现代数学基础丛书 165

散乱数据拟合的模型、方法和理论

（第二版）

吴宗敏 著

科学出版社

现代数学基础丛书 165

散乱数据拟合的模型、方法和理论

（第二版）

吴宗敏 著

科学出版社
北京

内 容 简 介

本书是应用数学与计算数学中有关曲面及多元函数插值、逼近、拟合的入门书籍，从多种物理背景、原理出发，导出相应的散乱数据拟合的数学模型及计算方法，进而逐个进行深入的理论分析.书中介绍了多元散乱数据拟合的一般方法，包括多元散乱数据多项式插值、基于三角剖分的插值方法、Boole 和与 Coons 曲面、Sibson 方法或自然邻近法、Shepard 方法、Kriging 方法、薄板样条方法、MQ 拟插值法、径向基函数方法、运动最小二乘法、隐函数样条方法、R 函数法等. 同时还特别介绍了近年来国际上越来越热并在无网格微分方程数值解方面有诸多应用的径向基函数方法及其相关理论.

本书可供应用数学与计算数学专业的研究生阅读，也可作为水文地质、预测预报、模式识别、统计学习等工程技术领域科技人员的参考用书.

图书在版编目（CIP）数据

散乱数据拟合的模型、方法和理论/吴宗敏著. —2 版. —北京：科学出版社，2016

(现代数学基础丛书)

ISBN 978-7-03-048902-9

I. ①散… II. ①吴… III. ①数据—拟合—计算方法 IV. ①O241.5

中国版本图书馆 CIP 数据核字 (2016) 第 136490 号

责任编辑：李静科　赵彦超／责任校对：彭　涛

责任印制：张　伟／封面设计：陈　敬

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京厚诚则铭印刷科技有限公司 印刷

科学出版社发行　各地新华书店经销

*

2007 年 1 月第 一 版　开本：B5(720×1000)

2016 年 6 月第 二 版　印张：12

2016 年 6 月第一次印刷　字数：227 000

POD定价： 68.00元

(如有印装质量问题，我社负责调换)

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言, 书籍与期刊起着特殊重要的作用. 许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍, 从中汲取营养, 获得教益.

20 世纪 70 年代后期, 我国的数学研究与数学书刊的出版由于文化大革命的浩劫已经破坏与中断了 10 余年, 而在这期间国际上数学研究却在迅猛地发展着. 1978 年以后, 我国青年学子重新获得了学习、钻研与深造的机会. 当时他们的参考书籍大多还是 50 年代甚至更早期的著述. 据此, 科学出版社陆续推出了多套数学丛书, 其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出, 前者出版约 40 卷, 后者则逾 80 卷. 它们质量甚高, 影响颇大, 对我国数学研究、交流与人才培养发挥了显著效用.

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者, 针对一些重要的数学领域与研究方向, 作较系统的介绍. 既注意该领域的基础知识, 又反映其新发展, 力求深入浅出, 简明扼要, 注重创新.

近年来, 数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用, 还形成了一些交叉学科. 我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科的各个领域.

这套丛书得到了许多数学家长期的大力支持, 编辑人员也为其付出了艰辛的劳动. 它获得了广大读者的喜爱. 我们诚挚地希望大家更加关心与支持它的发展, 使它越办越好, 为我国数学研究与教育水平的进一步提高做出贡献.

杨　乐

2003 年 8 月

第二版前言

本书第一版于 2007 年 1 月出版，主要介绍基于各种几何及物理模型的散乱数据拟合方法及背后的数学原理、函数空间分析以及在微分方程数值解中的应用，特别地利用 Hermit-Birkhoff 插值方法进行微分方程数值解，更具备微分方程数值化的本质意义.

第二版主要增加了利用 MQ 拟插值进行高阶的数值求导，主要为 3.8 节与 3.9 节，其中介绍了方法的引入、误差估计及稳定性分析. 理论与数值结果都显示了这是一种比差分更加稳定的数值求导方法，并且可以在低有效位的计算软件或计算机上实现高精度的计算.

吴宗敏

2016 年 6 月

第一版前言

数学, 特别是应用数学的任务有两个: 一个是自然现象或事件的规律性数学描述, 在应用数学上表现为数学模型的建立; 另一个是自然现象或事件的计量性数学描述, 就是要具体地得到上述方程的解或数值解.

前者从一些自然规律出发, 进行数学抽象, 得到某些自然现象之间满足的一些数学关系式, 这些关系式一般是一些代数方程式或者微分方程式, 然后对这些方程式进行变形演化得到另一些方程式, 用来解释自然现象服从某种导出的新的规律, 探索未知的自然规律; 或者从方程的形式出发研究方程解的渐近行为, 从解的存在性和形态性质来解释自然现象的存在性、形态及性质. 后者希望定量地描述自然现象或事件发生的状况, 或者估计某种自然现象发生的可能性大小, 从而比较确切且定量地描述自然现象.

Roussell 曾经说过这样的话: “描述自然的问题最后一般都归结为逼近问题.” 事实上, 在导出数学模型时, 人们必然地抓住那些重要的影响因子, 忽略 (或有意排除) 那些不重要的影响因子. 这是数学模型对自然现象所满足的实际数学关系的逼近 (用数学语言来说这是算子逼近). 在诸如生物、生命科学问题中, 有些过程是不可能用简单的数学方程描述的, 称之为盲模块或黑匣子, 对这样的问题, 只能用纯数学的方法进行算子逼近.

有了数学模型或方程之后, 解方程时, 如果不能得到方程的显式解, 或者显式解的数学表达式非常复杂, 一般 (不得不) 采用数值解方法, 这是利用数值解对方程理论 (真实) 解的逼近.

在众多实际问题中, 精确的数值解往往是很难得到的, 而且也可能没有实际意义. 有时获得的求解方程的信息本身就是不完全的, 这时就希望找到一个容易计算而又能比较合理地表现真实解的方法. 有时可以得到精确解, 但是获得精确解所要花费的时间或费用太大, 这时一般采用具有较高精确度而且更容易获得的数值解来模拟精确解, 这也就是对解函数用简单的函数进行模拟、拟合、逼近的背景. 在所有这类逼近问题中, 最简单的是函数的插值与逼近问题 (数学模型的建立可以看成是基于函数逼近的算子逼近问题, 一个更加抽象空间的逼近问题), 而反过来函数的插值与逼近问题也是这类相关问题 (包括算子逼近) 的基础.

本书主要讨论与多元散乱数据有关的插值和逼近问题, 进而讨论多元散乱数据插值或逼近在微分方程数值解的应用. 本书从问题的来源讲起, 将介绍多种散乱数据插值的模型. 每一种模型的导出都采用这样章节安排, 首先从某种自然现象的物

理规律出发, 找出某种关系, 建立一种数学模型, 然后从这个数学模型导出一种散乱数据插值的计算方法, 最后对这种插值算法的数学性质进行讨论. 对多元散乱数据计算机拟合算法感兴趣的工程应用人员, 可以只阅读算法的导出与计算公式的描述部分 (一般是每章的开始两节), 从而就可以利用这些公式为应用服务. 另一方面, 本书也可以作为研究生教材. 为了让读者能够进入有关问题的研究, 我们在介绍了一种模型及计算方法以后要对这个方法进行深入的理论探讨, 使读者对该方法的理论有比较深刻的了解, 从而可以进行深入的理论研究, 而其余的几节会进一步讨论该方法的数学性质. 在本书的最后部分我们还介绍了函数逼近或插值方法在微分方程数值解中的应用, 特别介绍一些不同于其他微分方程数值解方法书籍中的被称为无网格方法的新方法. 所以本书既可以作为这方面科技应用人员运用这些方法编程的参考书, 也可以作为该方面研究的入门教材.

作者感谢中国科学院科学出版基金、上海市教育委员会科技出版基金的资助.

吴宗敏

2006 年 6 月

目 录

第 1 章　多元散乱数据拟合与多项式插值

§1.1　问题的提出

在实际工作中经常会碰到这样的问题，希望用数学中的函数来描述某一研究对象，譬如用直线来描述光的行进路线. 但是物理学告诉我们：光只是在不受磁场影响的真空中表现为直线，而我们所研究的光都不是在真空中的，而且它还要受到地球、太阳的磁场影响. 所以甚至用直线描述光的行进路线这样一个一般认为是非常准确的基本方法，事实上也是一种逼近. 但是人们一般都能够接受光线是直线这样的概念，因为日常生活中的光线与直线非常接近，其差别对绝大多数的科学计算都是可以忽略的. 严格地说有两条理由：第一，直线非常简单，而要精确描述光的行进路线几乎是不可能的；第二，用直线描述光的行进路线有足够的精确度，而用非直线的复杂函数描述光线缺少实际应用意义. 从这个例子可见，求解逼近问题也有两条指导思想：第一，要选择简单的函数空间作为描述对象函数的逼近空间；第二，这个空间要有很好的逼近度. 如果知道某个研究对象是可以用某一类函数表示描述，甚至已经知道了这个函数的许多测量值，那么如何用数学方法来尽可能精确地描述这个函数呢？或者说，如何寻找尽可能地近似描述研究对象函数 $f(x)$ 的函数 $f^*(x)$ 呢？如果要找的函数还要求在一些测量点就等于这些测量值 (譬如 $f(x_j)=f_j$)，那么数学上就称为插值 (interpolation). 如果要求在某个函数类里寻找在某种度量下与测量值或者研究对象函数本身误差最小的函数 (譬如 $\sum(f^*(x_j)-f(x_j))^2$, $\displaystyle\int(f^*(x)-f(x))^2dx$ 取最小)，那么数学上就称为逼近 (approximation)，而在工程上一般通称为拟合. 显然插值是逼近的一个特殊的形式，并且是逼近的一个基础. 有时还要求有一定的形状限制，或者要求在形状上与研究对象函数有相像性，譬如计算机辅助几何设计中的保凸性问题，这类问题也统称为拟合问题. 本书主要研究对象是插值与逼近问题中与多元散乱数据有关的内容，即函数一般是多变量的，至少是两个变量的. 对一个变量情形或曲线的拟合问题假定读者已经有所了解，亦可参看基础的数值逼近及计算机辅助几何设计课程的参考书[1~7].

本书把重点放在数据是多元散乱的拟合问题上，即测量的数据一般是散乱的，也就是说测量点不一定落在某种网格上，否则可以用曲线的张量积方法 (或者说用曲线来织成曲面的方法) 解决. 进一步地，本书还讨论了测量数据甚至是一些一般线性泛函值的问题. 在本书中，我们通常是从一些物理模型出发，导出一系列的散

乱数据的拟合方法, 从而使得我们讨论的插值逼近拟合方法尽可能建立在实在的物理背景上, 然后用数学手段来分析这些方法的数学性质, 而且把最终目标放在计算机的实现上. 也就是说介绍的方法都是计算机可实现的, 很多还是当前国际流行的一些大型软件中所采用的方法. 下面通过给出一些简单的例子来说明问题的来源及应用背景, 并把这些问题抽象成数学问题.

(1) 来自应用的问题. 在石油勘探中, 经常把地层的地质渗透率作为研究对象, 从而可以判别诸如某地层是否可能蕴藏石油等问题. 也希望不仅采集油井附近的石油, 在三级采油中, 人们利用一些井灌水另一些井抽油的方法把石油赶出来. 要分析地层中石油及水的流动过程就必须研究地层的地质渗透率, 从而决定如何灌水与抽油的方案. 渗透率这个对象可以用一个三元变量的函数来表示. 实际问题中通过打井取芯获取一些井位在某些深度的数据, 要用数学方法描述这个函数. 井位一般来说不是网格型的, 有时由于岩芯的损坏, 某些深度的测量值也有缺损, 所以这是一个散乱数据的插值问题.

天气预报中需要地面气温的等值线图, 这是通过一些地面温度测量点得到的一些散乱的数据来描述温度对于地面坐标的两个变量的函数. 因为测量点一般不是网格型的, 所以这也是一个散乱数据的插值问题. 如果还要给出气温变化的趋势图, 那么我们还要通过譬如风向、地形等使温度变化的信息来描述未来的温度分布, 这是一个更加复杂的散乱数据的拟合问题.

在临摹、仿制及考古的古生物复原问题中, 人们经常要通过临摹、仿制对象的一些测量值来绘制该对象的外形, 从而制模. 由于临摹、仿制对象的形状的复杂性, 所以这也是一个散乱数据的拟合问题.

很多实际问题的数学模型可以归结为一个偏微分方程问题, 譬如传染病传染过程、热传导问题、流体问题、科学控制问题, 甚至股票涨跌、经济增长, 这些问题都可以数量化, 从而可以用函数表示. 这个函数满足某种偏微分方程, 我们还知道有关这个函数的信息, 而将该偏微分方程离散化后, 加上我们已经获得的有关这个函数的一些信息, 得到的是一系列该函数的泛函信息值. 当求解区域是不规则的时候, 实际上这也是一个散乱数据 (非函数值数据而是泛函数据) 的拟合问题. 这时如同温度的趋势图, 不仅数据是散乱的, 而且数据不是普通的函数值而是复杂的泛函值. 在本书最后也将涉及这样的问题.

(2) 数学本身要求解决的问题. 有很多复杂的函数很难用数学公式来描述, 如果还要求对这类函数积分、微分或者求与这类函数有关的微积分方程的解, 那么只能采用数值的方法, 也就是说要用简单的函数来描述、逼近复杂函数, 然后由对简单函数的运算来模拟、逼近对复杂函数的运算. 计算机虽然有很强的运算能力, 但是它只能处理离散的问题, 而且它处理的只能是有限位的二进制小数. 计算机不能精确表示三角函数或指数函数. 计算机的发展也对我们提出了这样的研究课题：如

何用简单的离散的函数表示复杂的连续函数, 这实际上都是散乱数据的拟合问题.

现在我们可以将上述问题数学抽象化了. 拟合问题的基础是插值问题.

问题 1.1.1 给定 $x_j \in \mathbb{R}^d$, $f_j \in \mathbb{R}$, $j = 1, \cdots, n$, 寻找函数 $f(x)$, 满足 $f(x_j) = f_j$, $j = 1, \cdots, n$.

进一步地, 我们可以讨论更一般的问题.

问题 1.1.2 λ_j 是 $C(\mathbb{R}^d)$ 上的线性泛函, 给定数据 $f_j \in \mathbb{R}$, $j = 1, \cdots, n$, 希望寻找 $C(\mathbb{R}^d)$ 中的函数 $f(x)$, 满足 $\lambda_j f(x) = f_j$, $j = 1, \cdots, n$.

上面的问题太庞大, 太复杂了. 有很多这样的函数 $f^*(x)$ 满足我们对要寻找函数的要求. 显然研究对象函数 $f(x)$ 本身就满足要求, 可惜这个函数往往找不到或者很难找到. 当然我们希望找到的函数与研究对象函数有很好的相似性质, 这包括误差小这个逼近性质以及形状相似这个几何性质. 另一方面, 如果找到的函数不能很好地或很快地用计算机数值描述, 也偏离了我们的目标. 所以在讨论上述问题时, 我们还要分析

(1) 在哪个函数类里寻找函数 $f^*(x)$. 我们要求这个函数类有计算机表示简单的优点, 而又能具备有效地描述研究对象函数的能力;

(2) 在函数类决定以后, 我们还要分析解的存在性、唯一性问题, 以及寻找求解的算法和讨论算法稳定性问题;

(3) 如果测量的数据越来越多, 越来越密, 那么找到的函数的渐近性质是怎么样的呢? 人们希望得到的函数越来越逼近研究对象函数, 并且有较快的收敛速度;

(4) 理论算法与计算机实现有时也会有非常大的差距, 所以我们还要讨论函数 $f^*(x)$ 有效的计算机计算与表示方法.

为了能更好地讨论多元散乱数据的拟合问题. 先引进一些多元记号:

向量及点由数组 $x^{\mathrm{T}} = (x_1, \cdots, x_d) \in \mathbb{R}^d$ 定义, 在这个 d 维空间上装备有欧几里得范数 $\|x\| = \sqrt{\sum_{j=1}^d x_j^2}$, 从而是一个距离空间.

多元记号 $\alpha = (\alpha_1, \cdots, \alpha_d) \in I^d$ 是 d 维空间的正整数点, $\alpha_j \geqslant 0$, $j = 1, \cdots, d$, 定义整点的模 $|\alpha| = \sum_{j=1}^d \alpha_j$.

定义 $x^\alpha = \prod_{j=1}^d x_j^{\alpha_j}$, $xy = \sum_{j=1}^d x_j y_j$.

梯度微分算子与偏导数记为

$$\nabla = D = \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_d}\right)^{\mathrm{T}},$$

$$D^\alpha f(x) = \partial^\alpha f(x) = \frac{\partial^{|\alpha|} f(x)}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_d} x_d}.$$

方向导数由

$$\frac{\partial f(x)}{\partial y} = \lim_{t \to 0} \frac{f(x+ty) - f(x)}{t}$$

定义. 用 $\Omega \subset \mathbb{R}^d$ 表示我们的讨论区域, 这一般是一个单连通的由简单连续闭曲线作为边界的区域. 上面 x_j 的下标 j 表示向量 x 的第 j 个分量. 在不引起误解的情况下也把一列点记成 $\{x_j\}$, 这时 j 表示点在点列中的编号. 当我们讨论的测量数据是函数在一些点上的取值时, $\{x_j\} \subset \Omega$ 是数据的测量点, 从而 $\{x_j, f_j\}$ 是采集的数据. 我们用

$$h = \max_{x\in\Omega} \min_{j} \{\|x - x_j\|\}$$

来描述数据点的稠密化程度, 称为数据的密度, 请注意这时数据越密 h 越小.

§1.2 插值问题的 Haar 条件

在单变量的插值问题中, 根据代数基本定理: "任给 $n+1$ 个两两不同的点 $x_j \in \mathbb{R}$ 及在 x_j 上的测量数据 f_j, 存在唯一的一个 n 次多项式, 满足 $f(x_j) = f_j$. " 从而导出了我们熟知的 Lagrange 插值方法. 当然, 在这里也可以用指数多项式代替代数多项式. 如果数据点都落在 $[0, 2\pi)$ 中, 我们还可以用三角多项式构造插值. 一般地, 如果 $b_j(x) \quad (j = 1, \cdots, n)$ 是一列函数 (基函数), 对任何两两不同的点的集合 $\{x_j\}_{j=1}^n$, 矩阵 $(b_j(x_k))$ 都是非奇异的, 那么就可以通过解一个线性方程组

$$\sum a_j b_j(x_k) = f_k,$$

得到一个形式为

$$f^*(x) = \sum_{j=1}^n a_j b_j(x)$$

的插值函数, 满足 $f^*(x_j) = f_j$. Euler 可能最先研究了这样的插值方法, 我们把这样的插值方法统称为 Euler 插值方法, 对任何两两不同的数据点都存在 Euler 插值的函数系称为 Tschebyscheff 系.

为了解决多元散乱数据的插值问题, 根据以往曲线拟合的经验, 我们首先想到的是 Euler 的插值方法, 即选择一组函数 $b_j(x)$ (最好是多项式, 但也可以是其他的函数, 称为基函数), 令

$$f^*(x) = \sum_{j=1}^n a_j b_j(x),$$

然后通过待定系数方程

$$f^*(x_k) = \sum_{j=1}^n a_j b_j(x_k) = f_k, \quad k = 1, \cdots, n$$

来决定系数 a_j. 用这样的方法, 要使线性方程组对任意选定的两两不同的测量点都有解, 则要求其系数矩阵的行列式对任何两两不同的测量点都不为零. 这导出了

我们必须要考虑的对任何的两两不同的数据点插值问题解的存在唯一性条件 —— Haar 条件.

定义 1.2.1 称函数族 $\{b_j(x)\} \in \mathbb{C}(\mathbb{R}^d)$, $j = 1, \cdots, n$ 满足 Haar 条件, 如果对任何两两不同的点 $x_j \in \mathbb{R}^d$, $j = 1, \cdots, n$, 矩阵 $(b_j(x_k))$ 都是非奇异的.

固定维数的空间表现函数的能力总是有限的. 如果我们要利用某个函数空间逼近任何的函数, 或者希望对任何多个数据点插值问题都有解, 那么就需要一个函数系.

定义 1.2.2 函数系 $\{b_j(x)\} \in \mathbb{C}(\mathbb{R}^d)$, $j = 1, \cdots, \infty$ 对任何 n 都满足 Haar 条件, 那么我们称这个函数系是一个 Tschebyscheff 系.

容易证明, 只有当函数系满足这样的条件时, 对任何位置、任何个数的数据点 x_j, 及测量值 f_j, Euler 方程有解且解是唯一的, 否则可能对某些测量值方程无解, 而对另一些数据可能是多解的. 即函数系 $\{b_j(x)\} \in \mathbb{C}(\mathbb{R}^d)$ 对任何的两两不同的数据点 Euler 方法插值存在唯一的充分必要条件是: 函数系 $\{b_j(x)\} \in \mathbb{C}(\mathbb{R}^d)$ 是一个 Tschebyscheff 系. 在一元曲线情形我们知道许多这样的函数族系, 譬如单项式 $x^j, j = 0, 1, \cdots, \infty$, 指数函数 $e^{jx}, j = 0, 1, \cdots, \infty$, 三角多项式 $e^{ijx}, j = -\infty, \cdots, 0, \cdots, \infty$; $x \in [0, 2\pi)$. 这样就可以导出许多类型的 Euler 插值公式. 可是在多变量情形我们却有下述定理.

定理 1.2.1 在多变量情形, 任何给定的函数族都不满足 Haar 条件. 当然更不可能存在 Tschebyscheff 系.

证明 用反证法. 如果有函数族 $\{b_j(x)\} \in \mathbb{C}(\mathbb{R}^d)$, $j = 1, \cdots, n$, 对某点列 $x_j \in \mathbb{R}^d$, $j = 1, \cdots, n$, 矩阵 $(b_j(x_k))$ 是非奇异的. 那么在 $\mathbb{R}^d$, $d \geqslant 2$, 可以找到两条连续的曲线连接 x_1 和 x_2, 使得这两条曲线除了起始点 x_1 和终点 x_2 外不相交, 并且也不经过其他的数据点 $x_j (j > 2)$. 让 x_1 沿着一条曲线走向 x_2, 而让 x_2 同时沿着另一条曲线走向 x_1, 那么在移动过程中 n 个点始终是两两不同的. 由 Haar 条件, 矩阵 $(b_j(x_k))$ 在点列移动过程中应该始终是非奇异的. 运动的结果是交换了系数矩阵的前两行, 也就是说开始与结束这两个矩阵的行列式有相同的绝对值而符号相反. 由于函数及曲线都是连续的, 从而行列式关于点 x_1, x_2 的移动是连续的. 即行列式是关于点 x_1, x_2 移动的连续函数. 由连续函数零点存在定理, 在点的移动过程中必定存在一个中间状况的 n 个点 x_j, 矩阵 $(b_j(x_k))$ 的行列式为零. 这与函数组满足 Haar 条件矛盾. □

上述定理表明多于两个变量的散乱数据插值问题与单变量插值问题有本质上的不同. 在一元情形得到的各种插值方法一般都不能直接推广到多元情形, 这也就是为什么要把多元散乱数据的拟合作为一个独立的课题研究的原因.

本书主要分成两个部分来解决这个问题. 一类是局部的方法, 主要是分片拼接法, 即首先把研究的区域分片, 分片地解决插值问题, 然后把这些曲面片 (超曲面片)

拼接起来, 由于高次连续拼接条件的复杂性, 这个方法较常用来处理只要求低次连续的问题. 因为实际问题通常也只有低次连续的要求, 所以这些方法也是经常地被应用于实际问题 (譬如微分方程数值解的有限元方法, 计算机辅助几何设计的样条方法); 另一类方法是根据物理模型来导出数据之间的关系, 从而根据测量点 x_j 来决定基函数 $b_j(x)$, 这种方法可以达到很高的连续要求和逼近要求, 也有计算公式简单的优点, 并且有明显的物理背景, 特别是近年来径向基函数插值的研究, 使得这个方法在理论分析与实际应用中都取得了很大的突破, 引起了理论研究工作者与实际应用人员的广泛的关注与兴趣. 本书还准备介绍其他的一些散乱数据的插值方法, 使得读者对散乱数据插值有一个整体的概念并掌握各类基本方法.

§1.3　多元散乱数据的多项式插值

从这一节开始我们介绍散乱数据的插值方法. 因为单变量情形的多项式空间满足 Haar 条件, 所以多项式的 Lagrange 插值是可以实现的, 反观多变量情形, 多项式函数空间已经不再满足 Haar 条件, 所以 Lagrange 插值不是一定存在的. 多变量函数中我们最熟悉的函数类还是多项式, 多项式函数还有计算机表示简单的优点, 这一点对当今计算机的发展及计算机的应用特别重要, 而且多项式插值还可以作为分片拼接法的基础, 所以这一节我们还是要来讨论用多项式对多变量函数散乱数据插值的可能性. 为了描述方便, 我们有时只讨论两个变量的情形, 而一般的多变量情形是可以相似解决的.

定理 1.3.1　任给两两不同的测量点 $\{(x_j, y_j)\}_{j=0}^n \in \mathbb{R}^2$ 及在测量点相应的函数测量值 $\{f_j\}_{j=0}^n$, 至少存在一个关于各分量次数都不超过 n 的多项式

$$p(x,y) = \sum_{j=0}^{n}\sum_{k=0}^{n} a_{jk}x^j y^k,$$

满足 $p(x_j, y_j) = f_j$.

此时函数空间是 $(n+1)^2$ 维的, 而约束条件只有 $n+1$ 个, 所以解的唯一性就不能保证了.

证明　存在一个 $(n+1)^2$ 的网格, 使得点集 $\{(x_j, y_j)\}_{j=0}^n \in \mathbb{R}^2$ 是该网格上的节点. 在这个网格的所有节点上, 有些是没有测量值的, 而有些已经有了测量数据. 任意给定无测量值点上的值, 就可以用张量积的方法, 由单变量的多项式插值公式得到问题的解. 因为在至少 $(n+1)^2-(n+1)$ 点上的函数值可以任意给定, 所以我们得到的解有 $(n+1)^2-(n+1)$ 个自由度. □

我们还可以把这个定理推广到一般的多变量情形, 并且用一元的 Lagrange 公式给出另外一个构造性的证明.

定理 1.3.2 任给两两不同的测量点 $\{x_j\}_{j=0}^n \in \mathbb{R}^d$ 及在测量点相应的函数测量值 $\{f_j\}_{j=0}^n$, 那么至少存在一个次数和不超过 n 的 d 元多项式 $p(x)$, 满足 $p(x_j) = f_j$.

证明 任取向量 r_{jk}, 满足 $\langle r_{jk}, x_j - x_k\rangle \neq 0$, 利用单变量的 Lagrange 的插值公式构造插值公式如下:

$$p(x) = \sum_{j=0}^{n} f_j \prod_{k\neq j} \frac{\langle x - x_k, r_{jk}\rangle}{\langle x_j - x_k, r_{jk}\rangle}.$$

显然这是一个次数不超过 n 的多项式. 把数据点代入上式, 容易验证, 这是关于测量数据的插值. 向量 r_{jk} 可以特别地取为 $r_{jk} = x_j - x_k$, 这时

$$p(x) = \sum_{j=0}^{n} f_j \prod_{k\neq j} \frac{\langle x - x_k, x_j - x_k\rangle}{\langle x_j - x_k, x_j - x_k\rangle}$$

是一个各项次数和不超过 n 的插值多项式. □

注 1.3.1 前一个定理的次数是指各个分量的次数中的最大次数, 而后一个定理的次数是指各个分量的次数和. 此时这个函数空间的维数是 $\binom{n+d}{d}$, 一般地要小于前一个定理函数空间的维数. 从某种意义上说, 后一方法要优于前一方法, 但是后一方法函数空间的维数仍然大于约束条件的个数, 所以唯一性还是不能保证. 上述证明同时给出了计算机赋值的计算公式.

正如一元的多项式插值有 Lunge 现象一样, 多元情形的多项式插值也有类似现象, 所以多项式插值一般只用来作为局部的公式. 譬如, 人们可以在每个格子点的邻近利用多项式插值得到在这个格子点的函数估计值, 然后利用张量积方法 (用曲线织成曲面的方法) 构造函数的逼近, 这时人们虽然得到的不是一个插值函数, 但一般还是一个好的逼近函数. 在计算机快速发展的时代, 一个算法的计算机可实现性、快速实现性、计算稳定性、可靠性是十分重要的, 而多项式插值在多元问题上作为局部方法也如一元情形是非常有用的, 所以我们还是要讨论更快捷稳定的算法.

1. 多项式插值的 Aitken 公式

如果 $p_{j,k}(x)$ 是关于点 $x_j, \cdots, x_k$ 及数据 $f_j, \cdots, f_k$ 的多项式插值函数, 那么由递推公式

$$p_{j,j}(x) = f_j,$$

$$p_{j,k+1}(x) = p_{j+1,k+1} \frac{\langle x - x_j, r_{k+1,j}\rangle}{\langle x_{k+1} - x_j, r_{k+1,j}\rangle} + p_{j,k} \frac{\langle x_{k+1} - x, r_{k+1,j}\rangle}{\langle x_{k+1} - x_j, r_{k+1,j}\rangle}$$

得到的 $p_{j,k+1}(x)$ 是关于点列 $x_j,\cdots,x_{k+1}$ 及数据 $f_j,\cdots,f_{k+1}$ 的次数不超过 $k+1-j$ 的多项式插值函数, 从而 $p(x)=p_{0,n}(x)$ 是一个次数不超过 n 的插值多项式. 同样地我们也可以导出 Newton 公式.

2. 多项式插值的 Newton 公式

由递推公式定义多变量函数关于一些方向 r_{jk} 的差商

$$\Delta^j(x_0,\cdots,x_j)f=\frac{\Delta^{j-1}(x_1,\cdots,x_j)f-\Delta^{j-1}(x_0,\cdots,x_{j-1})f}{\langle x_j-x_0,r_{j0}\rangle}.$$

那么 Newton 插值多项式可以表示为

$$p(x)=\sum_{j=0}^{n}\Delta^j(x_0,\cdots,x_j)f\cdot\prod_{k=0}^{j-1}\langle x-x_k,r_{jk}\rangle.$$

也如同单变量 Newton 公式一样, 我们可以获得误差估计式

$$f(x)-p(x)=\Delta^{n+1}(x_0,\cdots,x_n,x)f\cdot\prod_{k=0}^{n}\langle x-x_k,r_j\rangle,$$

其中 r_j 是任意固定 x 以后, 满足 $\langle x-x_j,r_j\rangle\neq 0$ 的向量. 读者可以利用一元情形的公式以及多元情形的不同点, 对这两个公式进行更细致的讨论与验证.

多项式插值公式虽然解决了散乱数据插值的存在性问题, 但是正如单变量多项式插值一样, 它不是一个稳定的方法, 在多变量情形甚至还会出现一些更为奇怪的结果, 所以在实际应用中, 人们并不完全采用这个方法. 但这些方法可以作为其他算法 (局部法或二步法) 的基础. 也正是这个原因, 我们把这个插值方法放在这一章中, 而不把它放在可以实际应用的方法章节中. 譬如讲我们可以采用多项式插值公式作为局部的解, 然后再用二步法构造整体解, 或者在一个局部用多项式插值估计函数在插值点邻近的高阶 (导数等) 行为, 为其他利用函数高阶信息的插值方法服务. 下面的几个章节将介绍更加实用有效的方法.