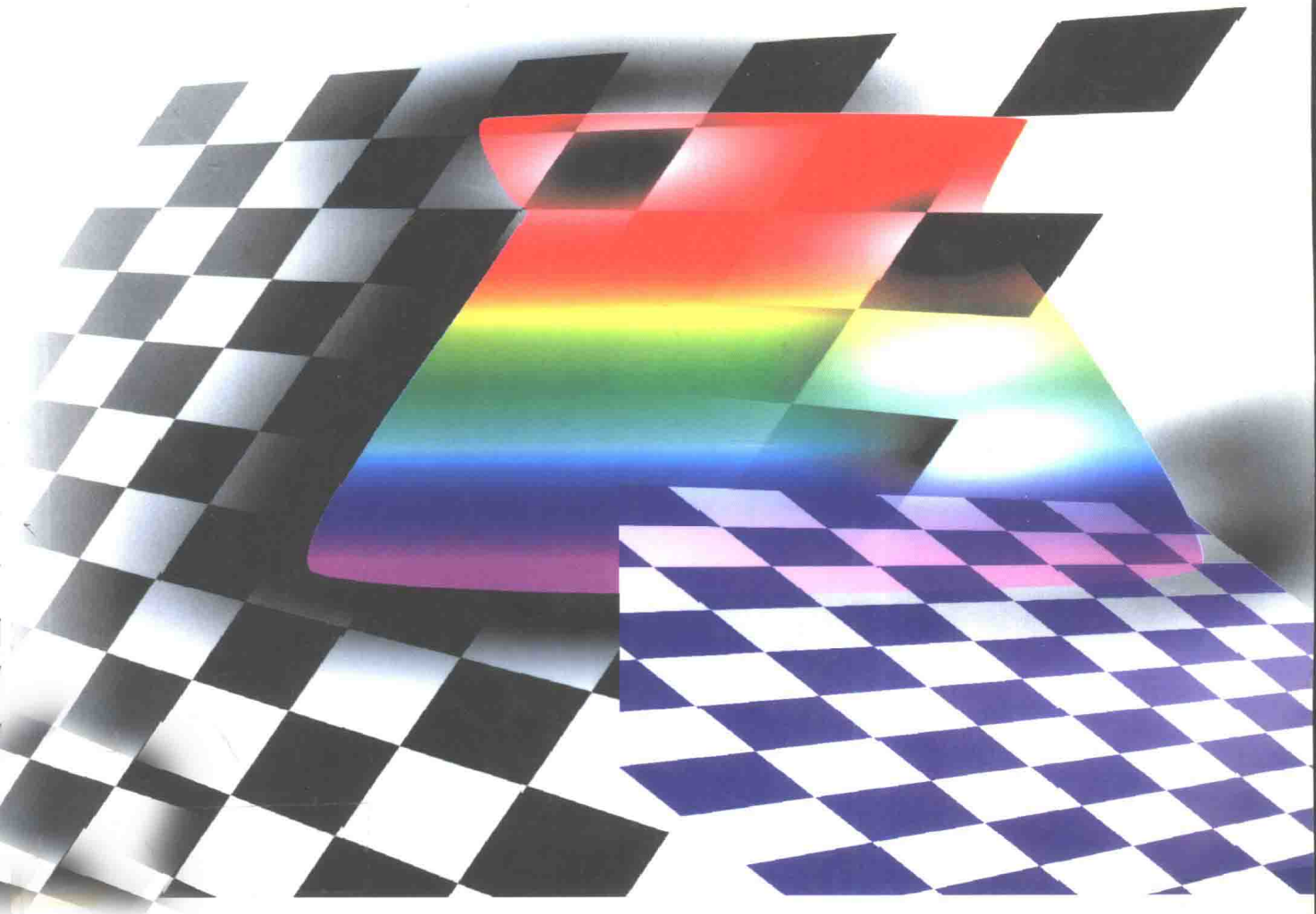
基础化学计量学

刘树深　易忠胜　编著



科学出版社

基础化学计量学

刘树深　易忠胜 编著

科学出版社

1999

内 容 简 介

本书是在化学计量学课程所用讲义基础上修、改、补而成，是编著者多年教学和科研的经验和成果的总结. 全书共六章，主要讲述化学试验设计与优化、分析信号处理、基础校正理论、化学因子分析基础和化学模式识别基础等. 书末还附有矩阵基本知识、常用正交表和均匀设计表、Ture BASIC 初步等 3 个附录和经过精选的参考文献.

本书可作为大专院校化学类专业的化学计量学教学用书，也可供有关专业的师生及从事于工程、化学、化工等科技人员参考.

图书在版编目(CIP)数据

基础化学计量学/刘树深，易忠胜编著. -北京：科学出版社，1999
ISBN 7-03-007271-5

Ⅰ. 基… Ⅱ. ①刘… ②易… Ⅲ. 化学计量学 Ⅳ. 06-04

中国版本图书馆 CIP 数据核字（1999）第 02723 号

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717

新蕾印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1999 年 8 月第 一 版 开本：787×1092 1/16
1999 年 8 月第一次印刷 印张：13
印数：1—2 700 字数：292 000

定价：20.00 元

（如有印装质量问题，我社负责调换〈新欣〉）

序　言

编写一本适用于本科化学类专业学生使用的化学计量学教学用书，是高等教育改革的需要，也是编著者多年的心愿．这不仅因为化学计量学作为化学、特别是现代分析化学的重要基础理论具有十分广阔的应用前景，更因为目前国内尚无相应教材出版．本书《基础化学计量学》是在桂林工学院工业分析专业本科生中开设的化学计量学课程所用讲义的基础上全面修订、补充加工而成的．

化学计量学是应用数学和统计学、计算机科学与化学、分析化学等多学科交缘的产物，内容非常丰富．在几十个学时中，要掌握化学计量学的全部内容是不可能的，本书仅选用一些基本定型，应用非常广泛且原理不很复杂的适用于化学类本科学生阅读的基础性内容，内容包括试验设计与优化、分析信号处理、基础校正理论、化学因子分析、化学模式识别基础等．考虑到多数化学系尤其是分析化学专业本科学生数学知识面较窄，难在短时间内领会化学计量学中的全部数学内容；同时，一般化学系学生都修过线性代数、概率统计和计算机应用基础等课程，本书以附录形式将必备的应用矩阵知识介绍给学生，以弥补学生因这些知识之不足而将有限光阴消耗于无关宏旨的数学填空．实验设计方法涉及许多标准表格，为便于使用，附录 B 中列出一部分正交表和均匀表．化学计量学是一门新兴学科，许多名词尚无定论，书中附录 C 给出计量学中出现的一些常用名词供参考．

编写计算机程序是实现化学计量学算法的前提．编写程序和调试虽然要花费大量时间，但对于同学们理解化学计量学算法却是一种有效且必不可少的练习手段．书中虽然没有给出程序清单，但书中所有例题都配有 True BASIC 计算程序（另册），供同学们上机调试、阅读、修改，最后达到理解各程序的目的．程序的编写反映了近年来编著者从事化学计量学研究和指导工业分析专业学生毕业实习的有关成果．在程序清单中，编著者有意让一些程序编制符号不同于原理所用符号，而另一些程序则留有空白，其目的在于启发和引导同学们在编制程序过程中真正理解化学计量学算法．考虑到部分学校未开设 True BASIC 语言，附录 C 介绍了该语言基本知识，并总结了部分表格，再结合对开始部分程序所作的注释，想必学过 BASIC 语言或 FORTRAN 语言的读者理解本书的有关程序是不成问题的．

在本书付梓之际，编者感谢中国科技大学张懋森教授、已故蒲国刚教授，是他们将编著者引进化学计量学学科的大门；感谢湖南大学俞汝勤院士、梁逸曾教授在全国高等学校化学计量学讲习班上给予了许多教诲与鼓励；感谢湖南大学李志良教授在作访问学者期间给予的指导和关心；感谢中国科学院盐湖研究所翟宗玺研究员、夏树屏研究员和高世扬院士十余年来一如既往的关心和鼓励，才得以能够完成本书．感谢化工学报主编叶铁林编审的无私推荐；感谢科学出版社李义发编审的指导和精心编辑、加工；感谢桂林工学院跨世纪人才基金和教材建设基金给予的支持使本书能够顺利出版．

由于编著者水平有限，书中缺点和错误在所难免，恳请读者批评指正．

编著者

目　录

序言

第一章　绪论 …… (1)

§1.1　什么是化学计量学? …… (1)

§1.2　化学计量学发展简况 …… (2)

§1.3　化学计量学实验程序 …… (2)

§1.4　本书各章内容简介 …… (4)

第二章　化学试验设计与优化 …… (6)

§2.1　试验设计基本概念 …… (6)

§2.2　析因设计(FD) …… (7)

2.2.1　析因设计表[$FD_n(2^m)$] …… (7)

2.2.2　析因设计试验一般步骤 …… (8)

§2.3　正交试验设计 …… (12)

2.3.1　正交表 …… (12)

2.3.2　正交试验设计的一般步骤 …… (13)

§2.4　均匀试验设计 …… (16)

2.4.1　均匀设计表 …… (16)

2.4.2　均匀试验设计一般步骤 …… (18)

§2.5　单纯形试验设计与优化 …… (23)

2.5.1　单纯形与单纯形试验设计 …… (23)

2.5.2　构造初始单纯形 …… (24)

2.5.3　单纯形推移(操作) …… (27)

2.5.4　单纯形优化算法与例题 …… (29)

§2.6　响应面分析简介 …… (35)

第三章　分析信号处理 …… (38)

§3.1　概述 …… (38)

§3.2　分析信号的均值滤波 …… (39)

§3.3　信号平滑与求导 …… (40)

3.3.1　Savitzky-Golay 多项式平滑与求导原理 …… (41)

3.3.2　五点二次平滑 …… (43)

3.3.3　一阶导数和二阶导数 …… (47)

§3.4　曲线拟合 …… (51)

3.4.1　非线性函数曲线拟合 …… (51)

3.4.2　峰参数估计 …… (55)

§3.5　谱峰面积估计(数值积分) …… (59)

3.5.1 梯形法(Newton 法) …… (59)
3.5.2 辛普森抛物线法(Simpson 法) …… (59)
第四章 基础校正理论 …… (63)
§4.1 概论 …… (63)
§4.2 单变量校正 …… (63)
4.2.1 校正曲线法(CCM) …… (63)
4.2.2 标准加入法(SAM) …… (67)
§4.3 化学量测数据的矩阵表示 …… (69)
§4.4 多元线性回归校正 …… (70)
4.4.1 多元校正概述 …… (70)
4.4.2 多元线性回归(MLR)校正 …… (71)
§4.5 卡尔曼滤波(KF) …… (74)
4.5.1 卡尔曼滤波用于校正基本原理 …… (74)
4.5.2 迭代算法 …… (75)
§4.6 **K** 矩阵和 **P** 矩阵间接校正 …… (80)
4.6.1 **K** 矩阵法 …… (80)
4.6.2 全校正和部分校正 …… (83)
4.6.3 **P** 矩阵法 …… (84)
§4.7 通用标准加入法(GSAM) …… (87)
§4.8 非线性校正 …… (90)
4.8.1 化非线性为线性(拟线性化) …… (91)
4.8.2 不化为线性的情况 …… (92)
第五章 化学因子分析基础 …… (94)
§5.1 概论 …… (94)
§5.2 模拟数据矩阵与协方差矩阵 …… (95)
5.2.1 模拟数据矩阵的产生 …… (95)
5.2.2 协方差矩阵 …… (96)
§5.3 主成分分析(PCA) …… (100)
5.3.1 主成分分析的基本思路 …… (100)
5.3.2 特征值与特征矢量 …… (101)
5.3.3 主成分数或重要因子数或组分数的判别 …… (107)
§5.4 因子分析的基本过程 …… (111)
5.4.1 预备 …… (111)
5.4.2 复原 …… (111)
5.4.3 变换 …… (112)
5.4.4 得分矢量 t_k 与载荷矢量 v_k …… (112)
§5.5 目标转换因子分析(TTFA) …… (114)
§5.6 主成分回归(PCR) …… (118)
§5.7 偏最小二乘法(PLS) …… (122)

5.7.1 基本原理 …… (122)
5.7.2 程序框图与应用 …… (124)
第六章 化学模式识别基础 …… (126)
§6.1 概论 …… (126)
§6.2 预处理和特征提取 …… (127)
6.2.1 数据预处理 …… (127)
6.2.2 特征提取或选择 …… (129)
§6.3 相似系数和距离 …… (131)
6.3.1 相似系数 …… (131)
6.3.2 距离 …… (132)
§6.4 线性学习机(LLM) …… (134)
§6.5 K 最近邻法(KNN) …… (140)
§6.6 SIMCA 方法 …… (145)
6.6.1 对训练集中每一类建立主成分模型 …… (145)
6.6.2 未知样本测试 …… (147)
§6.7 降维与显示技术 …… (152)
6.7.1 线性降维与显示 …… (153)
6.7.2 非线性降维与显示 …… (155)
§6.8 聚类分析——系统聚类方法 …… (158)
6.8.1 基本原理 …… (159)
6.8.2 类间距离的定义与系统聚类方法 …… (159)
附录 A 矩阵基本知识 …… (167)
§A.1 矩阵的简单运算 …… (167)
A.1.1 加法和数乘 …… (167)
A.1.2 矩阵乘法 …… (167)
A.1.3 矩阵的转置和对称性 …… (168)
A.1.4 某些特殊矩阵 …… (168)
A.1.5 矩阵的逆 …… (168)
A.1.6 矩阵表达式的转置和求逆 …… (169)
A.1.7 矩阵的秩 …… (169)
§A.2 特征值和特征矢量 …… (170)
A.2.1 矩阵的特征值 …… (170)
A.2.2 矩阵的特征矢量 …… (171)
A.2.3 对称矩阵特征矢量的正交性条件 …… (172)
附录 B 常用正交表和均匀设计表 …… (173)
§B.1 常用正交表 …… (173)
§B.2 常用均匀表 …… (178)
附录 C True BASIC 初步 …… (183)
§C.1 True BASIC 的基本语句与程序结构 …… (183)

C.1.1 3个简单语句 …… (183)
C.1.2 程序格式 …… (183)
C.1.3 二个结构控制语句 …… (184)
C.1.4 三种基本的程序结构 …… (185)
C.1.5 程序的书写 …… (185)
§C.2 True BASIC 的基本操作 …… (186)
C.2.1 上机操作步骤 …… (186)
C.2.2 一些常见的操作键 …… (186)
C.2.3 操作实例 …… (186)
§C.3 两个通用子程序 …… (188)
C.3.1 数据表格输出子程序(DATAB 子程序) …… (188)
C.3.2 图形输出子程序 …… (189)
§C.4 True BASIC 常用索引表 …… (190)
C.4.1 True BASIC 常用函数索引表 …… (190)
C.4.2 True BASIC 常用命令索引表 …… (191)
C.4.3 True BASIC 常用语句索引表 …… (191)
参考文献 …… (196)

第一章 绪 论

§1.1 什么是化学计量学?

科学的发展与技术的进步使得化学量测工作逐步仪器化、自动化和计算机化. 现代分析仪器能迅速、准确地为人们提供大量可靠的量测数据. 化学工作者面临着如何选择合适的实验方法和最优量测过程, 对原始量测数据进行再加工, 从而最大限度地提取有用的化学及其相关信息. 随着计算机科学、应用数学和统计学方法在化学中应用的日益广泛和深入, 一门崭新的化学分支学科即化学计量学 (chemometrics) 诞生了. 20 多年的学科实践证明, 化学计量学理论和方法已渗透到化学中的各个领域, 已经成为化学量测的基础理论和方法学.

化学计量学是数学和统计学、化学及计算机科学三者相互交缘而形成的一门边缘学科, 是化学中很具有魅力和应用前景十分广泛的新兴分支学科. 按照国际化学计量学学会 (International Chemometrics Society, 简称 ICS) 的定义: 化学计量学是化学的一门分支学科. 它应用数学和统计学方法, 设计或选择最优量测程序和实验方法, 并通过解析化学量测数据而获取最大限度的信息. 理解这个定义应考虑以下几点: 1. 要选择最优量测程序并获取最大限度信息, 必须也只有借助计算机技术; 2. 化学计量学研究和探讨各种化学量测过程的共性问题, 如化学试验设计与优化、化学数据解析及有用信息的提取等, 因而它是有关化学量测的基础理论和方法学; 3. 化学计量学是化学、数学和统计学以及计算机科学诸多学科的“接口”, 但同时应该注意到化学计量学又是一个学科总体, 有其自身的学科体系. 另一方面, 化学计量学可以更具体地表达为研究应用数学和统计学方法, 借助计算机技术, 进行化学量测的试验设计、数据处理、分类、解析和预测的一门学科.

应当指出, chemometrics 一词是 Wold, S. 仿 econometrics (经济计量学) 一词而提出的, 其中文译为“化学计量学”即反映了其原意. 中文文献中曾提出过“化学统计学”或“数统化学”等多种译法, 这些译法因较难反映化学计量学作为“接口”的内涵. 因为化学计量学不仅涉及统计学, 也包括逻辑学、拓扑学和图论方法等诸多数学方法的运用, 故上述译法现今很少被人们所采用.

化学计量学的研究范围极为广泛, 内容非常丰富. 化学试验设计与优化、定量校正理论、分析信号处理、化学模式识别、模型与参数估计、数据解析、过程模拟、人工智能、情报检索、实验室自动化等等都是化学计量学的研究范围. 化学计量学作为化学量测的基础理论与方法学, 其应用非常广泛. 可以说, 凡是施行化学量测的所有领域, 如工业过程采样、过程分析化学、过程控制、食品工业、海洋化学、地球化学、环境化学、造纸工业、石油勘探、临床诊断、制药工业、染料工业、有机合成化学、生物工程、材料工程等等都可以应用化学计量学.

§1.2 化学计量学发展简况

化学计量学诞生于70年代初期. 1971年，瑞典化学家Wold，S. 在为一项基金项目定名时，从“化学数据分析”(chemical data analysis)、“化学中的计算机”(computer in chemistry)和“化学计量学”三者中选定后者而正式宣布了化学计量学这门新兴学科的诞生. 三年后（1974年），他与美国华盛顿大学的Kowalski，B. R. 教授在美国西雅图成立了国际化学计量学学会（ICS). 随着80年代计算机的普及应用，化学工作者不仅应用现有的数学和统计学方法，而且根据化学学科特殊性要求创建了一系列化学量测数据的处理、分类、解析与预测等一大批化学计量学方法，编制了许多优秀的化学计量学软件，很多软件已成为现代化学量测仪器主要是分析仪器的有机组成部分. 80年代后期，化学计量学课程开始进入化学教学大纲，专门刊登化学计量学学术研究成果的学术期刊“*Journal of Chemometrics*”、“*Chemometrics & Intelligent Laboratory Systems*”等问世，检阅化学计量学研究成果的国际学术会议也陆续召开. 进入90年代，由于计算机及软件技术的飞速发展，符号处理高级语言的普及，使许多过去认为过于复杂难于普及的化学计量学算法逐步得到推广与应用，特别是信息高速公路的建立，广大化学计量学工作者可以更快更及时的了解化学计量学学科发展动态. 1994年和1996年已分别召开了两届Internet国际化学计量学学术会议. 毫无疑问，化学计量学作为计算机科学等多学科的“接口”正进入一个大发展时期.

§1.3 化学计量学实验程序

化学计量学实验程序的设计、编写、修改、调试、运行等是深入理解和掌握化学计量学方法的重要步骤，也是研究化学计量学方法性能、影响因素、应用条件等的重要手段，同时也是从事化学计量学研究或应用化学计量学方法的广大化学工作者最基本的训练内容. 本书作为基础化学计量学，切实考虑化学工作者的数理基础和计算机水平，应用具有语法简单、计算功能特别是矩阵运算功能和绘图功能较强等特点且已广泛应用的高级程序语言——True BASIC语言编制实验程序，以实现书中给出的各种化学计量学方法. 考虑到应用计量学方法解决化学实际问题的基本过程，我们将计算实验程序设计为数据输入、算法处理和结果输出等三大模块. 在程序结构上，将各种数据输入和输出，数据处理过程设计为单独的子程序（附录C给出两个输出“数据表格”和“二维图形”的通用子程序)，而主程序只包含一些调用语句和一些简单的输入输出语句. 此外，为减少读者在理解各实验程序时不断查阅参考书的麻烦，书末附录C介绍了True BASIC语言的基本知识.

作为示例，以通过化学量测数据估计线性模型各参数的过程编写True BASIC语言程序. 设经化学实验量测得到 n 个浓度分别为 x_i ($i=1,2,\cdots,n$)的某组分标准溶液的信号(比如在光度计上测得的吸光度信号 y_i)如表1.1所示.

表 1.1　某纯组分 5 个标准溶液的吸光度

x_i	20	40	60	80	100
y_i	0.120	0.244	0.359	0.502	0.599

设吸光度与浓度之间满足线性方程

$$\hat{y}_i = a + b \cdot x_i \qquad (i = 1,2,3,\cdots,n),$$

其中 $\hat{y}_i$ 为在任意给定 x_i 值下吸光度的估计值. 根据最小二乘原理可求得 a 和 b 的最佳估计及相关系数 r(参见§ 4.2)

$$b = \frac{L_{xy}}{L_{xx}}, a = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i\right), r = \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}},$$

其中

$$L_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i,$$

$$L_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2,$$

$$L_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2.$$

那么估计这个线性方程的参数(a, b 及 r)的实验程序可编制成 True BASIC 程序 PRAG1.1.

```
! PRAG1.1
! 主程序
DIM x(1),y(1)                          ! 定义存贮 x_i 和 y_i 的两个数组
READ n                                 ! 读入数据个数
MAT READ x(n)                          ! 读入 5 个 x_i 数据
MAT READ y(n)                          ! 读入 5 个 y_i 数据
DATA 5
DATA 20,40,60,80,100                   ! 与三个 READ 语句对应的数据语句
DATA 0.120,0.244,0.359,0.502,0.599
CALL LR(n,x,y,a,b,r)                   ! 调用计算线性回归参数的子程序 LR
PRINT "a=";a,"b=";b,"R=";r             ! 屏幕输出
END                                    ! 主程序结束

SUB LR(n,x(),y(),a,b,r)                ! 子程序 LR 开始
    LET x1,x2,y1,y2,xy=0               ! 赋初值
    FOR i=1 TO n
        LET x1=x1+x(i)                 ! 计算 Σx_i
        LET x2=x2+x(i)*x(i)            ! 计算 Σx_i^2
        LET y1=y1+y(i)                 ! 计算 Σy_i
        LET y2=y2+y(i)*y(i)            ! 计算 Σy_i^2
        LET xy=xy+x(i)*y(i)            ! 计算 Σx_iy_i
```

```
    NEXT i
    LET lxy = xy - x1 * y1/n                    ! 计算 L_xy
    LET lxx = x2 - x1 * x1/n                    ! 计算 L_xx
    LET lyy = y2 - y1 * y1/n                    ! 计算 L_yy
    LET b = lxy/lxx                             ! 计算 b
    LET a = (y1 - b * x1)/n                     ! 计算 a
    LET r = lxy/sqr(lxx * lyy)                  ! 计算 R
END SUB                                         ! 子程序结束
```

为便于初学者理解,程序中每一个语句以"!"形式对该语句的功能作了说明.在以后各章节中,这种说明将随着读者不断熟悉 True BASIC 而逐渐取消.

§1.4 本书各章内容简介

作为基础化学计量学，本书主要介绍化学计量学中比较成熟，原理相对比较简单，且易于为广大化学工作者所掌握的基础性内容. 全书共分六章，包括绪论，化学试验设计与优化，分析信号处理，基础校正理论，化学因子分析基础和化学模式识别基础等. 此外，书末列出矩阵基础知识，常见正交表与均匀表以及 True BASIC 语言基本知识 3 个简短附录.

在"绪论"中简要介绍了化学计量学产生背景、定义、内容与应用、发展简史等.

化学量测过程从选择实验对象、采样、检测、校正、预测直至解决实际问题等各个环节，无一不涉及试验设计与优化问题. "化学试验设计与优化"一章将首先介绍试验设计的一些基本知识；进而对两水平析因设计(FD)试验方法进行初步讨论;正交试验设计(OD)虽然是数理统计方法在化学中应用最为广泛的试验设计与优化方法之一，但有关应用文献与参考书较多，本书仅扼要讨论正交试验设计的基本操作过程；在介绍我国学者方开泰和王元等提出的均匀设计方法时，导出了多元线性回归的解析运算与矩阵运算之间的关系，给出了数学推导虽不甚严格但便于记忆的多元回归的矩阵运算方法. 通过初始单纯形的构造和单纯形的推移过程，结合实例，给出了单纯形优化的基本原理和具体推移过程；最后简要介绍了响应面分析方法.

化学量测所得到的分析信号一部分来源于有用组分和干扰组分的贡献，一部分来源于仪器噪声和测量误差. 如何消除量测误差，滤除仪器噪声，复原被扭曲的分析信号等等则是化学计量学中有关分析信号处理的主要内容. "分析信号处理"一章简要介绍均值滤波、卷积平滑和求导、曲线拟合、峰参数与峰面积估计等分析信号处理的基本内容.

校正主要是多元校正，它是化学计量学中最具特色且最有活力的重要分支之一. "基础校正理论"一章从经典分析化学中的单变量校正方法开始，在推导校正曲线法(CCM) 和标准加入法(SAM)时，定义了校正集、检验集和预测集等常用校正基本概念. 多元线性回归(MLR)技术或多元线性最小二乘法是一种经典优化技术，在自然科学与社会科学等各个领域都得到广泛应用. 本章讨论了 MLR、**K** 矩阵法、**P** 矩阵法、卡尔曼滤波(KF)等多元校正方法.对于存在基体效应,或者校正集标准样本很难得到的

多元体系，应采用通用标准加入法(GSAM).对于非线性多变量体系，一般先施行数学变换(比如对数变换等)，化简为线性校正模型予以校正和预测，或者直接采用非线性算法进行处理.

“化学因子分析基础”一章从三组分光谱模拟数据矩阵出发，首先探讨主成分分析(PCA)的基本思想，给出主成分分解(TV^T)和奇异值分解(USV^T)两种迭代算法求得特征矢量及特征值的基本过程；进而介绍确定化学量测体系或数据矩阵的重要因子数(主成分数)的基本方法；最后在因子分析的基础上介绍应用目标转换因子分析(TTFA)，主成分回归(PCR)和偏最小二乘法(PLS)于实际量测，如多组分体系的定性定量测定.本章内容难度较大，算法也比较复杂，但应用非常广泛，读者务必结合实验程序逐步加深理解，做到融会贯通，达到学以致用.

“化学模式识别基础”是从大量多维化学量测数据中提取有关物质隐含信息并予以归类的基本方法.化学量测数据可能来自不同总体，或采自不同方法或不同分析仪器，各组数据的方差不完全相同，对这些数据应先作必要的预处理.本章先介绍数据预处理方法和相似性量度指标；进而给出最早应用于化学领域的线性学习机(LLM)以及 K 最近邻法(KNN)法等经典模式识别方法；化学计量学创始人 Wold,S.基于主成分分析和偏最小二乘法提出的 SIMCA 方法具有更多的优点，本章作了适当的介绍.无教师分类方法的典型代表是基于“物以类聚”常识建立起来的聚类分析方法，原理简单，应用广泛，也是本章的重要内容.

第二章　化学试验设计与优化

§2.1　试验设计基本概念

要从化学量测数据中提取最大限度的有用信息，必须进行试验设计与优化．化学中涉及的试验设计问题甚为广泛．比如分析化学中分析方法选择、反应条件优化、光谱分析中波长选择、色谱分析中固定相选择以及各种仪器方法中操作参数选择等等都需要进行优化试验设计；一个化工合成反应，从原材料选取，实验条件确定，到获得产品，也都必须进行精心的试验设计，才可能达到节约能源、人力、物力、时间、降低成本等优化目的．

那么什么是试验设计？试验设计是指在试验域（因素可取值的区域）内，最有效地选择试验点，科学地安排试验，进而通过数据解析求得指标取最优值的条件的一种方法．顾名思义，试验设计研究的是如何设计试验条件使指标取得最优值．

为了便于说明各种试验设计方法，先介绍如下几个常用概念：

（1）试验指标　试验设计中用来衡量试验效果的物理量称为试验指标（简称指标）．指标可以是单一指标（包括综合评价指标），也可以是多个指标．试验设计按指标个数多少可分为单指标试验设计和多指标试验设计．试验指标也可按性质不同区分为定性指标与定量指标两类．定性指标是指不能直接用数值精确表达的指标，比如水质的恶臭程度、油漆的亮度等；而定量指标是指能用确定的数值来表示的指标．如吸光度、峰高、谱线强度、产率等．试验设计的目的是要使试验指标取得最优，要求指标具有可比性，因此，在试验设计中通常总是将那些定性指标进行定量化处理后以数值表示．应当指出，在最优化领域中，指标也常称为目标或目标函数．

（2）因素及水平　影响试验指标取值的物理量称为因素，有时亦称为因子．因素在试验中所处的状态，称为水平．在试验中有几种状态就称有几个水平．如考察温度的影响，温度即是因素，如要试验 80℃ 和 100℃ 两个温度的影响，则这 80 和 100 即是该温度因素的两个水平．所选因素的水平发生变化时，一般应引起试验指标的变化，否则就认为该因素对指标没有影响，应从试验中删去．如果所考察的两个因素在试验中相互影响，即一因素水平的高低对另一因素水平对指标的贡献有影响，这时称这两因素之间存在交互效应．根据影响指标的因素多少，试验设计可分为单因素试验（设计）和多因素试验设计．化学试验设计一般都是多因素试验设计．

（3）同时试验和序贯试验　所谓同时试验，就是通过试验设计对有关因素的水平进行规划后，同时进行诸因素各水平的试验(此处“同时”的意义是指在多次试验中,先做或后做哪一次试验不会影响试验效果与试验进程),然后综合分析得到的试验结果，获得最优条件．而序贯试验设计方法则是每进行一次试验或少数几次试验后，先分析已取得的试验结果，再根据这些结果规划下一次试验，其试验次序是不能改变的，即按设定的顺序依次地进行，逐步向最优条件靠近．目前，应用广泛的正交试验设计、析因设计和均

匀设计方法都属于同时试验法，而序贯试验的典型代表是单纯形优化法.

那么如何进行试验设计以获得关于指标取最优值的各因素水平取值？如果指标和因素之间存在确定的函数关系(解析式)，则可用解析法进行优化. 但在绝大多数化学量测试验中，很难直接构造指标关于诸因素的函数表达式，不能用解析法. 通常，可采用两种方法达到优化的目的. 一是通过大量试验构造一个函数来逼近(无限地接近)这些数据，然后再用解析法求这个逼近函数的最优解，同时进行试验验证. 另一种方法是不研究试验指标与因素之间确定性的函数关系，只寻求试验指标最优的因素取值，此法称“黑箱”方法.

本章主要讨论化学试验设计中最基本和最常用的析因试验设计、正交试验设计、均匀试验设计、单纯形试验设计以及响应面分析等试验设计方法.

§2.2 析因设计(FD)

2.2.1 析因设计表[$FD_n(2^m)$]

多因素试验不仅要研究各因素水平对指标的影响，还要着重分析诸因素对指标的作用. 因此，多因素试验设计亦称析因试验设计或析因设计(factorial design，简称 FD). 应当注意的是，FD 不是一般意义下的多因素试验设计，它是将各因素的全部水平按一定规则相互组合进行试验，以考察各因素的主效应(某因素对指标的影响程度)以及因素之间的交互效应(因素之间不同组合时对指标的影响程度)的优化试验设计方法. 换句话说，析因试验设计是按析因设计表设计试验方案，通过分析试验指标的变化决定各因素主效应和因素间交互效应的试验方法. 限于篇幅，本节只讨论两水平析因试验设计问题.

m 个试验因素，安排 n 次试验的两水平析因设计表写成通式为：$FD_n(2^m)$. 其中 2 表示因素的两个水平，$n=2^m$. 如果以“－”表示因素的低水平，而“＋”表示因素的高水平，则二因素二水平析因设计表[$FD_4(2^2)$]如表 2.1 所示. 表中第 1 列是试验序号；第二列是为了分析各因素对指标的平均影响而设计的，都以“＋”号(高水平)表示，记为 I；第三列是第 1 个因素(A)，从“－”即低水平开始，以“－”与“＋”相间的方式排列(其他二水平析因设计表也一样)；第四列是第 2 个因素(B)，试验安排按“－－”与“＋＋”相间的方式排列，实际上是在前一个因素(A)的水平上“加倍”后再以相间的方式排列. 纵观所有二水平析因表，这个基本规律都存在. 也即如果还有第三个因素存在，则再“加倍”以“－－－－”与“＋＋＋＋”相间的方式排列排列第三因素的实验顺序. 第五栏是两因素(AB)的交互效应列，其水平的排列遵守“乘法规则”. 即交互效应列的水平是两因素在同一试验中水平的乘积，且有同号相乘得正、异号相乘得负. 在运算中，把因素的 2 个水平“＋”与“－”看成是“正”与“负”. 比如第一个试验 AB 水平值为负(A 的“－”)与负(B 的“－”)相乘得正，即“＋”.

表 2.1 $FD_4(2^2)$析因设计表

No	I	A	B	AB
1	＋	－	－	＋
2	＋	＋	－	－
3	＋	－	＋	－
4	＋	＋	＋	＋

有了上述“相间加倍规律”和“乘法规则”，就可导出其他两水平析因表的试验设计方

案.例如 $FD_8(2^3)$析因设计表,应用上述两规则容易得出各次试验中因素的水平取值,结果见表 2.2.

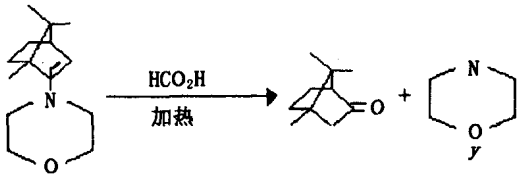
表 2.2　$FD_8(2^3)$析因设计表

No	*I*	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
1	+	−	−	−	+	+	+	−
2	+	+	−	−	−	−	+	+
3	+	−	+	−	−	+	−	+
4	+	+	+	−	+	−	−	−
5	+	−	−	+	+	−	−	+
6	+	+	−	+	−	+	−	−
7	+	−	+	+	−	−	+	−
8	+	+	+	+	+	+	+	+

2.2.2　析因设计试验一般步骤

首先根据化学经验或初步试验,挑选影响因素,确定大致范围,决定因素的两个水平(高水平"+"或"+1"和低水平"−"或"−1"),构造因素水平表;然后选择合适的析因设计表,据析因表安排试验并获得试验结果(指标);最后对指标进行分析,得出各因素主效应和交互效应.下面给出两个化学反应实例说明应用析因分析进行试验设计与优化的基本步骤.

例 2.1　有下列烯胺还原反应:



现考察甲酸(HCO_2H,因素 *A*)及加热温度(因素 *B*)对生成物产率的影响.请以析因设计方法分析各因素的主效应和交互效应.

解:先根据初步试验和化学经验确定因素的二个水平,所得因素水平见表 2.3(a).选择 $FD_4(2^2)$析因设计表安排试验方案,并进行试验获得生成物的产率 *y*(即指标),结果见表 2.3(b).

表 2.3(a)　因素水平表

因素	−1	+1
A(mol/L)	1.0	1.5
B(℃)	25	100

表 2.3(b)　烯胺还原反应试验方案与结果

No	*I*	*A*	*B*	*AB*	*y*(%)
1	+1	−1	−1	+1	80.4
2	+1	+1	−1	−1	72.4
3	+1	−1	+1	−1	94.4
4	+1	+1	+1	+1	90.6

要分析各因素主效应和交互效应,必须建立因素与指标之间的关系模型(此过程称建

模).设这个模型可以下列方程表示:

$$\boldsymbol{y} = a_0 + a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + a_3\boldsymbol{x}_1\boldsymbol{x}_2 + \boldsymbol{e}, \tag{2.1a}$$

式中 $\boldsymbol{y}$ 表示 4 次试验中生成物产率构成的矢量;$\boldsymbol{x}_j$ 表示第 j 个因素在 4 次试验中的水平矢量;$\boldsymbol{e}$ 为误差矢量,以 4 次试验误差为元素.则(2.1a)式可写成矩阵的形式如下:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ 1 & x_{31} & x_{32} & x_{31}x_{32} \\ 1 & x_{41} & x_{42} & x_{41}x_{42} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}, \tag{2.1b}$$

式中 x_{ij} 表示第 j 个因素在第 i 次试验中的水平取值("+1"或"−1");y_i 与 e_i 分别表示第 i 次试验产物产率和量测误差.以矩阵符号可将上式简记为

$$\mathbf{Y}_{4\times1} = \mathbf{X}_{4\times4} \cdot \mathbf{A}_{4\times1} + \mathbf{E}_{4\times1} \tag{2.1c}$$

利用析因设计方案中 **X** 矩阵和试验结果 **Y** 矩阵,对上式两边同时左乘 **X** 的逆矩阵 $\mathbf{X}^{-1}$ 即可求得系数矩阵 **A** 的估计值:

$$\mathbf{A} = \mathbf{X}^{-1}\mathbf{Y}. \tag{2.2a}$$

矩阵 **A** 中的第一个元素 a_0 是一个常量(平均贡献),a_1, a_2, a_3 分别是因素 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的主效应以及两者之间 $\boldsymbol{x}_1\boldsymbol{x}_2$ 的交互效应大小的量度.在析因试验设计中,**X** 矩阵的形式非常简单,仅由"+1"和"−1"为元素构成的矩阵,其逆存在、且其值是原矩阵 **X** 的转置矩阵的 $1/n$ 倍.其中 n 为析因设计的试验次数(本例 $n=4$).

由表 2.3(b)析因设计方案及式(2.1b),可得 **X** 阵

$$\mathbf{X} = \begin{pmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{pmatrix}. \tag{2.3}$$

通过求解该矩阵的逆矩阵 $\mathbf{X}^1$ 可得

$$\mathbf{X}^{-1} = \frac{1}{4}\begin{pmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 \end{pmatrix} = \frac{1}{4}\mathbf{X}^T, \tag{2.4}$$

式中矩阵符号右上角"−1"表示该矩阵的逆矩阵,而右上角"T"表示矩阵的转置矩阵.将式(2.4)代入式(2.2a)可得式(2.2b)

$$\mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}. \tag{2.2b}$$

根据矩阵乘法规则,由上式可以容易求得两因素的主效应及两者之间的交互效应(即各系数的大小),计算如下:

$$a_0 = \frac{1}{4}(+y_1 + y_2 + y_3 + y_4) = \frac{1}{4}(80.4 + 72.4 + 94.4 + 90.6) = 84.45,$$