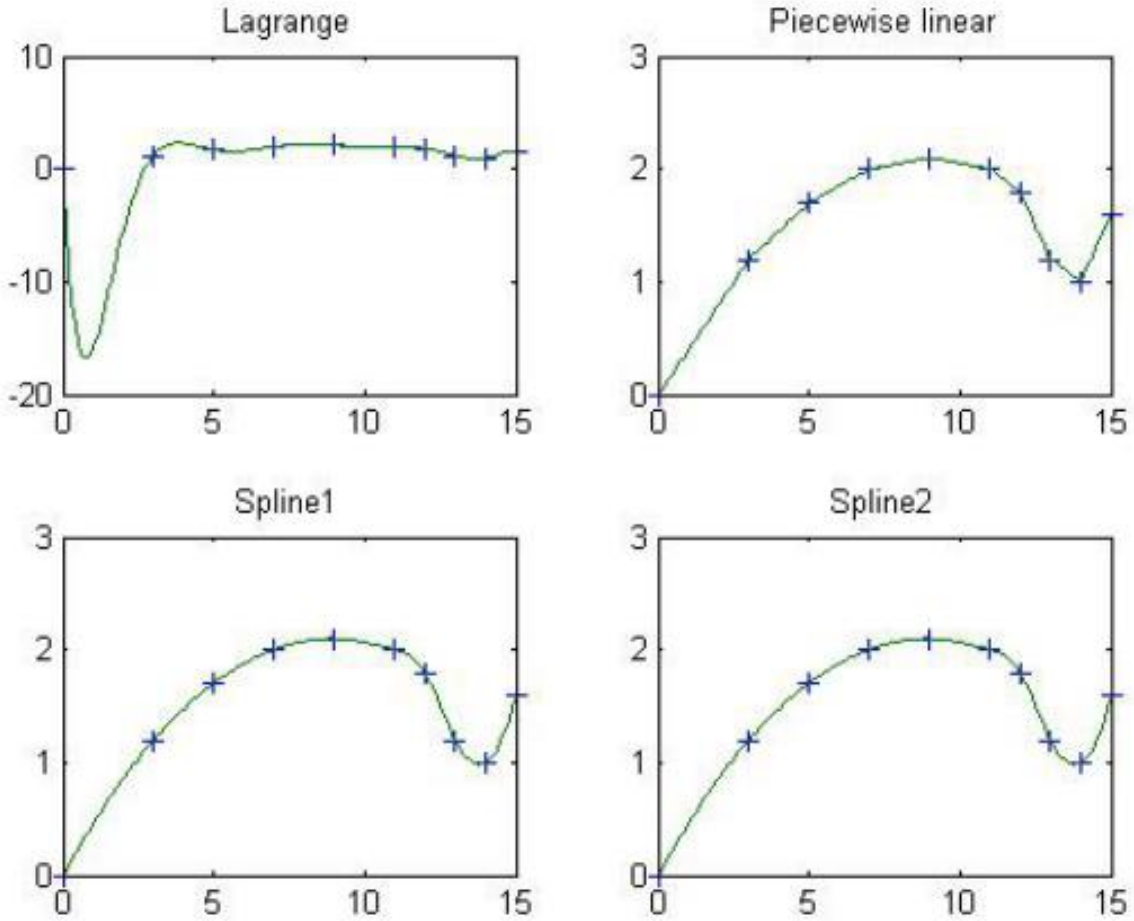
问题背景：在工业加工时，往往会对加工零件的外形有很高的要求．一般在单纯考虑平面的情况下，待加工零件的外形根据工艺要求由一组数据（x，y）给出，用程控铣床加工时每一刀只能沿 x 方向和 y 方向移动非常小的一步，这就需要从已知数据得到加工所要求的步长很小的（x，y）坐标．如下表所示给出的 x，y 数据是位于机翼断面的下轮廓上，假设需要得到 x 坐标每改变 0.1 时的 y 坐标．请完成加工所需数据并画出曲线。

x	0	3	5	7	9	11	12	13	14	15
y	0	1.2	1.7	2.0	2.1	2.0	1.8	1.2	1.0	1.6

模型构建与求解

根据题意进行分析，我们会发现这是一个简单的插值问题，用三次样条插值计算所得曲线如下图 2.13 所示。

因为所求的加工数据比较多，这就不一一列出了。


Lagrange
10
0
-10
-20
0
5
10
15
Piecewise linear
3
2
1
0
0
5
10
15
Spline1
3
2
1
0
0
5
10
15
Spline2
3
2
1
0
0
5
10
15

案例二 海底地貌探测（社科类）

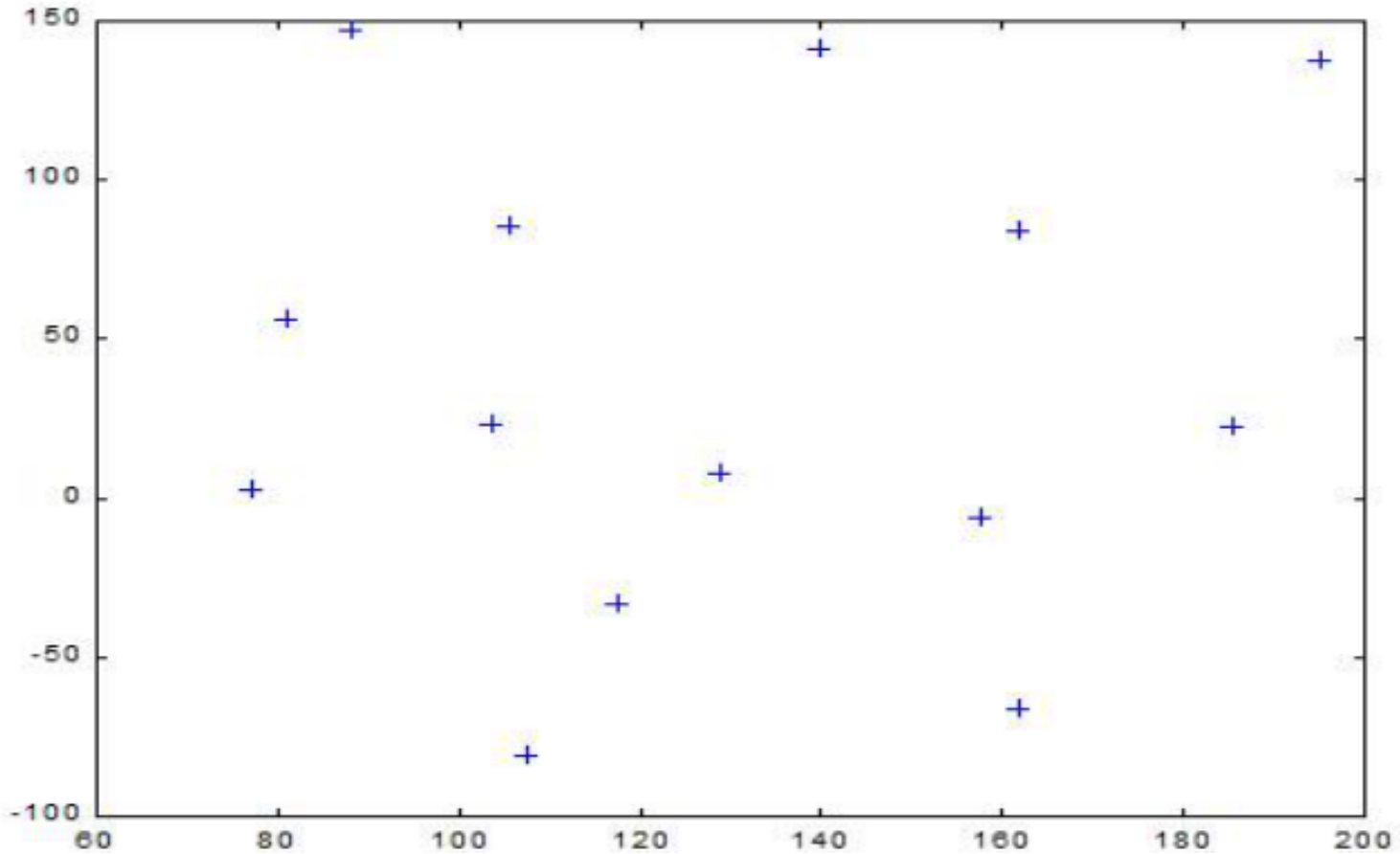
问题背景：现已知某海域测得一些点（x，y）处的水深 z 由下表给出，船的吃水深度为 5 英尺,所以船要避免进入水深在 5m 以下的水域．试画出海域的地貌图，以及在矩形区域(75，200）X（-50，150）里的哪些地方船要避免进入。

x	129	140	103.5	88	185.5	195	105
y	7.5	141.5	23	147	22.5	137.5	85.5
z	4	8	6	8	6	8	8
x	157.5	107.5	77	81	162	162	117.5
y	-6.5	-81	3	56.5	-66.5	84	-33.5
z	9	9	8	8	9	4	9

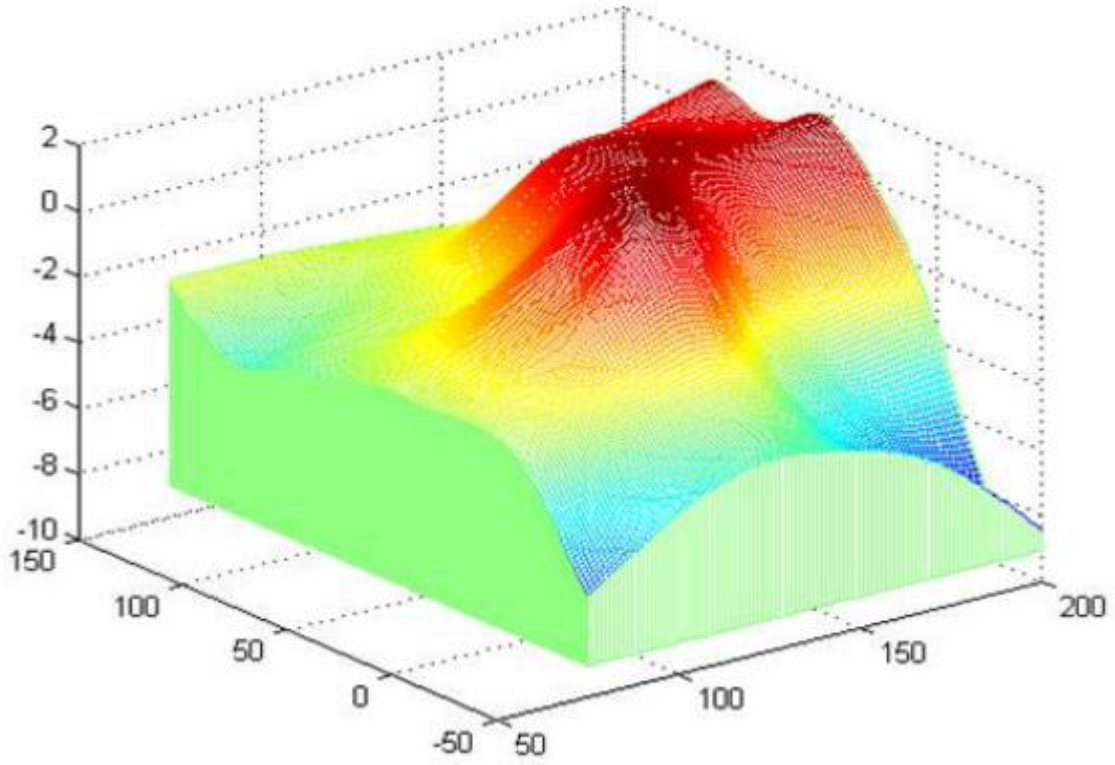
模型构建与求解

假设该海域海底是平滑的，由于测量点是散乱分布的，先在平面上作出测量点的分布图，再利用二维插值方法补充一些点的水深，然后作出海底曲面图和等高线图，并求出水深小于 5的海域范围.

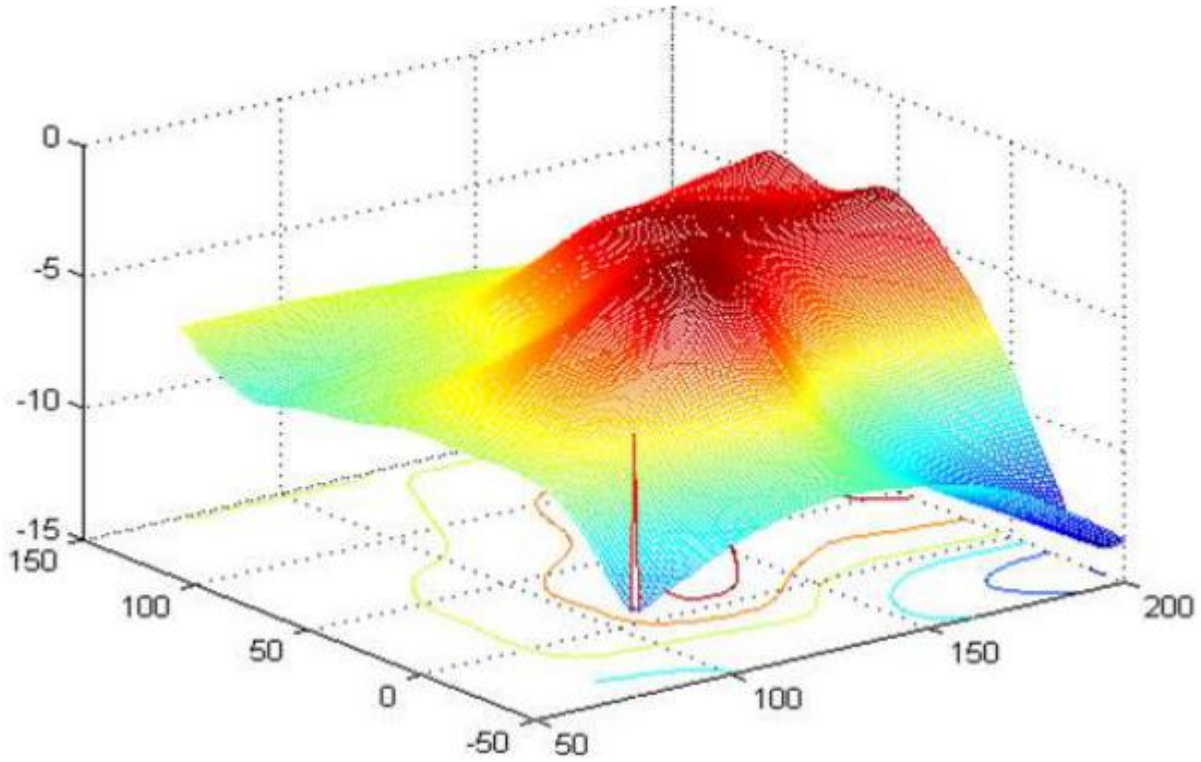
（1）画出散点图 2.15 如下所示:


150
100
50
0
-50
-100
60
80
100
120
140
160
180
200

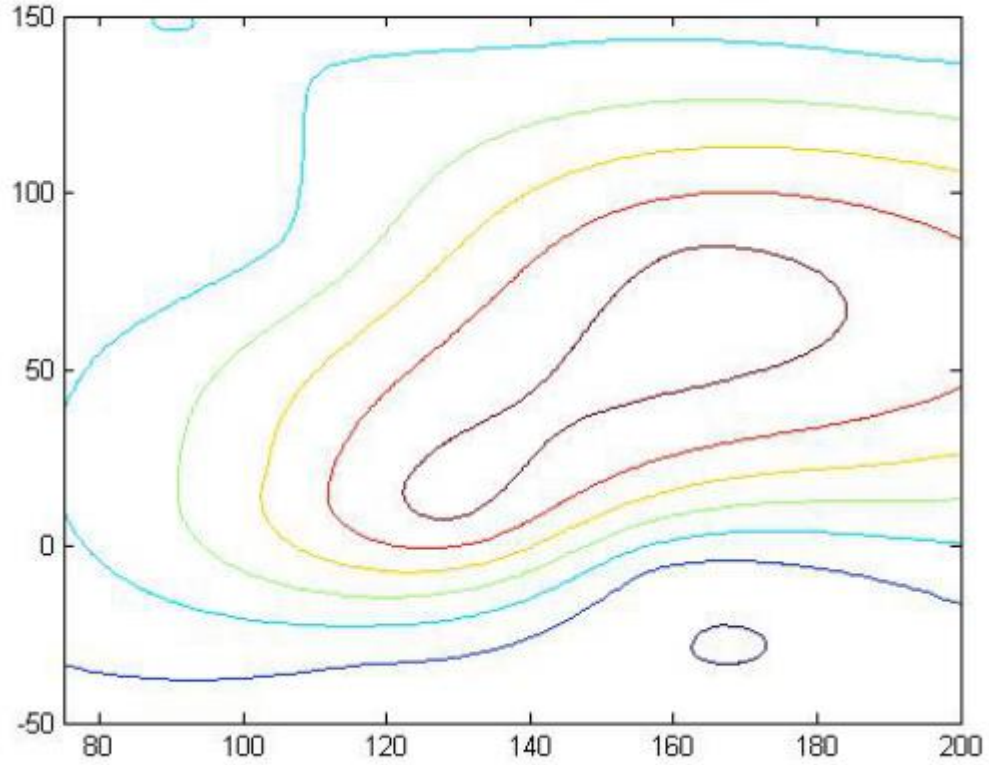
(2) 作出海底地貌图



（3）标识出危险海域



(4) 作出等高线图



案例三 醉汉行进路线（日常生活类）

问题背景：我国是酒的故乡，也是酒文化的发源地，是世界上酿酒最早的国家之一。酒的酿造，在我国已有相当悠久的历史。在中国数千年的文明发展史中，酒与文化的发展基本上是同步进行的。某个醉汉喝醉酒后的行走路线如下所示(已知一组数据点)，试着对该醉汉的行走路线进行分析。假设由下面的语句生成一组数据 x 和 y。

$$x = [1:0.1:5]$$

$$c = \begin{bmatrix} 2.5 & -0.5 & 1.3 & -0.1 \end{bmatrix}$$

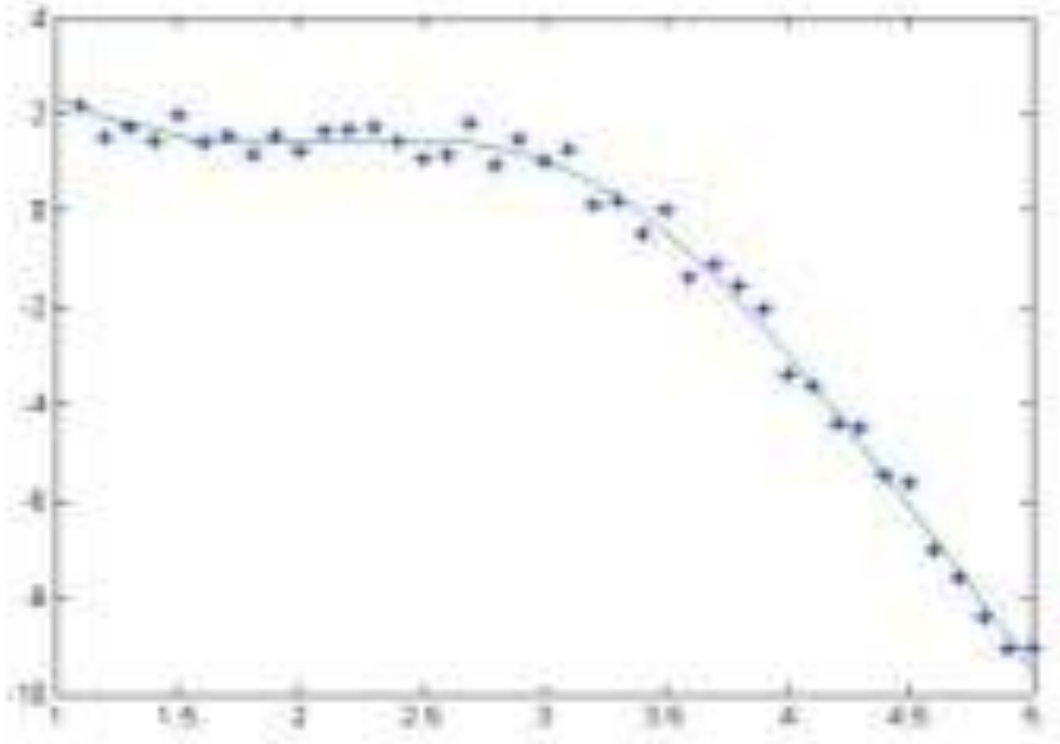
$$y = c(1) + c(2) * \sin(x) + c(3) * \cos(x).^2 + c(4) * x^3 + (\mathrm{rand}(\mathrm{size}(x)) - 0.5)$$

并已知该数据满足原型

$$y = c(1) + c(2) * \sin(x) + c(3) * \cos(x).^2 + c(4) * x^3$$

其中 c_i 为待定系数. 采用最小二乘曲线拟合画出拟合图。

解：采用最小二乘曲线拟合的目的就是获得这些待定系数。由题意定义M函数进行拟合，可以得到拟合图



感谢大家！