动态系统辨识

试验设计与数据分析

G. C. 哥德温　R. L. 潘恩 著

科学出版社

动态系统辨识

试验设计与数据分析

G. C. 哥德温　R. L. 潘恩 著

张永光　袁震东 译

楼启明 校

科学出版社

1983

内 容 简 介

系统辨识是现代控制理论中的一个重要组成部分．它涉及到如何根据控制对象输入和输出的信息推算出对象的动态数学模型，包括其结构和参数．有了较精确的模型以后，才有可能对它进行有效的控制．

本书对于建立这样的动态数学模型的一些基础理论进行了系统的和深入的探讨．特别着重于在有随机干扰下如何估算动态系统的参数，以及在不同的实验条件影响下估计所得到的模型的准确程度．

本书适用于作为本专业高年级研究生的课本和一些中高级研究人员和教师的参考书．在阅读时需具备一定的预备知识，其中包括概率论、数理统计、随机过程理论、估算理论、线性代数以及线性系统理论等等．

Graham C. Goodwin and Robert L. Payne

DYNAMIC SYSTEM IDENTIFICATION

Experiment Design and Data Analysis

Academic Press, 1977

动 态 系 统 辨 识

试验设计与数据分析

G. C. 哥德温　R. L. 潘恩 著

张永光　袁震东 译

楼启明 校

责任编辑　陈文芳　袁放尧

科 学 出 版 社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1983年 3月第 一 版　开本：787×1092 1/32

1983年 3月第一次印刷　印张：10 5/8

印数：0001—5,550　字数：238,000

统一书号：15031 · 478

本社书号：2966 · 15—8

定价：1.65 元

译 者 的 话

七十年代以来，"系统辨识"这个分支学科得到了迅速的发展，吸引了大批自动控制理论、过程控制、应用数学、数理统计、信息论等领域的科学家和工程师．近几年来，相继出现了若干种有关介绍和论述系统辨识问题的专著，使得这个领域所包含的内容和所使用的工具以及所研究的对象逐渐地清晰了．读者现在手中拿着的这本《动态系统辨识》就是其中的一种．这是一本为研究生所写的教材，适合于从事自动控制理论、应用数学等有关领域的研究人员和实际工作者参考．

1978 年夏天本书翻译初稿完成后，我们曾组织讨论班学习．通过使用，我们认为本书具有如下的特点：

(1) 取材新颖．本书包括了系统辨识近年来的主要成果，其中有作者尚未来得及发表的成果（如第七章的部分内容），也有作者本人的最新成果(如第六章的一些内容)．

(2) 结构完整．书中对系统辨识的三个主要方面：模型结构研究(标准形)、模型参数的估计方法、最优测试信号，作了介绍和研究，既使读者对系统辨识所包括的各个方面都有所了解，而内容又不庞杂．

(3) 本书比较注意推理，使用了较强的数学工具，因此有较为浓厚的应用数学的色彩．

(4) 本书每章都带有适量的习题，而且书后给出了解答和提示．这有助于学习和教学使用．

当然，本书也有一些不足之处．如有些定理的证明过于简略，必须查阅有关文献才可弄清（对于工程技术人员来说，

有些定理的证明是不必深究的).

本书的主体部分是四至七章. 前三章是概率论与数理统计的有关预备知识. 书中在个别地方用到的一些结论,如鞅的收敛定理、矩阵的特殊演算公式、谱分解定理、凸集的有关结果等,全放在书后的附录中.

限于译者水平,书中错误或不妥之处在所难免,敬请读者批评指正.

一九七九年九月

前　　言

本书研究利用试验数据构造数学模型的理论．在很多的场合下，都希望能提出系统的精确数学模型．例如，为了进行预测和控制，工程师们就常常需要模型．其他一些应用部门如经济学、生物学及社会学也是如此．

本书所包括的内容适合作为研究生关于动态系统的估计方法课程．这些内容曾在帝国大学，新南威尔斯大学及纽卡斯尔大学的高年级学生中讲过．全部内容可在一个学期内 40 个学时教授完毕．本书也可供在这个领域中从事工作的研究人员及实际工作者参考．

每一章都包括一部分习题．这些习题有一般的练习，也有促进读者深入理解内容的提问．在本书的末尾还提供了全部习题的扼要解答．所以，即便遇到最困难的问题，读者都可以得到解决的借鉴，从而满怀信心地把这本书读下去．

所需要的预备知识包括：概率论、数理统计、随机过程以及线性系统理论．可选用的背景材料包括参考文献 [10]，[11] 及 [5]．建议同时阅读的包括参考文献 [12]，[14]，[25] 及 [17]．附加的一些背景材料可以在参考文献 [16]，[21]，[23]，[24]，[29]，[32]，[33] 中找到．在技术性的文献中也有不少关于辨识的文章．研究辨识问题的一个有效起点是 IFAC（国际自动控制联合会）的四次关于辨识与系统参数估计的讨论会（它们分别是 1967 年和 1970 年在布拉格，1973 年在海牙/戴尔夫特，1976 在第比利斯召开的）．再有就是 1974 年 11 月 IEEE Transaction on A-C 关于系统辨识和

时间序列分析专辑.

在撰写这本书时，遇到了一些争议．作者在 Tukey[147] 的文章中发现了一段引文，对于解决这些争议是有用处的："近似地回答一个正确的、但常常是提得比较含糊的问题，要比严格地回答一个能精确表述然而是错误的问题好得多."作者衷心地感谢曾与之进行过有益讨论的同事们和学生们，这里不再一一列举他们的姓名．但要特别感谢的是：Brain Anderson 教授，Reg Brown 博士，Tony Cantoni 博士及 Paul Kabaila, David Mayne 教授，John Moore 教授及 Tung Sang Ng, Neville Rees 教授，Lennart Ljung 教授，Tom Kailath 教授及 Marting Zarrop.

原稿是由 Lorraine Pogonoski 及 Dianne Piefke 整理的，他们熟练的技术避免了许多校对工作. Kristene Sanson 出色地提供了所需要的资料副本.

Martin Ooms 提供了本书的插图.

最后，还要感谢我们自己的夫人 Rosslyn 及 Kären, 没有她们的支持和鼓励是写不出这本书的.

作　者

目　　录

第一章　引论与统计背景

§1.1　引　　论

对于大多数自然科学，数学模型的概念是一个基本的概念．例如，若目的在于预报和控制，模型则是必不可少的．模型的类别与精确度和心目中所应用的具体对象有关，如用于宇航的模型通常是需要非常准确的，而对于工业过程（象高炉）的模型常常是很粗糙的．模型可以由物理的推理得到，也可以从分析系统的实验数据得到．对后一种情况，所得模型的精确性受随机干扰的限制，如不可测的扰动和测量误差等．

本书研究从带有噪声的观测获取物理系统的数学模型问题．特别是我们要研究动态系统模型中的参数估计问题．我们也要探讨各种实验条件对模型精度的影响．

我们处理问题基本上是用统计的方法，即我们把扰动和测量误差作为随机过程的实现来处理．我们列出了概率论与数理统计的许多结论，同时也给出了一些线性系统理论的结果．我们认为读者以前都曾学习过这些内容，但为了完整起见，把它们概括汇总放在后面的附录中．我们首要的目的在于洞察其本质．因此，我们力图避免陷入不必要的一些数学术语，或者是为此目的而安排的实际的枝节问题．我们希望这本书可以作为研究人员的一个起始点，以及在实际中应用辨识方法的人员的一个指南．

我们给出了关于动态系统的实验设计与分析的一个巧妙的描述．我们着重于当前被人们关注的一些方面，也包括了

许多过去的书中所没有的结果. 下边简明地介绍一下本书的内容.

第二章中我们讨论最简单的估计方法，即最小二乘法. 我们研究这个估计量的性质，包括模型误差假设是否是高斯的两种情况. 最小二乘法是重要的，这是因为它本身固有的简单性，也因为它是许多比较复杂方法的基础. 鉴于它的重要性，我们也简明地介绍了解算最小二乘问题的一个优良的数值方法.

第三章我们介绍极大似然估计方法，对于非动态系统探讨了这个估计量的性质.

第四章研究对动态系统提供数学模型的问题. 我们讨论各种的规范形以及与线性系统的参数估计问题的关联，我们还介绍一种可描述很大一类随机系统的一般模型结构.

第五章介绍与动态系统有关的估计方法. 我们着重介绍最小二乘法，极大似然法，辅助变量法及广义最小二乘法. 我们还研究多输入多输出系统及最优化算法的一些细节问题. 我们也处理闭环系统，即有反馈的系统. 这个问题在实际中经常遇到，因为对一个系统进行开环运转往往是不太容易的.

第六章中考虑估计的精度问题. 证明了实验条件的选择(输入，采样间隔，预采样滤波器，等等)，对可达到的精度有明显的影响. 这一章提供了在时域和频域上的实验设计方法.

最后，在第七章中我们介绍递推的估计方法和递推的实验设计方法. 对于时变系统的情况特别有用. 我们把这个想法也应用到随机控制.

在本章后面几节，我们将复习几个统计推断理论的结果. 我们对这些结果处理得非常简单，因为我们认为许多读者以前就熟悉这些概念了. 但是，为了准确理解，我们把一些关键结果汇总起来，并给出表记. 比较熟悉统计推断背景的读者

可以直接阅读第二章.

§ 1.2 概 率 论†

1.2.1 概率空间

概率论研究的基础是概率空间 $(\Omega, \mathscr{A}, P)$ 的概念，Ω 称为样本空间，Ω 的元素 ω 称为实现. $\mathscr{A}$ 是 Ω 的一个集类，它对全部可数集运算是封闭的. $\mathscr{A}$ 称为 Ω 子集合的 σ-代数，$\mathscr{A}$ 的元素称为事件. P 是定义在 $\mathscr{A}$ 上且满足下列公理的集合函数:

(I) $0 \leqslant P(A) \leqslant 1$ 　　对一切 $A \in \mathscr{A}$

(II) $P(\Omega) = 1$

(III) $P(\bigcup A_n) = \Sigma P(A_n)$ 　　对 $\mathscr{A}$ 中的一切互不相交的集列 $\{A_n\}$，即当 $i \neq j$ 时，$A_i \cap A_j = \emptyset$.

满足 I, II 和III的集合函数称为一个**概率测度**.

1.2.2 随机变量与分布函数

定义在空间 $(\Omega, \mathscr{A}, P)$ 上的实值函数 $X(\cdot)$，如果对任意 x，$\{\omega: X(\omega) < x\} \in \mathscr{A}$ 成立，就称为一个**随机变量**，事件 $\{\omega: X(\omega) < x\}$ 的概率称为**分布函数**，且记为 $F_X(x)$. 我们总可以把 $F_X(x)$ 表示为两个函数之和，其中一个是绝对连续的，另一个是有可数多个不连续点的逐段为常数的函数(勒贝格分解引理[37]). 如果 $F_X(x)$ 绝对连续，那它就可以用**概率密度函数** $p_X(x)$ 如下表示出来:

$$F_X(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_X(y) dy_1 \cdots dy_n$$

† 见参考文献 [37, 38, 104—106, 40].

其中 x_i 和 y_i 分别为 x 和 y 的第 i 个分量. 在附录A中给出了一些常用的密度函数.

1.2.3 条件概率

考虑集合 A，$C \in \mathscr{A}$，且 $P(A) \neq 0$. 在给定 C 的条件下，A 的条件概率(记为 $P(A|C)$) 定义为:

$$P(A|C) = P(A \cap C)|P(A)$$

如果 $P(A|C) = P(A)$ 成立，我们就称 A 和 C 这两个事件是统计独立的.

对于随机变量，我们可定义条件分布函数为

$$F_{Y|X}(y|X<x) = P[(Y<y) \cap (X<x)]|P(X<x)$$

同样，我们可以定义条件概率密度(如果它是存在的):

$$p_{Y|X}(y|x) = p_{Y,X}(y,x)/p_X(x)$$

其中 $p_X(x)$ 是关于 x 的边际密度函数，如下:

$$p_X(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{Y,X}(y,x) dy_1 \cdots dy_n$$

1.2.4 数学期望

设 X 是一个随机变量. 我们定义 X 的数学期望值为

$$E_X(X) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x dF_X(x)$$

式中积分为 Stieltjes 型.

若 $F_X(x)$ 为绝对连续，则可以用密度函数来表达 $E_X(X)$，即

$$E_X(X) = \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} x p_X(x) dx$$

同样，可以定义条件期望

$$E_{X|Y}(X) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x p_{X|Y}(x|y)\, dx$$

§1.3 点估计理论†

考虑一个定义在概率空间 $(\Omega, \mathscr{A}, P)$ 上的随机变量 Y，其中 P 是带有参数的族 $\mathscr{P}$ 的一项．我们用 P_θ 表示 $\mathscr{P}$ 中的一般项，其中 $\theta\in\Theta$，$\Theta\subset R^p$．

我们得到的是 Y 的一个实现 y，我们称它为**数据**．我们定义**估计量**为随机变量 Y 的函数 $g(Y)$．对于给定的数据，我们称 $g(y)$ 为一个估计．

显然，我们希望 $g(y)$ 是符合于"自然的本来面貌(即 $P_\theta=P$) 的特定的 θ 的一个"好"的估计．但是，我们必须严格定义所谓的"好"意味着什么．下边定义通常刻划估计量好坏的一些性质．显然，具有一个或几个这种性质，并不意味着这个估计量对于给定的目的是"最好的"．

定义 1.3.1 如果对所有的 $\theta\in\Theta$，由概率 P_θ 得到的 $g(Y)$ 的数学期望就等于 θ，即

$$E_{Y|\theta}g(Y)=\theta \qquad \text{对一切 } \theta\in\Theta \tag{1.3.1}$$

则称 θ 的估计量 $g(Y)$ 是**无偏的**．

定义 1.3.2 如果对于 θ 的估计量 $g(Y)$ 和任意一个另外的估计量 $\tilde{g}$，满足：

$$\begin{aligned}&E_{Y|\theta}\{(g(Y)-\theta)(g(Y)-\theta)^T\}\\ \leqslant\ &E_{Y|\theta}\{(\tilde{g}(Y)-\theta)(\tilde{g}(Y)-\theta^T\}\\ &\qquad\qquad \text{对一切 } \theta\in\Theta\end{aligned} \tag{1.3.2}$$

则称此估计量是**一致最小均方误差的**．

注 对 θ 的关于最小均方误差一致性要求一般是太苛刻了，因为对于 $\tilde{g}=\theta$，(1.3.2) 式的右端为 0．如果我们限

---

† 见参考文献 [11, 15, 30, 27]．

定为某一类估计量，这个要求才较有意义.

定义 1.3.3 如果估计量 $g(Y)$ 在无偏估计类中关于 θ 具有一致最小均方误差，就称这个估计量是*最小方差无偏估计量*（MVUE）.

定义 1.3.4 在可表示为数据的线性函数的无偏估计类之中，如果估计量 $g(Y)$ 对于 θ 具有一致最小均方误差，则称它为*最优线性无偏估计*（BLUE）.

在无偏估计类中，找出适当的 MVUE 或 BLUE 估计量并不容易. 从实用的观点，只要说明所用的估计量的方差接近于所有无偏估计量方差的下界也就够了. 下面的定理给出了一个合适的下界.

定理 1.3.1 （Cramer-Rao 不等式）

设 $\{P_\theta:\theta\in\Theta\}$ 是样本空间 Ω 上的一个分布族，$\Theta\subset R^p$，并且对每一个 θ，P_θ 可表示为密度函数 $p_{Y|\theta}(\cdot|\theta)$. 则在正则性条件下，$\theta$ 的任一无偏估计量的协方差满足下述不等式:

$$\operatorname{cov} g \geqslant M_\theta^{-1} \tag{1.3.3}$$

其中

$$\operatorname{cov} g = E_{Y|\theta}\{(g(Y)-\theta)(g(Y)-\theta)^T\} \tag{1.3.4}$$

M_θ（称为 Fisher 信息矩阵）定义为

$$\begin{aligned} M_\theta = E_{Y|\theta}\{&[\partial \log p(Y|\theta)|\partial\theta]^T \\ &\times[\partial \log p(Y|\theta)|\partial\theta]\} \end{aligned} \tag{1.3.5}$$

证明 因为$g(Y)$ 是 θ 的无偏估计量，我们得到

$$E_{Y|\theta}\{g(Y)\} = \theta \tag{1.3.6}$$

即

$$\int_\Omega g(y)p(y|\theta)dy = \theta，\text{从而}\ \frac{\partial}{\partial\theta}\int_\Omega g(y)p(y|\theta)dy = I.$$

我们假设具有在积分号下求导的足够的正则性，于是得到