



数据拟合与不确定度

——加权最小二乘拟合及其推广

(第2版)

[德] Tilo Strutz 著

王鼎 唐涛 尹洁昕 杨宾 吴志东 译

吴瑛 审校

Data Fitting and Uncertainty:

A Practical Introduction to

Weighted Least Squares and Beyond(2nd Edition)



国防工业出版社
National Defense Industry Press



装备科技译著出版基金

数据拟合与不确定度

——加权最小二乘拟合及其推广（第2版）

**Data Fitting and Uncertainty: A Practical Introduction to
Weighted Least Squares and Beyond (2nd Edition)**

[德] Tilo Strutz 著

王鼎 唐涛 尹洁昕 杨宾 吴志东 译
吴瑛 审校

国防工业出版社

·北京·

著作权合同登记 图字：军-2018-071 号

图书在版编目（CIP）数据

数据拟合与不确定度：加权最小二乘拟合及其推广：第2版/（德）蒂罗·斯特鲁兹（Tilo Strutz）著；王鼎等译. —北京：国防工业出版社，2019.3

书名原文：Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond（2nd Edition）

ISBN 978-7-118-11834-6

I. ①数… II. ①蒂… ②王… III. ①最小二乘法—研究
IV. ①O241.5

中国版本图书馆 CIP 数据核字（2019）第 039718 号

First Published in English under the title

Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond(2nd Edition)

By Tilo Strutz

Copyright ©Springer Fachmedien Wiesbaden,2016

This edition has been translated and published under licence from
Springer Fachmedien Wiesbaden GmbH, part of Springer Nature.

※

国防工业出版社出版发行

（北京市海淀区紫竹院南路 23 号 邮政编码 100048）

天津嘉恒印务有限公司印刷

新华书店经售

*

开本 710×1000 1/16 印张 16¼ 彩插 6 字数 307 千字

2019 年 3 月第 1 版第 1 次印刷 印数 1—2000 册 定价 119.00 元

（本书如有印装错误，我社负责调换）

国防书店：(010) 88540777

发行邮购：(010) 88540776

发行传真：(010) 88540755

发行业务：(010) 88540717

译者序

众所周知，最小二乘方法已成功应用于各个工程科学领域，其重要性不言而喻。由 Tilo Strutz 教授所著的 *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond* 一书以数据拟合为视角，介绍了与最小二乘估计紧密相关的各种方法，阐述了参数估计的不确定度问题。此书内容自成体系，思想新颖，特点鲜明，理论与实际紧密结合，并提供了丰富的案例和软件代码。因此，这是一部在数据拟合和最小二乘估计领域中的优秀著作。译者团队于 2016 年完成了对此书（第 1 版）的翻译工作。

值得庆幸的是，Tilo Strutz 教授于 2015 年对此书又做了进一步完善，补充了很多新内容，并在 Springer 出版社出版了此书的第 2 版。鉴于 Tilo Strutz 教授在第 2 版中新增了很多重要知识点，并且改进了第 1 版中的部分内容，故译者团队决定继续对第 2 版进行翻译，以期能有更多中国学者、科研人员以及工程技术人员从此书中受益。

第 2 版新增的知识点主要包括：维纳滤波和最小均方（LMS）滤波、用于异常值检测的 M -score 方法和随机抽样一致性（RANSAC）方法、相关数据的加权最小二乘拟合问题、基于反向传播的神经网络参数训练算法等。此外，第 2 版还通过实验给出了上述新增方法的性能以及 Levenberg - Marquardt 优化方法的收敛性能，并且对与本书配套的软件进行了更新。

本书第 2 版的翻译工作由战略支援部队信息工程大学王鼎、唐涛、尹洁昕、杨宾、吴志东共同完成，并最终由王鼎统稿，吴瑛审校。在整个翻译过程中，除了补充原著新增和改进的内容以外，还对第 1 版中的语句做了进一步梳理，以期使其更符合中文的语言习惯，从而便于读者理解。本书既可以作为高等院校应用数学、物理学、经济学、工程学等专业高年级本科生的教材，也可以作为从事工程技术领域的科研工作者自学或者研究的参考书。本译著的出版得到了装备科技译著出版基金、战略支援部队信息工程大学优秀青年基金（项目编号：2016603201）和“2110”工程（项目编号：102063）的资助，在此一并感谢。

虽然经过了反复的校对和修改,但由于译者的水平有限,本书难免存在不足之处,恳请读者批评指正。

译者

2018年5月于战略支援部队信息工程大学

第 1 版译者序

本书原作者 Tilo Strutz 是德国莱比锡电信应用科技大学的教授, Tilo Strutz 教授学术造诣深厚, 在一些国际著名期刊上发表了很多高水平的学术论文, 并且出版了一些优秀教材和专著, 其研究方向涉及信号处理、图像处理以及数据压缩等诸多领域。

众所周知, 数据拟合问题出现在多个工程技术和应用科学领域, 而最小二乘方法则在求解各类数据拟合问题中发挥着基础性作用。本书正是一本专门讲述数据拟合和最小二乘求解方法的教材, 书中的内容并不侧重于对复杂数学理论的分析 and 推导, 而是侧重于方法的应用和性能的评估, 书中讨论了大量朴素但却不失代表性的应用实例, 并且阐述了一些新颖的思想和理念, 相信无论是哪类专业和层次的读者都能从中获益。此外, 作者还发布了与本教材相配套的 C 语言程序, 可供实践型读者参考使用。

本书的内容主要包括:

第 1 章: 数据拟合问题的引入。该章描述了与数据拟合相关的基本概念, 定义了书中的一些常用变量, 并讨论了线性和非线性拟合问题之间的本质区别, 阐述了如何将非线性模型函数转化为线性模型函数, 并列举了若干线性和非线性拟合问题的应用实例。

第 2 章: 利用最小二乘方法求解模型参数。该章描述了求解数据拟合问题的最小二乘优化模型, 介绍了误差曲面的基本概念, 给出了求解最小二乘问题的一般性方法以及针对线性拟合问题的简化方法, 并讨论了在未知模型函数条件下的数据拟合问题, 最后针对若干具体的应用实例分别给出其求解方法。

第 3 章: 权值和异常值。该章讨论了加权带来的好处, 描述了异常值的基本概念及其产生的机理, 介绍了两种权值估计方法 (分段估计权值和基于偏差估计权值), 给出了两种异常值检测方法 (标准残差法和聚类检测法), 最后针对若干应用实例讨论加权和异常值检测所带来的改进效果。

第 4 章: 拟合结果的不确定度。该章描述了拟合优度、精确度以及准确度的相关概念, 介绍了参数估计值的不确定度和模型预测不确定度的计算公式, 并阐述了如何通过观察拟合图形来检验拟合效果, 最后针对一些应用实例讨论

了拟合结果的评估和模型失配的检测。

第5章：矩阵代数。该章描述了与矩阵相关的基本概念，介绍了矩阵行列式的计算方法，并给出了4种矩阵求逆方法（包括伴随矩阵法、Gauss - Jordan消元法、LU分解法，以及奇异值分解法）。

第6章：最小二乘方法背后的理念。该章描述了正态分布的基本概念，讨论了最小二乘方法与最大似然原理之间的关系，给出了两种求解线性问题的计算方法（标准方法和奇异值分解法），并阐述了条件缩放的原理及其所带来的好处，最后介绍了3种求解非线性问题的数值迭代方法（包括Gauss - Newton方法、梯度下降方法以及Levenberg - Marquardt方法），并通过一个计算实例比较了3种方法的特点。

第7章：补充工具和方法。该章给出了3种特殊的参数估计方法（递推自适应参数估计方法、迭代梯度下降方法，以及进化方法），介绍了异常值检测的Chauvenet准则，并描述了误差传播原理，给出了求解线性问题的手工推演方法，阐述了不同模型函数的联合处理方法，最后给出了总体最小二乘拟合方法。

附录A：两种异常值检测方法的比较。该章基于不同分布的测试数据集，利用蒙特卡罗仿真比较了标准残差法和聚类检测法的异常值检测性能。

附录B：软件实现。该章描述了本书附带的C语言源代码的功能、使用说明、组织结构、程序优化等，并针对一些测试数据集验证了该软件的性能。

本书的翻译工作由解放军信息工程大学王鼎、唐涛、吴志东、杨宾、吴江共同完成，并由王鼎统稿，吴瑛完成审校。另外，博士研究生尹洁昕同学对于本书图形翻译和公式编校做了大量工作。本书既可以作为高等院校应用数学、物理学、经济学、工程学等专业高年级本科生的学习教材，也可作为从事工程技术领域的科研工作者自学或者研究的参考书。本译著的出版得到了信息工程大学优秀青年基金（项目编号：2016603201），中国博士后科学基金（项目编号：2016M592989）和“2110”工程（项目编号：102063）的资助。

虽然经过了反复的校对和修改，但由于译者的水平有限，难免存在不足之处，恳请读者批评指正。

译者

2016年10月于解放军信息工程大学

前 言

基于最小二乘方法的优化技术一直都具有广泛应用，并且具有前沿性，这激励我更新本书第 1 版内容，并进而撰写第 2 版。与第 1 版相比，第 2 版不仅做了很多改进，还新增了一些内容。

第 3 章增加了两种鲁棒的异常值检测方法。第一种是对 z -score 方法进行了修正，修正方法是利用观测量的中值和绝对偏差的中值分别替换观测量的均值和标准差。第二种是随机抽样一致性 (RANSAC) 方法，该方法可在异常值所占比例较高的情况下给出最合理的模型参数集合。这两种方法都已经通过编程实现，并且利用软件对它们的性能进行了测试。软件代码与本书一起发布，相比于第 1 版软件，第 2 版软件允许在输入文件中加入权值。

6.4.5 节扩展了两个例子，用于说明 Levenberg - Marquardt 方法求解非线性优化问题的性能。

第 7 章新增了一节 (7.6 节)，用于阐述如何对相关数据进行拟合。如果观测量中的误差具有相同符号 (例如，误差由相同的机制所产生)，那么数据间就会存在相关性。此外，第 2 版对 7.1 节进行了修改，并且讨论如何利用反向传播原理对神经网络中的参数进行训练。

除此以外，第 2 版还新增了维纳滤波和最小均方 (LMS) 滤波，并且提供了很多例子，这两种滤波器与最小二乘方法具有很强的相关性。

感谢 Rainer Feistel 博士针对不确定度问题提供的宝贵建议。

Tilo Strutz

2015 年 7 月于莱比锡电信应用科技大学

第 1 版前言

最小二乘方法是由卡尔 - 弗里德里希 · 高斯于 1795 年提出, 并于 1809 年在其关于天体力学的著作《天体运动论 (第 2 卷)》中正式发表。尽管最小二乘方法是一门相当古老的技术, 但现如今仍然广泛应用于实际中, 其重要性未受到任何影响, 并且在很多方面还有着进一步发展。

最小二乘方法的基本思想是在给定有限个采样值 (即若干 “输入 - 输出” 对) 的条件下, 确定系统参数。如果已知该系统的模型函数, 就以误差平方和最小化为准则来寻求模型函数中的参数, 从而能够精确刻画该系统。对于一些简单情况, 最小二乘问题将简化为寻求一条通过数据散点图的近似曲线。

最小二乘方法早期应用于天文学 (确定天体运动模型) 和测地学 (消除测量误差)。现如今, 这种对离散数据点进行拟合的方法可应用于很多科学领域, 尤其是那些需要处理统计数据的学科, 例如物理学、生物学、经济学、心理学等。除此以外, 它还可用于一些最新发展起来的科学领域, 例如计算机视觉。当利用最小二乘方法进行数据分析时, 在实验数据和由计算机模拟所产生的数据之间基本是没有差异的。

最小二乘方法的原理十分简单, 其计算复杂性主要取决于系统模型函数的特征和复杂度。因此, 自从 19 世纪引入数据拟合方法以来, 已经发展出很多有效且简练的数值技术。除此以外, 还衍生出一些评估拟合结果的数学工具。

本书以便于理解的方式引入最小二乘这一知识体系, 以帮助读者解决具体的数据拟合问题。书中描述了最小二乘方法, 并且按步骤详细阐述了如何处理不确定度。本书的重点在于阐述统一的、普遍的方法, 从中可以得到每个具体问题的解。需要指出的是, 有些书仅描述了非常简单的数据拟合技术, 还有些书则从非常高的数学层次来解释这一问题, 本书旨在弥补它们之间的空缺。本书讨论了很多现实的和仿真的例子, 还配有 C 语言编写的源代码, 并发布在出版社网站上 (见 “OnlinePlus” 区域)。

本书主要面向工程师、软件程序员以及相关专业的本科生, 他们的主要兴趣可能并不在于公式推导和数学证明, 而是想快速熟知本领域的相关知识以及其中需要警惕的问题, 从而能够解决某些具体的数据拟合问题。熟悉矩阵、

(偏)导数以及关于方差和标准差的知识对于阅读本书来说是十分有益的。对于不熟悉最小二乘方法的读者来说,建议从第1章开始,并按照各章的顺序来阅读。每一章的最后都给出了一些测试题,可以检测读者是否理解了各章的内容。

本书的内容由两部分构成。

第一部分介绍了利用最小二乘方法进行数据拟合的基础知识,其中详细讨论了一些线性和非线性系统的例子,以使读者能够解决类似的问题。

第1章描述了数据拟合的基本思想,并且定义了一些术语和符号。第2章给出了应用最小二乘方法的全部重要公式。第3章提出了权值估计方法,这对于观测量的权值不能事先获知,以及数据集中含有异常值的情况是必要的,除此以外,还讨论了一种聚类方法,该方法能够将可能存在的异常值和大部分“好”数据点区分开。第4章则描述如何对数据拟合的结果进行评估。

本书的第二部分主要描述了基本理论,并且更详细地讨论了一些数值方法。从第5章开始,介绍了基本的线性代数与矩阵求逆算法。第6章描述了最小二乘方法的基本思想,并引出最大似然原理,在此基础上,还讨论了一些求解线性和非线性拟合(优化)问题的数值技术。第7章则整理了一些对于数据拟合有用的补充方法和工具。

附录中共包含两部分。附录A研究了基于聚类的异常值检测方法。附录B描述了数据拟合软件,并讨论了最小二乘方法的实现问题,该章的结尾还详细测试了该软件的性能。

Tilo Strutz

2010年7月于莱比锡电信应用科技大学

目 录

第 I 部分 最小二乘方法的框架

第 1 章 数据拟合问题的引入	3
1.1 什么是数据拟合	3
1.2 符号说明	4
1.3 线性与非线性问题	5
1.4 线性数据拟合的应用实例	8
1.4.1 估计常数	8
1.4.2 估计直线中的参数 (线性回归)	9
1.4.3 多项式函数	9
1.4.4 多元线性回归	10
1.4.5 维纳滤波	16
1.5 若干非线性数据拟合问题	16
1.5.1 指数函数	16
1.5.2 复合高斯贝尔函数	17
1.5.3 圆周函数	18
1.5.4 神经网络	19
1.6 测试题	21
第 2 章 基于最小二乘方法估计模型参数	22
2.1 “最小二乘”的含义	22
2.2 求解最小化问题的常规算法	24
2.3 需要注意的问题	26
2.4 线性模型函数条件下的简化处理	27
2.5 未知模型函数条件下的曲线拟合	28

2.6 计算实例	30
2.6.1 常数拟合	30
2.6.2 直线拟合	31
2.6.3 多项式函数拟合	32
2.6.4 平面拟合	32
2.6.5 线性预测	33
2.6.6 余弦函数拟合	33
2.6.7 坐标旋转和平移	34
2.6.8 指数函数拟合	35
2.6.9 复合高斯贝尔函数拟合	36
2.6.10 圆周拟合	37
2.6.11 神经网络	38
2.7 测试题	40
第3章 加权和异常值	41
3.1 加权的好处是什么?	41
3.2 异常值	42
3.3 权值估计	43
3.3.1 分段估计权值	44
3.3.2 利用偏差估计权值	45
3.4 异常值检测方法	48
3.4.1 标准残差法	48
3.4.2 聚类检测法	51
3.4.3 随机抽样一致性 (RANSAC) 方法	59
3.5 加权数据拟合与异常值检测的应用实例	64
3.5.1 常数拟合	65
3.5.2 直线拟合	66
3.5.3 平面拟合	72
3.5.4 坐标变换	74
3.5.5 线性预测	77
3.5.6 余弦函数拟合	78
3.5.7 指数函数拟合	83
3.5.8 复合高斯贝尔函数拟合	86
3.5.9 圆周拟合	90

3.5.10 比较分段估计权值与基于偏差估计权值的性能	91
3.6 结论	93
3.6.1 权值评估	93
3.6.2 比较各种异常值检测方法	94
3.6.3 权值的作用	95
3.7 测试题	95
第4章 拟合结果的不确定度	97
4.1 拟合优度、精确度与准确性	97
4.1.1 统计模型和数据的一致性	97
4.1.2 样本方差	98
4.2 参数估计值的不确定度	100
4.3 模型预测的不确定度	101
4.4 拟合图形检查	102
4.5 计算实例	103
4.5.1 常数拟合	104
4.5.2 直线拟合	105
4.5.3 余弦函数拟合	107
4.5.4 模型失配	108
4.6 测试题	114
 第II部分 数学、优化方法以及辅助内容	
第5章 矩阵代数	117
5.1 基础知识	117
5.2 行列式	120
5.3 矩阵求逆的数值计算方法	122
5.3.1 伴随矩阵法	122
5.3.2 高斯-约当消元法	123
5.3.3 LU分解方法	125
5.3.4 奇异值分解(SVD)方法	131
5.4 测试题	132

第 6 章 最小二乘方法的思想	133
6.1 正态分布	133
6.2 最大似然原理	134
6.3 拟合线性模型函数	135
6.3.1 标准方法	135
6.3.2 利用奇异值分解 (SVD) 进行求解	136
6.3.3 条件缩放	137
6.4 拟合非线性模型函数	138
6.4.1 误差曲面的近似	138
6.4.2 高斯 - 牛顿方法	139
6.4.3 梯度下降方法	141
6.4.4 Levenberg - Marquardt 方法	142
6.4.5 寻求极小值点的例子	143
6.5 测试题	153
第 7 章 补充工具和方法	154
7.1 其他参数估计方法	154
7.1.1 递归自适应参数估计方法	154
7.1.2 迭代的梯度下降方法	155
7.1.3 进化方法	156
7.1.4 利用反向传播算法对神经网络进行训练	157
7.2 用于异常值检测的 Chauvenet 准则	162
7.3 误差传播原理	164
7.4 利用手工推导线性最小二乘问题的解	167
7.4.1 直线拟合	167
7.4.2 维纳滤波	169
7.5 联合处理多种模型函数	171
7.5.1 例子 1: 坐标变换	173
7.5.2 例子 2: 圆周运动	174
7.6 针对相关数据进行加权最小二乘拟合	176
7.7 总体最小二乘 (TLS) 拟合	181
7.7.1 圆周的正交拟合	182
7.7.2 通用性的方法	183
7.8 测试题	186

附录 A 异常值检测方法的比较	187
A.1 正态分布数据集	187
A.1.1 没有异常值的数据集	187
A.1.2 含有异常值的数据集	189
A.2 非正态分布数据集	192
A.2.1 拉普拉斯分布	192
A.2.2 均匀分布	193
A.3 讨论	195
附录 B 实现	197
B.1 功能	197
B.2 使用说明	197
B.2.1 输入和输出	199
B.2.2 模型参数的初始化	206
B.2.3 处理控制	207
B.2.4 权值和异常值	207
B.3 源代码的通用架构	212
B.4 模型函数	214
B.4.1 数值微分	214
B.4.2 多维条件处理	215
B.4.3 参数取值范围	215
B.4.4 参数初始化	216
B.5 特殊算法	217
B.5.1 LU 分解	217
B.5.2 奇异值分解	217
B.5.3 排序	217
B.6 程序的优化	217
B.7 性能测试	218
B.7.1 拟合线性系统	218
B.7.2 拟合非线性系统	221
部分习题解答	226
参考文献	240

第 1 章 数据拟合问题的引入

1.1 什么是数据拟合

经典的数据拟合场景需要进行实验或者观测，在不同的实验条件下会得到不同的观测结果。例如，想象一下现要测量经过某个电阻 R 上的电流 I （称为观测量），该电流显然取决于电阻两端的电压 U （称为实验条件）。

一般而言，观测结果与实验条件之间的基本函数关系是已知的或是被假定的，该关系式可以利用一个数学模型进行刻画，本书称其为模型函数。例如，如果假设上面讨论的电阻 R 是个常数，那么根据欧姆定律可知其函数关系式为 $I = U/R$ 。如果是其他类型的电器元件， I 和 U 之间的关系式可能会有所不同，此时就需要采用其他更为适合的模型函数。

数据拟合的目的就是要确定模型函数中的参数，以使其能够最好地刻画观测量与实验条件之间的关系，其数学模型如下。

将测量或者实验结果记为观测量 y ，它与实验条件有关，不妨用向量 $\mathbf{x}$ 来表征条件。模型中的全部参数包含在向量 $\mathbf{a}$ 中，假设其中共有 M 个元素 $\{a_j\}_{1 \leq j \leq M}$ ，并且这些参数都是未知的，需要利用数据拟合方法进行求解。于是，可将模型函数表示为

$$y = f(\mathbf{x} | \mathbf{a}) \quad (1.1)$$

在前面提到的欧姆定律的例子中，电流 I 就是这里的观测量 y ，条件向量退化为标量 $x = U$ ，而 $\mathbf{a} = a = R$ 是唯一需要确定的参数。

在单次实验中往往可以获得多个观测量。例如，除了可以测量经过电阻的电流 I 以外，还可以测量其温度 T 。然而，相应的模型函数 $T = f(I)$ 是不同的，此刻先将 y 考虑为一个标量，而非向量。

在实际中无法连续改变实验条件，通常用变量 i 来刻画离散的实验条件。此外，观测量会受到随机误差（即噪声）的影响，于是可将上述问题重新描述为

$$y_i = f(\mathbf{x}_i | \mathbf{a}) + \varepsilon_i \quad (1.2)$$

式中： $\mathbf{x}_i$ 为第 i 个实验条件，可以将其看成自变量； y_i 可以看成是应变量。

假设 $\mathbf{x}_i$ 是精确已知的，或者 $\mathbf{x}_i$ 中的误差相比于观测量 y_i 中的误差足够小，