

系统辨识(1):辨识导引

丁锋

摘要

系统辨识是研究建立系统数学模型的理论与方法.虽然数学建模有很长的研究历史,但是形成系统辨识学科的历史才几十年.在这短暂的几十年里,系统辨识得到了长足的发展,一些新的辨识方法相继问世,其理论与应用成果覆盖了自然科学和社会科学的各个领域,包括物理学、生物学、地球科学、气象学、计算机科学、经济学、心理学、政治学等.在这样的背景下,回过头来深思系统辨识的一些基本问题,对系统辨识的发展不无裨益.作为系统辨识的导引论文,简单介绍了辨识的定义,系统模型与辨识模型,辨识的基本步骤与辨识目的,包括辨识试验设计与数据预处理,以及辨识方法的类别,包括最小二乘辨识方法、梯度辨识方法、辅助模型辨识方法、多新息辨识方法、递阶辨识方法等.

关键词

系统辨识;递推辨识;参数估计;辅助模型辨识;多新息辨识;递阶辨识

中图分类号 TP273

文献标志码 A

0 引言

控制理论与自动化技术的伟大成就是现代科学文明的标志.控制论辉煌的六十年^[1],控制论在社会和经济发展中的作用^[2-3],现代宇航升天、太空探索,高集成度计算机芯片的生产和广泛应用,无不说明控制科学与自动化技术的辉煌成就与贡献^[4].而控制理论与自动化的基础是数学模型.

数学模型(如牛顿第二定律 $F = ma$, 气态方程 $\frac{PV}{T} = nR$ 等)是所有科学中或特定对象中定量研究事物运动规律的基础.系统辨识是研究建立系统数学模型的理论和方法.

系统辨识、控制理论、状态估计是现代控制论中三个密切相关的研究领域,它们的关系如图1所示.控制理论和状态估计离不开系统数学模型,因此,辨识是控制理论和状态估计的基础,是现代控制论的基石.许多控制理论分析方法和综合方法、状态估计算法(如典型的观测器设计)都假设系统模型结构和参数是已知的.实际控制系统的分析和设计的首要任务是建立系统的数学模型,所以说系统辨识是自动化控制的一门基础学科.



图1 系统辨识、控制理论与状态估计三者之间的关系

Fig. 1 The relationship between identification, control and state estimation

系统建模与模型辨识(模型参数估计)是一切控制问题的基础.对于实际控制问题,系统建模和模型参数确定时间往往占整个过程的80%甚至更高,占整个工程工作量的80%以上.系统模型参数确定后,许多现成的控制系统设计、分析和综合方法都可套用,因为实际系统不是追求一个新的控制方法,而是一个比较成熟控制方法的可靠应用.

如果用传递算子 $G(s)$ 表示系统模型, $u(t)$ 表示系统输入, $y(t)$ 表示系统输出,在三元组 $\{u(t), y(t), G(s)\}$ 中,已知其中两个,就可以求另外一个.这个三元组元素有如下关系:

收稿日期 2011-01-11

资助项目 国家自然科学基金(60974043)

作者简介

丁锋,男,博士,教授,博士生导师,主要从事系统辨识、过程建模、自适应控制方面的研究. fdling@jiangnan.edu.cn

$$y(t) = G(s)u(t), \quad u(t) = \frac{y(t)}{G(s)}, \quad G(s) = \frac{y(t)}{u(t)}.$$

第1个方程是控制问题,第2个方程是自校正控制问题(自校正解决跟踪问题,方程中的 $y(t)$ 一般改为期望的轨迹 $y_r(t)$),第3个方程是辨识问题.这只是简单说明了它们的关系,它们各形成了一门学科,有其各自的研究目标和研究内容.

有了这个三元组,就可以给控制理论、系统辨识、状态估计下一个简单的定义.

1) 控制理论(Control Theory)是假设系统模型 $G(s)$ 已知,研究一定输入下系统输出的行为特性,如系统的动态品质(上升时间、超调量、调节时间等)、稳定性等,并通过反馈来改善系统的动态品质和稳定性能(即改善系统的闭环传递函数).也就是说:控制理论是已知输入 $u(t)$ 和传递函数 $G(s)$ 求系统输出 $y(t)$ 的过程.

2) 系统辨识(System Identification)是假设系统输入 $u(t)$ 和输出 $y(t)$ 已知,求系统的传递函数 $G(s)$,即通过采集系统的输入输出数据,研究确定系统数学模型的理论和方法.

一个连续时间动态系统一般可用高阶微分方程来描述(离散系统用差分方程来描述),通过引进一些中间变量(状态变量) $x_i(t) \in \mathbf{R}$,可以化为一些一阶微分方程(差分方程),这个表示系统输入输出关系的方程组称为状态空间模型.单输入单输出连续时间线性时不变系统的状态空间模型如下:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \\ \mathbf{y}(t) = \mathbf{c}\mathbf{x}(t) + \mathbf{d}u(t). \end{cases}$$

其中: $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T \in \mathbf{R}^n$ 为状态向量; $u(t) \in \mathbf{R}$ 和 $y(t) \in \mathbf{R}$ 分别为系统的输入和输出变量; $\mathbf{A} \in \mathbf{R}^{n \times n}$, $\mathbf{b} \in \mathbf{R}^n$, $\mathbf{c} \in \mathbf{R}^{1 \times n}$ 和 $\mathbf{d} \in \mathbf{R}^1$ 是系统参数(矩阵).

3) 状态估计(State Estimation)是假设系统模型参数已知,利用系统的输入输出估计系统的状态,即假设系统参数 $[\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{d}]$ 和输入输出 $\{u(t), y(t)\}$ 已知,估计系统的状态向量 $\mathbf{x}(t)$.

辨识广泛应用于许多科学和工程领域,如化工过程、电力系统、航空航天、社会经济系统、生物生态系统等.凡是需要应用观测数据确定数学模型和估计参数的问题都要利用辨识技术.自适应控制系统是辨识与控制结合的一个典范,形成了自校正控制;辨识与预测控制结合形成自校正预测控制;辨识与鲁棒控制结合形成自校正鲁棒控制;辨识与 H_∞ 控制结合形成自校正 H_∞ 控制等.

系统辨识包括系统结构和阶次的识别、系统参数的估计,以及估计方法的收敛性等.本文简单介绍辨识的定义、数学模型与辨识模型、辨识步骤和辨识方法的类别.在后继的系列文章中,将依次介绍有关辨识精度与辨识的基本问题、辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识原理,以及相应的辨识方法.

本文较长,为了便于阅读,特将本文框架结构列示如下.

0 引言

1 辨识的定义

1.1 系统结构和参数

1.2 机理辨识方法或机理建模方法

1.3 统计辨识方法或统计建模方法

1.4 阶跃响应辨识方法

1.5 辨识的定义与辨识的4要素

1.6 一些学者的辨识定义

1.7 关于测量误差问题

2 数学模型与辨识模型

2.1 数学模型

2.2 辨识模型

3 辨识步骤与辨识目的

3.1 辨识的基本步骤

3.2 辨识的目的

3.3 实验设计

3.4 数据预处理

4 辨识方法的类别

4.1 最小二乘辨识方法

4.2 梯度辨识方法

4.3 辅助模型辨识方法

4.4 多新息辨识方法

4.5 递阶辨识方法

5 结语

1 辨识的定义

1.1 系统结构和参数

通过对控制理论方面课程的学习,我们对“系统”和“控制系统”的概念有了比较深入的了解.那么什么是“系统辨识”呢?从字面上讲,系统辨识(System Identification)就是识别一个系统、辨识一个系统.系统通常是由表征系统输入输出关系的数学模型描述的,这个模型有其特定的结构和参数.因此,系统辨识包含系统结构辨识(System Structure

Identification) 和参数估计 (Parameter Estimation).

系统结构(或模型结构)就是系统数学表达式的形式. 对单输入单输出线性系统而言,模型结构就是系统的阶次(Order);对多变量线性系统而言,模型结构就是系统的能控性结构指数(Controllability Structure Index)或能观测性结构指数(Observability Structure Index),系统阶次等于系统的能控性结构指数或能观测性结构指数之和. 对传递函数而言,系统参数就是传递函数分子分母多项式的系数(Coefficient),系统阶次就是传递函数分母多项式的次数(Degree);对状态空间模型而言,系统参数就是状态空间模型的 A, B, C, D 矩阵,系统阶次就是状态向量的维数或矩阵的维数,它等于系统的能控性结构指数或能观测性结构指数之和.

1.2 机理辨识方法或机理建模方法

有些系统可以用机理方法(如物理学定理)推导出其数学模型结构形式,余下的问题就是识别系统参数. 例如,对于 $R-L-C$ 电路,如图 2 所示,输入为电源电压 $u(t)$,输出是电容 C 上电压 $y(t)$,这个系统的传递函数为

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\frac{1}{sC}}{R + sL + \frac{1}{sC}} = \frac{\frac{1}{LC}}{s^2 + \frac{R}{L}s + \frac{1}{LC}},$$

系统的阶次为 $n = 2$,系统的参数为 $\frac{R}{L}$ 和 $\frac{1}{LC}$. 如果可以用仪表来测量这个电路系统的电阻、电感和电容值,就可得到模型的参数 $\frac{R}{L}$ 和 $\frac{1}{LC}$. 这种方法称为机理辨识方法或机理建模方法. 机理方法就是运用物理学定理(电流电压定理)、能量质量平衡定律、热力学定律等,推导系统模型的解析表达式,获得模型的结构形式(结构参数),采用测量的方法获得系统的参数.

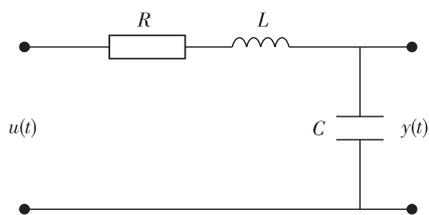


图 2 电阻-电感-电容($R-L-C$)电路

Fig. 2 The resistor, inductor, capacitor circuit

Modeling), 要求对系统的运动规律有清晰的了解,可以利用系统的内部机理、物料和能量的平衡关系、物理学定律等来建立系统的数学模型.

对于一些复杂的系统,要用机理方法细致分析系统的内部行为特征是极其困难的,甚至是不可能的. 因此,机理建模方法有其局限性,因而提出了统计辨识方法,它仅仅根据系统的输入输出数据建立系统的数学模型.

1.3 统计辨识方法或统计建模方法

尽管可以用仪表来测量 $R-L-C$ 电路系统的电阻、电感和电容,然而在系统运行情况下,不允许断电进行测量,如何辨识系统结构和参数,正是系统辨识要回答的问题. 系统辨识解决的问题之一,就是系统在正常运行情况下(稳定工作点),在系统输入端加载一个不影响系统运行的输入信号(持续激励信号),通过量测系统的输入和输出数据,利用大量的观测数据,研究辨识方法来识别系统的结构和参数. 这就是通常所说的统计辨识方法或统计建模方法.

统计建模方法是基于实验的辨识方法,也称为实验建模方法或黑箱建模方法(Black Box Modeling). 所谓黑箱是指系统内部行为对建模者来说是未知的,只能根据外部的系统输入和输出数据序列,确定系统行为的数学模型.

还有一类灰箱建模方法(Gray Box Modeling),就是白箱与黑箱相结合的建模方法. 实际中,如果知道系统的运行规律,就用机理方法推导描述系统行为的数学表达式,然后用实验的方法估计模型的参数.

值得指出的是,虽然有的系统可以根据反应机理推导系统的模型,但这种模型可能是分布参数的偏微分方程、高度非线性,不利于系统分析和综合,不利于用线性控制理论方法设计控制器,对这样的系统我们也采用统计实验方法建立其数学模型.

1.4 阶跃响应辨识方法

对于上述电路系统例子,系统阶次 n 已知,辨识问题就是用系统输入输出数据 $u(t)$ 和 $y(t)$ 辨识和估计这个系统的参数 $\frac{R}{L}$ 和 $\frac{1}{LC}$. 当然,可以给系统施加一个单位阶跃激励(输入)信号

$$u(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0 \end{cases}$$

测量系统的输出 $y(t)$. 对输入和输出进行拉普拉斯变换(Laplace Transform),得到 $U(s) = L[u(t)]$ 和 Y

机理建模方法又称为白箱建模方法(White Box

$(s) = L[y(t)]$, 求它们之比, 就可得到系统的传递函数(阶次和参数), 这是经典阶跃响应辨识方法确定系统的传递函数(还有脉冲响应辨识方法). 可以用计算机产生一个具有解析表达式的阶跃函数作为输入信号, 但是实际中难以得到(不可能得到)输出 $y(t)$ 的解析式(尽管可以使用函数记录仪记录输出曲线). 因此, 这是一种理论方法, 实际中不可行. 解决这一问题就是统计辨识方法: 利用系统的离散输入输出数据序列 $\{u(t), y(t) : t = 0, 1, 2, \dots\}$ 确定或估计系统参数, 就是辨识要研究的基本问题.

1.5 辨识的定义与辨识的 4 要素

辨识就是从系统的观测数据中, 确定一个与系统特性“等价”的模型. 这个定义包括没有输入时的时间序列模型的辨识在内. 模型只能描述系统的主要特征. 如果观测数据包含系统的输入和输出, 那么系统辨识就是拟和一个与系统输入—输出外特性最接近的数学模型. “与系统外特性最接近”可以理解为在同一输入下, 模型的输出 $\hat{y}(t) = Gu(t)$ 最接近系统的真实输出 $y(t)$, 如图 3 所示(G 表示输入输出传递关系).

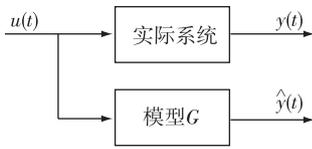


图 3 辨识原理

Fig. 3 The identification principle

这种接近可以用下列误差准则函数 (Error Criterion Function) 来描述:

$$\text{连续系统 } J_1 = \int_0^{\infty} [y(t) - \hat{y}(t)]^2 dt =$$

$$\int_0^{\infty} [y(t) - Gu(t)]^2 dt,$$

$$\text{或离散系统 } J_1 = \sum_{t=0}^{\infty} [y(t) - \hat{y}(t)]^2 =$$

$$\sum_{t=0}^{\infty} [y(t) - Gu(t)]^2.$$

这只是一形式准则函数. 准则函数在有的资料上也称为损失函数 (Cost Function). 如果 $J_1 = 0$, 那么对任意的 t , 模型输出完全等于系统输出 $\hat{y}(t) = Gu(t) = y(t)$, 这是一种理想情况. 实际中, 对任意的 t , 这个等式是不可能成立的, 因为系统总是受到各种各样的干扰, 测量总是包含误差. 因此, 总是希望准则函数 J_1 越小越好, 也就是说误差 $y(t) - Gu(t)$

越小越好, 这个误差具有随机性, 简称为噪声 (Noise), 用 $v(t)$ 表示, 有

$$v(t) = y(t) - Gu(t),$$

或

$$y(t) = Gu(t) + v(t).$$

这就是随机系统产生的缘由. 这个准则是极小化系统输出与模型输出之差, 所以这个模型称为输出误差模型, 可以绘制成图 4 所示. 其中 $u(t)$ 和 $y(t)$ 为系统的输入和输出, $v(t)$ 为随机干扰噪声, G 为要辨识的等价模型.

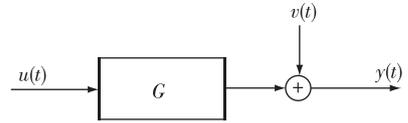


图 4 待辨识系统

Fig. 4 The system to be identified

辨识可以定义为从给定的输入输出数据中, 通过极小化某一误差准则函数, 确立系统的模型. 因为模型有多种形式, 如连续系统传递函数、离散系统差分方程、状态空间模型等, 准则函数的选取依赖于模型形式, 即模型类别. 输入输出数据也是有要求的, 就是期望系统的输入激励信号能够激发出系统的本质特征, 使得系统的输出包含足够的信息来确定系统的数学模型, 这样的输入信号称为持续激励信号. 当然, 不同的辨识方法对输入信号的要求也是不一样的.

准则函数采用不同的优化方法 (如梯度方法、牛顿方法、最小二乘方法等), 可导致不同的辨识方法. 因此, 本文总结出辨识的 4 要素: 输入输出数据、模型、准则函数、优化方法, 简称为数据、模型、准则和优化方法.

1.6 一些学者的辨识定义

许多控制理论专家给系统辨识下过定义, 这些定义各有千秋, 下面介绍 Zadeh^[5] 与 Ljung^[6] 的定义.

美国学者扎德 (L. A. Zadeh) 是模糊系统专家 (University of California, Berkeley, USA), 他于 1962 年在其论文“From circuit theory to system theory (从电路理论到系统理论)”中曾对“系统辨识”给出了定义: 系统辨识是在输入和输出数据基础上, 在指定的一类模型中, 确定一个与被识别系统等价的模型 (System identification: Determination, on the basis of observation of input and output, of a system within a specified class of systems to which the system under test

is equivalent)^[5]. 这个定义明确了辨识的 3 个要素, 即输入输出数据、模型类和等价原则. 扎德的定义极为苛刻: 确定一个与被识别系统的完全等价的模型无疑是极为困难的, 甚至是不可能的.

瑞典学者荣(L. Ljung)是国际上著名辨识专家之一(Linköping University, Linköping, Sweden). 1978 年他给系统辨识下了更为实用的定义, 即系统辨识有 3 要素: 数据、模型类和准则. 系统辨识是按照一个准则, 在模型类中选择一个与数据拟合得最好的模型(The identification procedure is based on three entities: The data, the set of models, and the criterion. Identification, then, is to select that model in the model set that describes the data best, according to the criterion)^[6].

从上述系统辨识定义中可以了解到: 3 个要素中, 数据是辨识的基础; 准则是辨识的优化目标; 模型类是所寻找的模型的范围^[7]. 他们的定义中强调了数据、模型和准则 3 要素. 但是优化手段也很重要, 它是提出新的辨识方法的基础. 新的辨识方法的提出, 以及辨识方法的收敛性分析是系统辨识的主要任务. 从这个层面讲, 辨识方法的优化手段也极其重要, 这就是本文提出辨识的 4 要素的缘由. 基于 4 要素的定义: 系统辨识是通过设计适当的输入信号, 利用实验的输入输出数据, 选择一类模型, 构造一误差准则函数, 用优化方法确定一个与数据拟合得最好的一个模型.

系统辨识是设计输入信号, 通过实验数据建立系统数学模型的理论和方法. 大千世界每一事物都

有其运动规律, 不同学科领域的研究对象, 将其运动规律用方程描述, 就是数学模型. 从这个层面讲, 不同学科的发展过程就是建立它的数学模型的过程, 物理学更是如此, 如万有引力定律. 因此, 系统建模和系统辨识在各个学科中都有广泛的应用.

1.7 关于测量误差问题

学了系统辨识, 我们就走进了一个很不精确的世界: 除了自然数外, 几乎所有测量都是不准确的, 是有误差的, 可以说一切都是测不准的. 测量的精度取决于使用仪器的精度和方法, 连我们最熟悉的光速 30×10^4 km/s 也是不精确的. 创新需要大胆的假设. 爱因斯坦假设光速是绝对的, 即在一个相对的世界里, 假设一个绝对的光速. 正是这一大胆的假设, 便产生了相对论. 在他的相对论中, 大于光速就走进一个虚幻的世界(实际上, 超过光速的射线是存在的). 近代测量真空中光速的实际测量值如表 1 所示(http://baike.baidu.com/view/18638.html?tp=3_01). 随着科学技术的进步, 测量值会不断逼近真值, 精度越来越高, 即误差越来越小. 这就是为什么在辨识模型中, 要引入随机干扰噪声的原因.

2 数学模型与辨识模型

由相互作用又相互依赖的诸要素组成的具有相对独立的整体可以认为是一个系统. 通常我们把所研究的对象看成一个系统. 为了能有效地、定量地研究系统的性能, 从而进行系统分析、预报等, 数学模

表 1 光速测量值年代表

Table 1 The speed of light

年代	测量者	测量方式	光速/(km/s)	不确定度/(km/s)
1907	Rosa, Dorsey	ESU/EMU ^①	299 784	15
1928	Karolus 等	克尔盒	299 786	15
1947	Essen 等	谐振腔	299 792	4
1949	Aslakson	雷达	299 792. 4	2. 4
1951	Bergstand	光电测距仪	299 793. 1	0. 26
1954	Froome	微波干涉仪	299 792. 75	0. 3
1964	Rank 等	带光谱	299 792. 8	0. 4
1972	Bay 等	稳频氦氖雷射器	299 792. 462	0. 018
1973		平差	299 792. 458 0	0. 001 2
1974	Blaney	稳频 CO ₂ 雷射器	299 792. 459 0	0. 000 6
1976	Woods 等		299 792. 458 8	0. 000 2
1980	Baird 等	稳频氦氖雷射器	299 792. 458 1	0. 001 9
1983	国际协议	(规定)	299 792. 458	(精确值)

注: ①ESU 即 Electrostatic Units 的缩写, EMU 为 Electromagnetic Units 的缩写.

型的引入是相当重要的. 模型类别有多种, 如直觉模型、概念模型、结构模型、层次模型、物理模型、图表模型、数学模型、辨识模型等. 系统辨识主要考虑描述系统运动规律的数学模型.

2.1 数学模型

数学模型是描述系统内在规律用的数学表达式, 反映实际系统的行为特性. 控制系统数学模型是指能用来描述系统的输入—输出关系的动态和或静态特性和行为的数学表达式或方程. 同一个研究对象的数学模型不是唯一的, 如线性系统的传递函数模型、状态空间模型等. 按模型属性划分, 数学模型有下列类别.

2.1.1 静态模型和动态模型 (Static Model and Dynamic Model)

模型有静态和动态之分, 用代数方程表示的模型称为静态模型, 如线性弹簧所受的力 F 与弹簧伸长(可正可负) x 和弹性系数 k 的关系模型为 $F = kx$; 用微分方程或差分方程表示的模型称为动态模型, 对应的系统称为动态系统.

2.1.2 参数模型和非参数模型 (Parameter Model and Non-Parametric Model)

具有揭示系统输入输出关系的解析表达式的模型称为参数模型, 模型辨识就是确定这个解析模型的参数, 如传递函数模型、状态空间模型、差分方程模型、微分方程模型等. 难以用参数模型描述的称为非参数模型, 如图表曲线和数据模型、阶跃响应、脉冲响应、频率响应曲线等.

2.1.3 线性模型和非线性模型 (Linear Model and Nonlinear Model)

线性模型用于描述线性系统, 非线性模型用于描述非线性系统. 输入—输出关系满足叠加原理的系统称为线性系统(在忽略系统中的随机干扰噪声时), 否则称为非线性系统. 实际系统都是非线性系统, 非线性系统的控制方法比较复杂, 而线性系统理论研究比较成熟, 因而经常用线性化的方法, 把有些非线性系统近似为线性系统来处理.

2.1.4 线性参数模型和非线性参数模型 (Linear Parameter Model and Nonlinear Parameter Model)

如果系统输出(或经过变化后系统的输出)可以表示为参数的线性函数, 就称为线性参数模型 (Linear Parameter Model), 对应的系统是线性参数系统 (Linear Parameter System), 也称为本质线性模型;

否则称为非线性参数模型(本质非线性模型). 有些非线性系统经过参数化可以化为线性参数系统. 如, 非线性系统

$$y(t) = a_1 y(t-1) + a_2 y(t-2) y(t-3) + b_1 u(t)$$

是一个线性参数系统, 因为输出 $y(t)$ 是参数 (a_1, a_2, b_1) 的线性函数. 又如, 非线性系统

$$y(t) = \frac{b}{u^2(t) + a}$$

是一个本质线性系统(线性参数系统), 因为这个模型等价于

$$\frac{1}{y(t)} = \frac{u^2(t)}{b} + \frac{a}{b},$$

如果把 $\frac{1}{y(t)}$ 看作新系统的输出, 它是参数 $\theta_1 = \frac{1}{b}$ 和 $\theta_2 = \frac{a}{b}$ 的线性函数. 当辨识出 θ_1 和 θ_2 , 就可以求出原系统的参数 a 和 b .

同样, 非线性回归问题(忽略误差)

$$y = a^{bx}$$

通过两边取对数, 可以化为一个线性回归问题

$$\ln y = \ln a + bx.$$

尽管不是所有非线性系统都可以转化为线性参数模型, 但有些复杂非线性系统可以转化为简单的非线性模型, 这个简单的非线性模型可能更容易求解.

2.1.5 连续模型和离散模型 (Continuous Model and Discrete Model)

系统的运动规律是随时间演化的, 如果模型中时间是一个连续变量, 就称为连续(时间)模型, 对应的系统称为连续时间系统 (Continuous-Time System), 简称连续系统; 如果模型中时间取离散值, 就称为离散(时间)模型, 对应的系统称为离散时间系统 (Discrete-Time System), 简称离散系统.

2.1.6 确定性模型和随机模型 (Deterministic Model and Stochastic Model)

如果模型中的变量都是确定性的(即模型中不含有随机干扰), 就称为确定性模型; 如果模型中含有随机噪声项, 就称为随机模型(这里是指系统输入和(或)输出变量中含有随机噪声, 系统参数是确定性的). 如

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_1 u(t-1) + b_2 u(t-2) \quad (1)$$

是一个确定性模型. 因为其参数 a_i, b_i 不随时间变化, 也是一个线性时不变确定性系统. 假设 $v(t)$ 是一

个随机噪声,那么

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_1 u(t-1) + b_2 u(t-2) + v(t) \quad (2)$$

就是一个随机模型(线性时不变随机系统). 系统参数是随机变化的,也是一类随机系统. 一般的辨识著作很少深入讨论这类随机系统.

2.1.7 时不变参数模型和时变参数模型(Time-Invariant Parameter Model and Time-Varying Parameter Model)

如果模型参数不随时间变化,就是时不变参数模型,对应的系统称为时不变系统. 式(1)和(2)是时不变模型. 如果模型参数随时间变化,就是时变参数模型,对应的系统称为时变系统. 下面两个系统是时变参数模型,

$$y(t) + a_1(t)y(t-1) + a_2(t)y(t-2) = b_1 u(t-1) + b_2 u(t-2), \quad (3)$$

$$y(t) + a_1(t)y(t-1) + a_2(t)y(t-2) = b_1(t)u(t-1) + b_2(t)u(t-2) + v(t), \quad (4)$$

因为其参数 $a_i(t)$ 和(或) $b_i(t)$ 是随时间变化的. 只要有一个或一个以上参数随时间变化的系统,就是时变系统.

我们知道,指数函数的泰勒展开(Taylor Expansion)有很好的收敛性,即对于任意 t , 下列级数收敛, $e^t = 1 + t + \frac{1}{2!}t^2 + \frac{1}{3!}t^3 + \dots + \frac{1}{n!}t^n + \dots, -\infty < t < +\infty$, 且收敛半径是无穷大.

假设一个时变参数是

$$a(t) = 1 - e^{-t},$$

这是一个有界时变参数,因为当时间 t 趋于无穷时, $a(t)$ 收敛于 1, 即 $a(t) \rightarrow 1$.

快速变化时变参数的辨识是极其困难的. 一个方法是把时变参数展开成级数,取其有限项近似,经过对辨识模型重新参数化,就将对时变参数的辨识化为对时不变参数的辨识. 这样做可使辨识问题简单,但导致下列问题.

把 $a(t)$ 展开成级数为

$$a(t) = 1 - e^{-t} = t - \frac{1}{2!}t^2 + \frac{1}{3!}t^3 - \frac{1}{4!}t^4 + \dots + \frac{(-1)^{n-1}}{n!}t^n + \dots, -\infty < t < +\infty,$$

因为这个级数收敛,所以人们理所当然地取其有限项近似(比如 n 项):

$$a(t) \approx t - \frac{1}{2!}t^2 + \frac{1}{3!}t^3 - \frac{1}{4!}t^4 + \dots + \frac{(-1)^{n-1}}{n!}t^n.$$

尽管该级数对任意的 t (很大的 t) 都成立,但上式只对很小的 t 成立,因为 t 是随时间增大的,对任意有限的 n , 上式右边趋于无穷,左边是有界的. 这导致矛盾. 这是用级数展开时变参数进行辨识应注意的问题. 因此,用傅立叶级数(Fourier Series)的有限项来近似时变参数,是一种研究时变参数辨识的途径之一.

2.1.8 频域模型与时域模型(Frequency Domain Model and Time-Domain Model)

2.1.9 集中参数模型和分布参数模型(Lumped Parameter Model and Distributed Parameter Model)

2.1.10 微观模型与宏观模型(Micro Model and Macro Model)

2.1.11 不确定性模型(Uncertainty Model)

2.1.12 混杂模型(Hybrid Model)

2.1.13 模糊模型(Fuzzy Model)

按照上述分类,神经网络模型不能成为一类模型,它是以复合函数形式表示输入输出关系,就像我们没有列出抛物线模型、多项式模型、差分方程模型一样.

2.2 辨识模型

为了辨识系统参数,一般把系统的参数写成一个参数向量形式,输入输出写成一个信息向量形式,就得到系统的辨识模型(Identification Model),也称为辨识表达式(Identification Representation). 一般单输入单输出线性系统的辨识模型可以表达为

$$y(t) = \varphi^T(t)\theta + v(t). \quad (5)$$

其中: $y(t) \in \mathbf{R}$ 为系统输出; $v(t) \in \mathbf{R}$ 是随机干扰噪声; $\varphi(t) = [\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t)]^T \in \mathbf{R}^n$ 是由系统输入输出数据构成的信息向量; $\theta = [\theta_1, \theta_2, \dots, \theta_n]^T \in \mathbf{R}^n$ 是待辨识系统的参数向量,上标 T 表示向量或矩阵转置, n 是参数数目.

式(5)描述的辨识模型示意如图 5 所示,它把输出表示为参数的线性组合形式,是辨识问题的一个基本表达形式,称为最小二乘格式(Least Squares Form),因为这个最小二乘格式辨识模型的参数向量易于用最小二乘方法辨识. 一般线性系统和本质线性系统(线性参数模型)都可以化成这种格式.

在这个最小二乘辨识模型中, $y(t)$ 和 $\varphi(t)$ 是观测量. 如果 $y(t)$ 和 $\varphi(t)$ 都可测,则最小二乘算法能够估计参数向量 θ ; 如果信息向量 $\varphi(t)$ 里有部分项不可测,就需要用增广最小二乘算法、辅助模型最小二乘算法或其他辨识算法进行求解.



图5 辨识模型示意

Fig. 5 The identification model

如果参数是随时间 t 变化的, 参数向量就写为 $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \dots, \theta_n(t)]^T \in \mathbf{R}^n$, 时变参数系统辨识模型为

$$y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta}(t) + v(t). \quad (6)$$

下面给出一些系统的辨识模型.

2.2.1 线性系统辨识模型

1) 线性系统脉冲响应模型

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + \dots + b_n u(t-n) + v(t) =$$

$$[u(t-1), u(t-2), \dots, u(t-n)] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + v(t)$$

对应的辨识模型为

$$\begin{cases} y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + v(t), \\ \boldsymbol{\varphi}(t) = [u(t-1), u(t-2), \dots, u(t-n)]^T \in \mathbf{R}^n, \\ \boldsymbol{\theta} = [b_1, b_2, \dots, b_n]^T \in \mathbf{R}^n. \end{cases}$$

2) 时变系数差分方程描述的线性系统

$$y(t) + a_1(t)y(t-1) + a_2 y(t-2) = b_1 u(t-1) + b_2 u(t-2)$$

对应的辨识模型为

$$\begin{cases} y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta}(t), \\ \boldsymbol{\varphi}(t) = [-y(t-1), -y(t-2), u(t-1), \\ u(t-2)]^T \in \mathbf{R}^4, \\ \boldsymbol{\theta}(t) = [a_1(t), a_2, b_1, b_2]^T \in \mathbf{R}^4. \end{cases}$$

3) 线性动态系统(CAR 模型)

$$A(z)y(t) = B(z)u(t) + v(t), \quad (7)$$

其中 z^{-1} 为单位后移算子 [$z^{-1}y(t) = y(t-1)$, $zy(t) = y(t+1)$], $A(z)$ 和 $B(z)$ 是 z^{-1} 的常系数时不变多项式, 定义如下:

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a};$$

$$B(z) := b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b}.$$

将多项式 $A(z)$ 和 $B(z)$ 的表达式代入式(7)

可得

$$(1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a})y(t) = (b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b})u(t) + v(t),$$

利用移位算子的性质, 上式可以写为

$$y(t) + a_1 y(t-1) + a_2 y(t-2) + \dots + a_{n_a} y(t-n_a) = b_1 u(t-1) + b_2 u(t-2) + \dots + b_{n_b} u(t-n_b) + v(t),$$

移项得到

$$y(t) = -a_1 y(t-1) - a_2 y(t-2) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + b_2 u(t-2) + \dots + b_{n_b} u(t-n_b) + v(t).$$

对应的辨识模型为

$$\begin{cases} y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + v(t), \\ \boldsymbol{\varphi}(t) = [-y(t-1), -y(t-2), \dots, \\ -y(t-n_a), u(t-1), u(t-2), \dots, \\ u(t-n_b)]^T \in \mathbf{R}^n, \\ \boldsymbol{\theta} = [a_1, a_2, \dots, a_{n_a}, b_1, b_2, \dots, b_{n_b}]^T \in \mathbf{R}^n, \\ n := n_a + n_b. \end{cases}$$

2.2.2 非线性系统辨识模型

1) 静态非线性系统

非线性系统(线性参数系统)

$$y = ax^2 + bx + c$$

可以写为

$$y = [x^2, x, 1] \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

对应的辨识模型为

$$\begin{cases} y = \boldsymbol{\varphi}^T \boldsymbol{\theta}, \\ \boldsymbol{\varphi} = [x^2, x, 1]^T \in \mathbf{R}^3, \\ \boldsymbol{\theta} = [a, b, c]^T \in \mathbf{R}^3. \end{cases}$$

当量测值为 (x_i, y_i) , $i = 1, 2, \dots, L$ 时, 得到方程

$$\begin{cases} y_i = \boldsymbol{\varphi}_i^T \boldsymbol{\theta}, \\ \boldsymbol{\varphi}_i = [x_i^2, x_i, 1]^T \in \mathbf{R}^3. \end{cases}$$

又考虑非线性时变系统

$$y(t) = a(t)x^2(t) + bx(t) + c + de^t$$

可以写为

$$y(t) = [x^2(t), x(t), 1, e^t] \begin{bmatrix} a(t) \\ b \\ c \\ d \end{bmatrix}.$$

对应的辨识模型为

$$\begin{cases} y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta}(t), \\ \boldsymbol{\varphi}(t) = [x^2(t), x(t), 1, e^t]^T \in \mathbf{R}^4, \\ \boldsymbol{\theta}(t) = [a(t), b, c, d]^T \in \mathbf{R}^4. \end{cases}$$

2) 动态非线性系统

$$y(t) = a_1 y(t-1) + a_2 y(t-2)y(t-3) + b_1 u(t) + b_2 u^2(t-1) + c \cos(\sqrt{t}) + v(t) = [y(t-1), y(t-2)y(t-3), u(t), u^2(t-1), \cos(\sqrt{t})]^T \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ c \end{bmatrix} + v(t)$$

对应的辨识模型为

$$\begin{cases} y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + v(t), \\ \boldsymbol{\varphi}(t) = [y(t-1), y(t-2)y(t-3), u(t), u^2(t-1), \cos(\sqrt{t})]^T \in \mathbf{R}^5, \\ \boldsymbol{\theta} = [a_1, a_2, b_1, b_2, c]^T \in \mathbf{R}^5. \end{cases}$$

3) 非线性系统

$$y(t) = \frac{b[e^t + u^2(t-3)]}{u^2(t) + a_1 u(t-1) + a_2 u(t-2)}.$$

通过定义新的输出 $y_1(t) := 1/y(t)$, 新的参数 $\theta_1 := 1/b, \theta_2 := a_1/b, \theta_3 := a_2/b$, 那么这个模型可变化为线性参数系统:

$$y_1(t) = \frac{u^2(t) + a_1 u(t-1) + a_2 u(t-2)}{b[e^t + u^2(t-3)]} = \frac{1}{b} \frac{u^2(t)}{e^t + u^2(t-3)} + \frac{a_1}{b} \frac{u(t-1)}{e^t + u^2(t-3)} + \frac{a_2}{b} \frac{u(t-2)}{e^t + u^2(t-3)} = \left[\frac{u^2(t)}{e^t + u^2(t-3)}, \frac{u(t-1)}{e^t + u^2(t-3)}, \frac{u(t-2)}{e^t + u^2(t-3)} \right] \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}.$$

对应的辨识模型为

$$\begin{cases} y_1(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}, \\ \boldsymbol{\varphi}(t) = \left[\frac{u^2(t)}{e^t + u^2(t-3)}, \frac{u(t-1)}{e^t + u^2(t-3)}, \frac{u(t-2)}{e^t + u^2(t-3)} \right]^T \in \mathbf{R}^3, \\ \boldsymbol{\theta} = \left[\frac{1}{b}, \frac{a_1}{b}, \frac{a_2}{b} \right]^T \in \mathbf{R}^3. \end{cases}$$

变化后的系统是一个线性参数系统, 当辨识出 θ_1, θ_2 和 θ_3 , 就可以求出原系统的参数:

$$b = \frac{1}{\theta_1}, \quad a_1 = \frac{\theta_2}{\theta_1}, \quad a_2 = \frac{\theta_3}{\theta_1}.$$

4) 状态方程

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) + \mathbf{w}(t), \\ y(t) = \mathbf{c}\mathbf{x}(t) + \mathbf{d}u(t) + v(t). \end{cases} \quad (8)$$

其中: $u(t) \in \mathbf{R}$ 和 $y(t) \in \mathbf{R}$ 分别为系统的输入和输出变量; $\mathbf{x}(t) \in \mathbf{R}^n$ 为状态向量; $\mathbf{A} \in \mathbf{R}^{n \times n}$, $\mathbf{b} \in \mathbf{R}^n$, $\mathbf{c} \in \mathbf{R}^{1 \times n}$ 和 $\mathbf{d} \in \mathbf{R}$ 为常数矩阵; $\mathbf{w}(t) \in \mathbf{R}^n$ 为过程噪声向量; $v(t) \in \mathbf{R}$ 为观测噪声项. 定义

$$\mathbf{Y}(t) := \begin{bmatrix} \mathbf{x}(t+1)y(t) \\ y(t) \end{bmatrix} \in \mathbf{R}^{n+1},$$

$$\boldsymbol{\varphi}(t) := \begin{bmatrix} \mathbf{x}(t) \\ u(t) \end{bmatrix} \in \mathbf{R}^{n+1},$$

$$\boldsymbol{\theta}^T := \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{bmatrix} \in \mathbf{R}^{(n+1) \times (n+1)},$$

$$\mathbf{e}(t) := \begin{bmatrix} \mathbf{w}(t) \\ v(t) \end{bmatrix} \in \mathbf{R}^{n+1},$$

则状态空间模型(8)对应的辨识模型为

$$\mathbf{Y}(t) = \boldsymbol{\theta}^T \boldsymbol{\varphi}(t) + \mathbf{e}(t).$$

这是一个线性参数系统. 利用移位算子的性质, 由式(8)的第1个方程可得

$$z\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) + \mathbf{w}(t),$$

$$\mathbf{x}(t) = (z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}u(t) + (z\mathbf{I} - \mathbf{A})^{-1} \mathbf{w}(t).$$

将上式代入(8)的第2式可得

$$y(t) = [\mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} + \mathbf{d}]u(t) + \mathbf{c}(z\mathbf{I} - \mathbf{A})^{-1} \mathbf{w}(t) + v(t).$$

这里 $y(t)$ 是参数 $\mathbf{A}, \mathbf{b}, \mathbf{c}$ 和 \mathbf{d} 的高度非线性参数系统. 因此, 经过模型变化, 线性参数系统可以变为非线性参数系统.

3 辨识步骤与辨识目的

系统辨识包括结构辨识和参数估计. 实际中两个问题密切相关, 比如一个传递函数的阶次和参数, 阶次高参数就多. 在理论研究中, 经常把两个问题分开来研究, 也有的方法同时辨识系统阶次和参数. 本节介绍辨识的基本步骤与辨识目的.

3.1 辨识的基本步骤

辨识的基本步骤如图6所示, 包括实验设计、数据采集、数据预处理、结构辨识(阶次辨识)、参数估计、模型验证等.

3.1.1 实验设计(Experimental Design)

辨识是从系统的输入输出数据提取数学模型的过程. 设计实验就是使所得到的数据能包含系统的充分多的信息. 实验设计的主要内容包括输入信号的选择、采样周期等. 如果有了系统的先验知识, 知道系统运动规律的一些信息, 将有利于设计合理的

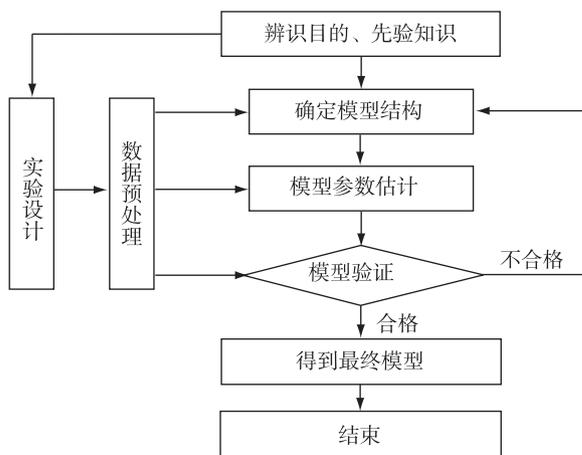


图6 辨识的基本步骤

Fig. 6 The basic steps of identification

实验(如系统的快慢性可以帮助选择输入信号频带和输入输出信号的采样周期). 系统先验知识可以向实际工程师了解, 可以通过多次反复实验, 摸索系统的运动规律, 进一步设计一个合理的实验.

设计实验对系统辨识是极其重要的. 系统日常的运行数据可能并不包含系统的本质特征, 用于辨识往往达不到预期效果, 原因在于系统稳态运行, 输入信号幅值变化不够大, 也不满足持续激励信号要求, 造成噪信比太大、信息被噪声淹没等.

3.1.2 数据预处理(Data Preprocessing)

用于辨识的数据需要经过数据处理, 因这个处理在辨识之前, 故称为数据预处理. 输入输出数据通常都含有直流成分或高频成分, 或由于数据采用了不同量纲, 可能导致错误的辨识结果. 因此, 为使所辨识的模型不受这些因素的影响, 必须对数据进行预处理(如数据零均值化、数据滤波等), 以提高辨识的精度和辨识模型的可用性.

假定设计一个实验研究气态常数 R (辨识气态常数 R). 根据气态方程

$$\frac{PV}{T} = nR,$$

其中 P 为压强 (Pa), V 为体积 (立方米; m^3), n 为物质的量, R 为气体常数, T 为绝对温度 (K). 绝对温度等于摄氏温度加上 273.15 (如摄氏 10°C 即为 283.15 K).

在单位压力 ($P = 1$)、单位质量 ($n = 1$) 时, 气体体积 V 与温度 T 有线性关系:

$$V = RT.$$

如果我们使用摄氏温度作为 T 的单位, 测量不

同温度 T_i 下气体体积 $V_i, i = 1, 2, \dots, m$, 根据最小二乘方法, 极小化误差平方和准则函数

$$J(R) = (V_1 - RT_1)^2 + (V_2 - RT_2)^2 + \dots + (V_m - RT_m)^2,$$

令其对 R 的导数为零:

$$\frac{dJ(R)}{dR} = -2T_1(V_1 - RT_1) - 2T_2(V_2 - RT_2) - \dots -$$

$$2T_m(V_m - RT_m) = 0,$$

求解可得到 R 的估计:

$$\hat{R} = \frac{T_1 V_1 + T_2 V_2 + \dots + T_m V_m}{T_1^2 + T_2^2 + \dots + T_m^2}.$$

尽管这个计算似乎很完美, 但计算结果却是错误的, 因为在热力学中, T 的单位应该采用绝对温度. 这个辨识错误, 可以采用数据零均值化或在线性模型方程中加一个常数项来纠正, 参见后面的方程常数方法.

3.1.3 结构辨识 (Structure Identification) 或阶次辨识 (Order Identification)

模型结构辨识依赖于辨识的目的或模型的用途. 如果模型用于预报或者监测过程变量 (软测量: Soft Sensing), 模型阶次就可以选择大一些; 如果模型用于控制, 模型阶次就应适当, 因为阶次太高会给系统设计和分析造成困难.

系统阶次辨识的方法很多. 一般的模型结构主要通过先验知识来得到. 对于离散时间线性系统和线性参数模型, 一种方法是通过分析输入输出数据的相关性确定系统的阶次.

如果知道系统运动规律的解析表达式:

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), u(t), \boldsymbol{\theta}), \\ y(t) = g(\mathbf{x}(t), u(t), \boldsymbol{\theta}), \end{cases}$$

其中 $\mathbf{x}(t)$ 是系统状态向量, $u(t)$ 和 $y(t)$ 是系统的输入和输出, $\boldsymbol{\theta}$ 是系统参数向量, 函数 \mathbf{f} 和 g 的形式就是模型的结构. 那么, 线性参数系统 (或经过变换获得的线性参数系统) 的结构可以通过输入输出数据进行辨识.

3.1.4 参数估计 (Parameter Estimation) 或模型参数辨识 (Model Parameter Identification)

知道模型的结构后, 模型中可能还会有一些参数的值是未知的, 例如前面所列举的动态模型中的参数向量 $\boldsymbol{\theta}$. 用输入输出数据确定模型中的未知参数 $\boldsymbol{\theta}$ 的过程, 就是参数估计. 之所以用“估计”一词, 是因为实验数据总是有误差的. 这种误差具有随机性, 所以参数估计以统计方法为主, 只有使用大量的数据才能得到良好的结果.

参数估计算法按执行的方式可分为一次完成算

法、递推算法和迭代算法 3 类。

1) 一次完成算法是将所有的数据一次进行处理,即根据全部数据计算得到参数的估计,如最小二乘估计。一次完成算法一般用于(干扰为白噪声的)线性参数模型,如方程误差模型、有限脉冲响应模型等。

当辨识模型中 $v(t)$ 是零均值白噪声, $\varphi(t)$ 是可测的, $\varphi(t)$ 与 $v(t)$ 不相关,那么一次完成算法给出的参数估计是无偏估计,也是一致估计,即在信息向量 $\varphi(t)$ 是持续激励的,随着数据增加,参数估计收敛于真参数。

2) 递推算法是随时间推移,采集数据的增加,不断更新参数估计,即下一时刻的参数估计等于前一时刻的参数估计加上修正项。这是一个递推计算过程,故称为递推算法。递推算法的递推变量与时间有关,故适合在线辨识(On-Line Identification)或实时估计(Real-Time Estimation)。递推算法可以用于(存在白噪声干扰或有噪声干扰的)线性参数模型(如方程误差模型和 CARMA 模型等),也可用于(白噪声干扰和有噪声干扰的)非线性参数模型。本文中的递推最小二乘算法、随机梯度算法都是递推辨识算法。

对于实际系统,随着时间的推移,采集的输入输出数据 $u(t)$ 和 $y(t)$ 不断增多,可以采用在线的实时算法计算参数估计,这种算法通过递推辨识(即递推参数估计算法)实现。递推算法的每一步计算量比较小(与一次完成算法相比,每一步的计算量比解一个线性代数方程组要少得多),特别适用于估计时变参数,也可以实时估计适应控制器的参数^[8]。

20 世纪 60 年代起,递推估计算法得到迅速发展,随后产生了许多不同的方法,如有限数据窗最小二乘算法、协方差修正算法、卡尔曼滤波算法、随机逼近方法、辅助模型辨识算法、多新息辨识算法、递阶辨识算法等。

3) 迭代算法主要用于非线性参数模型或信息向量中存在未知项的线性模型(如时间序列 MA 模型和 ARMA 模型、有色噪声干扰的方程误差类模型、输出误差类模型等)的参数估计。迭代算法是同时考虑所有量测的数据,定义一个准则函数,采用梯度搜索原理、最小二乘原理或牛顿迭代方法,求解优化问题得到的。随着迭代次数增加更新参数估计,如输出误差模型的最小二乘迭代算法和梯度迭代算法。迭代算法中数据长度是不变的,迭代变量与时间无关,

一般用于离线辨识(Off-Line Identification)。本文中的最小二乘迭代算法、梯度迭代算法等都是迭代算法。

3.1.5 模型验证(Model Validation)

辨识获得的模型必须进行模型验证。模型验证就是检验模型的有效性和适应性。通过参数估计得到的模型,虽然是在某种准则下优化得到的“最好”的模型,但是并不一定能达到建模的目的,所以还必须进行适用性检验,即模型验证。这是辨识过程的重要一环,只有通过适用性检验的模型才是最终的模型。就像一个满足一定条件的函数都可以展开成泰勒级数(Taylor Series),但是泰勒级数是否收敛到原函数,还需要进行收敛性分析以确定级数的收敛域。

模型验证的基本原则:必须承认辨识得到的模型只是实际系统的一个近似,不能期望找到一个和实际系统完全一致的模型。如果模型的输入输出特性与实际系统基本一致,那么就认为模型是满意的。

影响模型效果的因素是多方面的。主要有模型结构选择不当、实验数据误差过大或数据代表性太差、辨识算法存在问题等。

对于一个物理系统,如果模型结构或模型解析式已知,余下的任务就是确定模型中的参数。在这种情况下,目标就是参数估计精度。参数估计精度 δ 可以使用参数估计 $\hat{\theta}$ 与真参数 θ 之相对误差来衡量:

$$\delta = \frac{\|\hat{\theta} - \theta\|}{\|\theta\|} \times 100\%.$$

然而真参数 θ 是要估计的,是未知的,上式不可计算(上式只在仿真中)。

模型验证主要有利用先验知识检验和利用数据检验两类。利用先验知识是模型适用性检验的一条重要途径。比如通过计算和观察获得模型的阶跃响应,有经验的工程师可以大致判断模型的适应性。一些模型从数据的拟合上看不出问题,但是根据对模型已有的知识却可以断定模型是否适用。

1) 数据检验

数据检验有离线数据检验和在线数据检验。通常先进行离线数据检验,基本合格后,再进行在线数据检验。

① 离线数据检验。当采集到一批数据,把数据分为两组,一组数据用于辨识建模(参数估计),另一组用于模型验证。离线数据检验就是在同一输入下,计算模型的输出,与系统真实输出进行比较,通过输出残差的白色性来检验。模型验证必须使用与建模

不同的数据集,因为建模一般是极小化一误差准则函数,用于建模的数据往往有很好的拟合性.

② 在线数据检验.就是将离线检验后的模型接于实际系统,用于输出预测,模型输入就是系统的输入,如果模型输出与系统真实输出很接近,得到实际工程师的认可,就可进行模型检验的最后一步——模型的应用检验.

2) 模型验证的基本方法^[7]

① 利用在不同时间区段内采集的输入输出数据,分别建立系统模型.如果模型的特性(如极点、零点分布,系统增益等)基本相符,则可认为模型是可靠的.

② 利用两组长度不同的数据集,独立辨识出模型,并分别计算它们的准则函数,然后将两组数据交叉使用,再计算各自的准则函数.如果对应的准则函数值没有明显的变化,则可认为模型是可靠的.

③ 增加辨识中使用的数据长度,如果准则函数值除以数据长度不再明显变化,则可认为模型是可靠的.

④ 检验模型与过程输出残差序列的白色性.如果残差序列可以视作零均值白噪声序列,则可认为模型是可靠的.

3.2 辨识的目的

在进行建模时,需要明确模型的用途,即辨识的目的.辨识目的不同对模型的要求也不同.例如:估计一个物理系统的参数(电阻、电容或电感),希望参数估计越接近真参数越好;用于控制的模型希望输出越接近期望输出.下面简单介绍辨识建模的目的.

3.2.1 物理量的估计

如果辨识的目的是为了估计一个物理量(估计模型的参数),则希望估计的参数的精度越高越好,如设计实验测量或估计光在真空中的速度.一些物理系统的模型结构往往已知,辨识建模的目的就是估计模型中参数(物理常数),如前述的设计实验估计气体方程的气态常数 R ,估计万有引力定律

$$F = G \frac{mM}{r^2}$$

中的万有引力常数 G ,其中 m 和 M 是两个物体的质量(kg), r 是两个物体之间的距离(m), F 是两个物体间的引力(N).

牛顿发现了万有引力定律,但引力常量 G 这个数值是多少,连他本人也不知道.按说只要测出两个

物体的质量,测出两个物体间的距离,再测出物体间的引力,代入万有引力定律,就可以计算出这个常量,但因为一般物体的质量太小了,它们间的引力无法测出,而天体的质量太大了,又无法测出质量.万有引力定律发现后的 100 多年里,万有引力常量仍没有一个准确的结果,这个公式就仍然不是一个完善的等式.直到 1798 年英国物理学家卡文迪许利用著名的卡文迪许扭秤(即卡文迪许实验),才巧妙地测出了这个常量.

卡文迪许测定的 G 值为 6.754×10^{-11} ,现在公认的 G 值为 6.67×10^{-11} .需要注意的是:这个引力常量是有单位的,它的单位应该是乘以两个质量的单位(kg),再除以距离的单位(m)的平方后,得到力的单位牛顿(N),故应为 Nm^2/kg^2 ,即

$$G = 6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2.$$

到目前为止,引力常数是物理学常数中最不精确的一个.由于引力常量的数值非常小,所以一般质量的物体之间的万有引力是很小的.我们可以估算一下,质量均为 50 kg 的两人相距 0.5 m 时之间的万有引力大约 6.67×10^{-7} N,这么小的力是根本感觉不到的.只有质量很大的物体对一般物体的引力我们才能感觉到,如地球对我们的引力大致就是我们的重力,月球对海洋的引力导致了潮汐现象.而天体之间的引力由于星球的质量很大,又是非常惊人的,如太阳对地球的引力达 3.56×10^{22} N (<http://baike.baidu.com/view/1108890.htm>).

不知道卡文迪许做了几次实验,如果是一次实验,那误差太大了怎么办?如果是多次实验,那么他又以哪次实验结果为准,是否使用了最小二乘法?当然可以使用多次计算结果取平均值,也可以定义一个准则函数求最小二乘估计.

1) 平均法

最简单的是使用多次计算结果的一个简单平均.把万有引力公式改为

$$G = \frac{F_i r_i^2}{m_i M_i}$$

因此,多次测量(假设 n 次)的万有引力常数的平均值为

$$\bar{G} = \frac{\frac{F_1 r_1^2}{m_1 M_1} + \frac{F_2 r_2^2}{m_2 M_2} + \dots + \frac{F_n r_n^2}{m_n M_n}}{n}$$

这相当于使用下列准则函数

$$J(G) = \left(G - \frac{F_1 r_1^2}{m_1 M_1}\right)^2 + \left(G - \frac{F_2 r_2^2}{m_2 M_2}\right)^2 + \dots + \left(G - \frac{F_n r_n^2}{m_n M_n}\right)^2.$$

2) 最小二乘估计 I

尽管平均值可以消除干扰因素,增加准确性,学了系统辨识就知道,简单的平均不一定是最好的办法.应该使用系统辨识的最小二乘方法.定义误差平方和准则函数

$$J(G) = \left(F_1 - G \frac{m_1 M_1}{r_1^2}\right)^2 + \left(F_2 - G \frac{m_2 M_2}{r_2^2}\right)^2 + \dots + \left(F_n - G \frac{m_n M_n}{r_n^2}\right)^2,$$

令其对 G 导数为零:

$$\frac{dJ(G)}{dG} = -2 \frac{m_1 M_1}{r_1^2} \left(F_1 - G \frac{m_1 M_1}{r_1^2}\right) - 2 \frac{m_2 M_2}{r_2^2} \left(F_2 - G \frac{m_2 M_2}{r_2^2}\right)^2 - \dots - 2 \frac{m_n M_n}{r_n^2} \left(F_n - G \frac{m_n M_n}{r_n^2}\right)^2 = 0.$$

求解可得到万有引力常数 G 的最小二乘估计:

$$\hat{G} = \frac{\frac{m_1 M_1}{r_1^2} F_1 + \frac{m_2 M_2}{r_2^2} F_2 + \dots + \frac{m_n M_n}{r_n^2} F_n}{\left(\frac{m_1 M_1}{r_1^2}\right)^2 + \left(\frac{m_2 M_2}{r_2^2}\right)^2 + \dots + \left(\frac{m_n M_n}{r_n^2}\right)^2}.$$

3) 最小二乘估计 II

按照前述的气态方程的气体常数的估计方法,把万有定律方程改写成下列形式:

$$Fr^2 = GmM.$$

定义误差平方和准则函数

$$J(G) = (F_1 r_1^2 - G m_1 M_1)^2 + (F_2 r_2^2 - G m_2 M_2)^2 + \dots + (F_n r_n^2 - G m_n M_n)^2.$$

令其对 G 导数为零:

$$\frac{dJ(G)}{dG} = -2 m_1 M_1 (F_1 r_1^2 - G m_1 M_1) - 2 m_2 M_2 (F_2 r_2^2 - G m_2 M_2) - \dots - 2 m_n M_n (F_n r_n^2 - G m_n M_n) = 0,$$

求解可得到万有引力常数 G 的最小二乘估计:

$$\hat{G} = \frac{m_1 M_1 F_1 r_1^2 + m_2 M_2 F_2 r_2^2 + \dots + m_n M_n F_n r_n^2}{m_1^2 M_1^2 + m_2^2 M_2^2 + \dots + m_n^2 M_n^2}.$$

上述 3 个估计精度是不一样的,那么就需要分析哪一个精度最高,在什么情况下高?

3.2.2 预测(Prediction)

预测和控制是建模的两个重要应用.根据系统的历史数据建立系统的模型,就可以用系统的可测输入和输出去预测系统输出的未来的演变,例如最常见的气象预报、洪水预报、市场价格的预测等.辨识用于预测的模型的参数数目可以大一些,阶次可

以高一些,准则仍然是预测误差平方和最小.只要预测误差小就是好的预测模型,对模型的结构及参数则很少再有其他限制(此段参阅 <http://baike.baidu.com/view/320825.htm>).

3.2.3 控制(Control)

如果建模的目的是为了分析系统的运动规律,设计控制器对系统实行有效的控制,那么模型就不应该太复杂,在满足精度要求下模型阶次不应该太高,否则会给系统控制器设计带来困难.

线性系统阶次一般可根据系统输入输出数据构成矩阵的秩加以判断.对于确定性系统,系统阶次容易判断.由于实际系统存在噪声干扰,数据矩阵总是满秩的,就通过随阶次增加(即数据矩阵维数的增加)比较数据矩阵行列式值的显著变化情况来确定系统阶次;或者构造一个维数大的数据矩阵(维数高于系统阶次),使用奇异值分解(SVD, Singular Value Decompositon)方法,根据奇异值大小的显著变化情况来确定系统阶次.总之,对于控制问题,阶次选择小一些;对于预测和软测量问题,模型阶次可以选择大一些.

3.3 实验设计

辨识实验设计包括输入信号的设计(信号类型、幅度和频带等)、采样周期的选择、实验时间长度(数据长度)的确定、开环或闭环辨识、离线辨识或在线辨识.目的是使采集的输入输出数据尽可能多地包含系统的内在特性.

3.3.1 输入信号设计(Design of Input Signals)

为了使系统是可辨识的,输入信号必须满足一定的条件,最低的要求是在实验期间,输入信号必须充分激励系统的所有模态(过程动态).从谱分析角度看,这就意味着输入信号的频谱必须足以覆盖系统的频谱.理想情况下,输入信号的频谱应该是无限的,白噪声就是这样一种信号.这就引出了持续激励输入信号的概念.

考虑一个最简单的有限脉冲响应模型描述的系统:

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + \dots + b_n u(t-n) + v(t), \quad (9)$$

其中 $u(t)$ 和 $y(t)$ 分别为系统的输入和输出, $\{v(t)\}$ 是零均值、方差为 σ^2 的白噪声序列:

$$E[v(t)] = 0, E[v^2(t)] = \sigma^2,$$

$$E[v(t)v(s)] = 0, s \neq t,$$

且与输入信号 $\{u(t)\}$ 不相关: $E[v(t)u(i)] = 0$.

定义参数向量 θ 和信息向量 $\varphi(t)$ 如下:

$$\theta := [b_1, b_2, \dots, b_n]^T \in \mathbf{R}^n;$$

$$\varphi(t) := [u(t-1), u(t-2), \dots, u(t-n)]^T \in \mathbf{R}^n.$$

式(9)可以写为最小二乘格式:

$$y(t) = \varphi^T(t)\theta + v(t). \quad (10)$$

假设数据长度为 L . 采用最小二乘法选择估计参数, 定义误差平方和准则函数

$$\begin{aligned} J(\theta) &:= [y(1) - \varphi^T(1)\theta]^2 + [y(2) - \varphi^T(2)\theta]^2 + \\ &\quad \dots + [y(L) - \varphi^T(L)\theta]^2 = \\ &\quad \sum_{t=1}^L [y(t) - \varphi^T(t)\theta]^2. \end{aligned} \quad (11)$$

极小化 $J(\theta)$ 得到 θ 的估计值 $\hat{\theta}$. 有时为了简化, 这个优化问题可以等价写为

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{t=1}^L [y(t) - \varphi^T(t)\theta]^2,$$

符号 argmin 是极小化准则函数获得的解的意思. 令 $J(\theta)$ 对 θ 的导数为零, 得到

$$-2 \sum_{t=1}^L \varphi(t) [y(t) - \varphi^T(t)\theta] = 0.$$

或

$$\left[\sum_{t=1}^L \varphi(t) \varphi^T(t) \right] \theta = \sum_{t=1}^L \varphi(t) y(t).$$

定义数据乘积矩阵阵 (Data Product Moment Matrix)

$$\mathbf{R}_L := \sum_{t=1}^L \varphi(t) \varphi^T(t) \in \mathbf{R}^{n \times n}.$$

如果矩阵 \mathbf{R}_L 是可逆的 (在这种情形, 我们说 $\varphi(t)$ 是持续激励的, 输入 $u(t)$ 是持续激励信号), 那么 θ 的估计为

$$\begin{aligned} \hat{\theta} &= \left[\sum_{t=1}^L \varphi(t) \varphi^T(t) \right]^{-1} \sum_{t=1}^L \varphi(t) y(t) = \\ &\quad \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) y(t). \end{aligned} \quad (12)$$

将式(10)的 $y(t)$ 代入式(12)得到

$$\begin{aligned} \hat{\theta} &= \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) [\varphi^T(t)\theta + v(t)] = \\ &\quad \theta + \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) v(t). \end{aligned} \quad (13)$$

利用 $v(t)$ 是零均值白噪声的性质, 假设输入是确定性的 ($\varphi(t)$ 和 \mathbf{R}_L 也是确定性的), 式(13)取数学期望得到

$$\mathbf{E}[\hat{\theta}] = \theta + \mathbf{E} \left[\mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) v(t) \right] = \frac{\mathbf{R}_L}{L} = \mathbf{I}_n \quad (18)$$

$$\theta + \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) \mathbf{E}[v(t)] = \theta. \quad (14)$$

这说明 $\hat{\theta}$ 是 θ 的无偏估计 (Unbiased Estimate).

定义参数估计误差向量 $\hat{\theta} - \theta$ 的协方差矩阵 (Covariance Matrix)

$$\mathbf{P}_L := \mathbf{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \in \mathbf{R}^{n \times n}.$$

注意到 \mathbf{R}_L 是对称阵, 利用式(13)可得

$$\begin{aligned} \mathbf{P}_L &= \mathbf{E} \left[\left(\mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) v(t) \right) \left(\mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) v(t) \right)^T \right] = \\ &\quad \mathbf{E} \left[\left(\mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) v(t) \right) \left(\sum_{t=1}^L v(t) \varphi^T(t) \mathbf{R}_L^{-1} \right) \right]. \end{aligned}$$

上式中两个和式交叉项相乘的数学期望为零, 所以可以简化为

$$\begin{aligned} \mathbf{P}_L &= \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) \mathbf{E}[v^2(t)] \varphi^T(t) \mathbf{R}_L^{-1} = \\ &\quad \mathbf{R}_L^{-1} \sum_{t=1}^L \varphi(t) \varphi^T(t) \mathbf{R}_L^{-1} \sigma^2 = \\ &\quad \mathbf{R}_L^{-1} \sigma^2 = \left(\frac{\mathbf{R}_L}{L} \right)^{-1} \frac{\sigma^2}{L}. \end{aligned} \quad (15)$$

从式(15)可以看出, 参数估计误差的协方差矩阵 \mathbf{P}_L 取决于输入数据矩阵 \mathbf{R}_L 和噪声方差 σ^2 .

把矩阵 \mathbf{R}_L 显示表达出来:

$$\begin{aligned} \frac{\mathbf{R}_L}{L} &= \frac{1}{L} \sum_{t=1}^L \varphi(t) \varphi^T(t) = \\ &\quad \frac{1}{L} \sum_{t=1}^L \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n) \end{bmatrix} [u(t-1), u(t-2), \dots, u(t-n)] = \\ &\quad \frac{1}{L} \sum_{t=1}^L \begin{bmatrix} u^2(t-1) & \cdots & u(t-1)u(t-n) \\ u(t-2)u(t-1) & \cdots & u(t-2)u(t-n) \\ \vdots & & \vdots \\ u(t-n)u(t-1) & \cdots & u^2(t-n) \end{bmatrix}. \end{aligned} \quad (16)$$

当数据长度 L 很大时, 我们选择输入信号 $\{u(t)\}$ 满足

$$\frac{1}{L} \sum_{i=1}^L u(t-i)u(t-j) = \begin{cases} 1, & i=j, \\ 0, & i \neq j. \end{cases} \quad (17)$$

1) 单位方差白噪声输入信号

如果 $L \rightarrow \infty$, 由式(17), 输入 $\{u(t)\}$ 可以看作是零均值、单位方差不相关可测随机变量序列 (相当于单位方差白噪声). 在这种情况下, 有

和

$$P_L = \left(\frac{R_L}{L}\right)^{-1} \frac{\sigma^2}{L} = \frac{\sigma^2}{L} I_n. \quad (19)$$

式(19)说明:参数估计误差协方差阵 P_L 随数据长度 L 增大而减小,即协方差阵 P_L 随 $1/L$ 衰减. 参数估计误差向量的方差为

$$E[\|\hat{\theta} - \theta\|^2] = \text{tr}\{E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T]\} = \text{tr}[P_L] = \frac{n\sigma^2}{L}.$$

上式说明:参数估计向量均方误差 $E[\|\hat{\theta} - \theta\|^2]$ 随数据长度 L 增大而减小,随噪声方差 σ^2 增大而增大,即反比于数据长度 L ,正比于噪声方差;可近似认为参数估计误差 $\hat{\theta} - \theta$ 以 $1/\sqrt{L}$ 速度收敛于零.

在式(17)条件下,利用式(18),由式(12)可知参数估计有下列简单表达式:

$$\hat{\theta} = R_L^{-1} \sum_{t=1}^L \varphi(t)y(t) = \frac{1}{L} \sum_{t=1}^L \varphi(t)y(t).$$

2) 方差为 σ_u^2 的白噪声输入信号

由上述推导可知:如果输入 $\{u(t)\}$ 是零均值,方差为 σ_u^2 的不相关可测随机变量序列,那么有

$$\begin{aligned} \frac{R_L}{L} &= I_n \sigma_u^2, \\ \hat{\theta} &= \frac{1}{L\sigma_u^2} \sum_{t=1}^L \varphi(t)y(t), \\ E[\|\hat{\theta} - \theta\|^2] &= \frac{n\sigma^2}{L\sigma_u^2} = \frac{n}{L} \frac{\sigma^2}{\sigma_u^2}. \end{aligned}$$

即:如果噪声方差不变,加大输入信号方差可提高参数估计精度.

3) 幅值为 M 的伪随机二进制持续激励信号

类似地,如果输入 $\{u(t)\}$ 采用零均值,幅值为 M 的 $+M$ 或 $-M$ 的持续激励信号,即 $\{u(t)\}$ 满足

$$\frac{1}{L} \sum_{i=1}^L u(t-i)u(t-j) = \begin{cases} M^2, & i=j, \\ 0, & i \neq j. \end{cases} \quad (20)$$

那么有

$$\frac{R_L}{L} = I_n M^2, \quad (21)$$

$$\hat{\theta} = \frac{1}{LM^2} \sum_{t=1}^L \varphi(t)y(t), \quad (22)$$

$$E[\|\hat{\theta} - \theta\|^2] = \frac{n\sigma^2}{LM^2} = \frac{n}{L} \frac{\sigma^2}{M^2}. \quad (23)$$

即:加大输入信号的幅值 M ,有利于提高参数估计精度. 在获得参数估计 $\hat{\theta}$ 后,根据式(11)可求得噪声方差的估计:

$$\hat{\sigma}^2 = \frac{1}{L-n} J(\hat{\theta}), \quad n: = \dim\theta. \quad (24)$$

因此,实验中使用输入信号 $u(t)$ 的幅度不宜过小. 输入信号幅度过小,会使系统中的噪声占主导地位,淹没有用的信号. 输入信号幅度也不宜过大,否则使系统进入非线性区域,更重要的是系统在稳定运行时,在不影响系统正常生产下,实验输入信号幅值应尽量大(谈不上进入非线性区域);另外,输入 $\{u(t)\}$ 对系统的“净扰动”要小,即 $\{u(t)\}$ 中所含的直流分量(即 $\{u(t)\}$ 的均值)要尽可能地小,以保证系统不会偏离正常的稳定工作状态.

上述分析表明:尽管白噪声作为持续激励输入信号有很好的性质(在计算机仿真中经常用作输入信号),参数估计也容易计算,但是白噪声在工业中实现也是不允许的,比如阀门开度如果剧烈地变化,容易受损坏. 实验信号通常采用幅值为 M 的伪随机二进制码形成的伪随机二进制序列(PRBS, Pseudo-Random Binary Sequence),简称 M 序列.

M 序列是 PRBS 的一种形式,是一个周期性持续激励序列. 它仅有两个值,由一个值变到另一个值要经过一段时间(Δt 的整数倍,一个 Δt 称为一个节拍(bit))才发生,且下一个 Δt 的状态与前一个 Δt 的状态相互独立. 这方面的材料可参见文献[7].

在辨识仿真中,可以使用 Matlab 函数 idinput 产生输入信号. 该函数可以产生随机高斯信号、随机二进制信号、伪随机二进制信号、正弦信号等,用法如下:

- ① $U = \text{IDINPUT}(N, \text{TYPE}, \text{BAND}, \text{LEVELS})$
- ②
- ③ U : The generated input signal. A column vector or a N-by-nu matrix.
- ④ N : The length of the input.
- ⑤ $N = [N \text{ Nu}]$ gives a N-by-Nu input (Nu input channels).
- ⑥ $N = [P \text{ Nu } M]$ gives a $M * P$ -by-Nu input, periodic with period P
- ⑦ and with M periods.
- ⑧ Default values are $\text{Nu} = 1$ and $M = 1$;
- ⑨ TYPE : One of the following:
- ⑩ 'RGS': Generates a Random, Gaussian Signal.
- ⑪ 'RBS': Generates a Random, Binary Signal.
- ⑫ 'PRBS': Generates a Pseudo-random, Binary Signal.
- ⑬ 'SINE': Generates a sum-of-sinusoid signal.
- ⑭ Default: $\text{TYPE} = \text{'RBS'}$.

3.3.2 采样周期选择(Choice of Sampling Periods)

一般实际系统都是连续时间系统,采集数据需

要经过采样处理. 采样周期的选择, 就是确定输入输出采样时间的间隔. 采样周期的选择直接影响辨识精度, 甚至影响系统的可辨识性.

实际系统选择采样周期, 应该考虑下列因素.

1) 采样周期应满足奈奎斯特-香农采样定理 (Nyquist-Shannon Sampling Theorem), 即采样速度不应低于系统截止频率的两倍. 这只是一个理论结果, 无法使用, 因为系统的传递函数是要辨识的, 其截止频率不可能知道.

2) 采样周期 h 的选取不宜太小, 这是因为当 h 太小时, 如果系统是一个慢过程, 相邻时刻的输出采样值 $y(t)$ 与 $y(t-1)$ 可能没有差别, 导致包含输入输出数据乘积矩阵接近奇异, 导致病态方程组, 使辨识的解不可靠. 采样周期也不宜过大, 不然, 采集的信号的信息量损失太大, 直接影响辨识精度.

系统的快慢是由系统的时间常数决定的, 即由稳定系统特征值实部的倒数决定的 (系统极点的位置决定的).

3) 从信号分析的角度, 采样应尽可能快, 这样能采集到更多的信息, 精度才可能高. 然而, 对于实际问题, 只要能满足控制或预测精度要求即可, 因为采样周期太小, 单位时间内数据量太大, 计算处理太耗时 (Time-Consuming), 增加设备负担.

4) 采样周期难以有一个严格的规定. 它依赖系统的快速性, 即系统的上升时间和过渡过程时间 (5% 误差内的调节时间). 采样周期 h 的选择是很宽的, 工程上可采用经验公式

$$h = \frac{t_s}{5 \sim 20},$$

其中 t_s 为过渡过程时间 (5% 误差内的调节时间). 此公式说明: 在过渡过程内, 应该采样 5 到 20 个数据. 这只有通过多次实验得到.

3.3.3 实验时间或确定数据长度 (Test Time or Determining the Data Length)

理论分析表明: 参数估计误差一般随数据长度增加而减小. 用于辨识的数据长度 L 越长, 辨识精度越高. 因此, 数据长度应根据要求的辨识精度 (参数估计精度) 确定. 然而, 实际中不允许实验时间太长, 再说实验时间太长也耗费人力和物力等. 在任何情况下, 数据长度一般应远远大于系统参数数目.

下面以有限脉冲响应模型描述的系统 (9) 为例说明实验时间长度 (数据长度) 的确定方法.

假设输入 $\{u(t)\}$ 采用零均值, 幅值为 M 的 $+M$

或 $-M$ 的持续激励信号, 且对某个数据长度 L , 我们已用某个辨识方法求出参数估计 $\hat{\theta}$, 使用式 (24) 可以初步估算出噪声方差:

$$\hat{\sigma}^2 = \frac{1}{L-n} J(\hat{\theta}) = \frac{1}{L-n} \sum_{t=1}^L [y(t) - \varphi^T(t)\hat{\theta}]^2.$$

当参数估计值不准确时, 上式计算出的噪声方差估计只会偏大. 假设给定的参数估计精度为 ε , 即要求误差满足 $\|\hat{\theta} - \theta\| \leq \varepsilon$, 近似把 $E[\|\hat{\theta} - \theta\|^2]$ 认为是 $\|\hat{\theta} - \theta\|^2$, 由式 (23) 可知, 参数估计误差应满足

$$\frac{n \hat{\sigma}^2}{LM^2} \leq \varepsilon^2.$$

因此, 在给定的辨识精度下, 实验数据长度 L 必须满足:

$$L \geq \frac{n \hat{\sigma}^2}{\varepsilon^2 M^2}.$$

至于 ARX 模型的实验时间长度的确定, 情况要复杂一点.

3.3.4 开环辨识还是闭环辨识 (Open-Loop Identification or Closed-Loop Identification)

如果系统允许, 一般应进行开环辨识实验, 等系统稳定后 (此时 $u(t) = u_0$ 和 $y(t) = y_0$ 为系统的稳定运行工作点), 再产生一个许可幅值 M 的 M 序列 $u^*(t)$ 作为实验输入信号, 叠加在 $u(t)$ 上, 记录系统的输出 $y(t)$ (输出的变化值为 $y^*(t) = y(t) - y_0$). 利用输入信号和输出信号的变化值 $\{u^*(t), y^*(t)\}$ 进行模型辨识, 这相当于已经对输入输出数据进行零均值化了.

如果开环系统不稳定时, 就要进行闭环辨识. 对于闭环系统来说 (参见图 7), 对象的输入 $u(t)$ (即控制输入) 是参考输入 $r(t)$ 与输出反馈信号 $f(t)$ 的叠加: $u(t) = r(t) - f(t)$, 同样是等闭环系统稳定后 (此时 $r(t) = r_0, u(t) = u_0$ 和 $y(t) = y_0$ 为系统的稳定运行工作点), 改变参考输入信号 (即在参考输入上加上一个 M 序列), 记录对象两端的输入输出数据, 利用其变化值进行模型辨识.

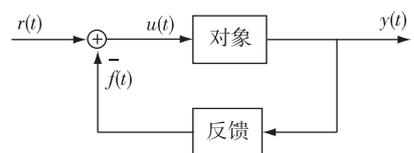


图 7 环系统辨识示意

Fig. 7 The closed-loop system identification

3.3.5 离线辨识还是在线辨识 (Off-Line Identification or On-Line Identification)

离线辨识是在所有实验数据采集完了之后才计算参数估计结果,而在线辨识中采集数据和计算结果是同时进行的.在实际中,如果在系统的稳定工作点附近,可看作一个时不变系统,离线辨识即可,也可进行在线辨识,通过递推算法实现.但在基于辨识的自适应控制系统中,辨识必须是在线的.对于时变参数系统,辨识也必须在线进行.

3.4 数据预处理

从辨识实验中收集到的输入输出数据在用于辨识前,需要进行预处理.基本的数据预处理方法有零均值化方法和数据滤波方法等.

采集的数据一般包含干扰噪声,如常见的低频干扰、偏离、飘移和周期性变化等.处理这种信号干扰有两种基本方法:去掉干扰或用噪声模型将干扰考虑进去.

零均值化方法可以去掉信号中的直流成份,解决数据飘移、偏离问题;数据滤波方法可以去掉信号中的高频干扰.因此人们总是先处理数据,然后进行辨识.

输入输出数据通常都含有直流成分,用任何辨识方法都无法消除它们对辨识精度的影响.此外,数据中的高频成分对辨识也是不利的.因此对输入输出数据要进行零均值化和剔除高频成分的预处理.处理得好,就能显著提高辨识的精度^[7].

下面介绍数据预处理的基本方法.

3.4.1 零均值化方法

不管是开环辨识实验,还是闭环辨识实验,一般要等系统稳定运行后再进行.系统的稳定工作点记作 $u(t) = u_0$ 和 $y(t) = y_0$. 采集的数据通常是绝对的输入输出(即观测到的输入输出)值,记作为 $u^*(t)$ 和 $y^*(t)$. 辨识一般使用输入输出的变化量 $u(t) := u^*(t) - u_0$ 和 $y(t) := y^*(t) - y_0$. 如果采用零均值的输入信号(如 M 序列)叠加在 u_0 上,那么 u_0 和 y_0 可认为是输入和输出的均值.

设实际观测到的输入输出为 $u^*(t)$ 和 $y^*(t)$, 则零均值化后的输入输出数据为

$$u(t) = u^*(t) - \bar{u}, \quad y(t) = y^*(t) - \bar{y}, \quad (25)$$

其中 \bar{u} 和 \bar{y} 分别为输入输出数据的直流分量(均值). \bar{u} 和 \bar{y} 就是观测值的平均,计算公式为

$$\bar{u} = \frac{1}{L} \sum_{i=1}^L u^*(i), \quad \bar{y} = \frac{1}{L} \sum_{i=1}^L y^*(i), \quad (26)$$

其中 L 为所考虑的数据长度. 这是一个离线计算均值的方法,把所有的观测数据进行平均.

还有一种在线计算均值的方法. 实际中,当前时刻 t 时的均值定义为直到时刻 t 的数据的平均,因此这个在线均值与 t 有关,其计算公式为

$$\bar{u}(t) = \frac{1}{t} \sum_{i=1}^t u^*(i), \quad \bar{y}(t) = \frac{1}{t} \sum_{i=1}^t y^*(i). \quad (27)$$

在线零均值化后的输入输出数据为

$$u(t) = u^*(t) - \bar{u}(t), \quad y(t) = y^*(t) - \bar{y}(t), \quad (28)$$

由式(27)可以得到时刻 t 在线均值的递推计算式:

$$\begin{aligned} \bar{u}(t) &= \frac{1}{t} \sum_{i=1}^t u^*(i) = \frac{1}{t} u^*(t) + \frac{1}{t} \sum_{i=1}^{t-1} u^*(i) = \\ &= \frac{1}{t} u^*(t) + \frac{t-1}{t} \frac{1}{t-1} \sum_{i=1}^{t-1} u^*(i) = \\ &= \frac{1}{t} u^*(t) + \frac{t-1}{t} \bar{u}(t-1) = \\ &= \bar{u}(t-1) + \frac{1}{t} [u^*(t) - \bar{u}(t-1)], \end{aligned} \quad (29)$$

$$\bar{y}(t) = \bar{y}(t-1) + \frac{1}{t} [y^*(t) - \bar{y}(t-1)]. \quad (30)$$

3.4.2 差分方法

由式(25)可得

$$u^*(t) = u(t) + \bar{u}, \quad y^*(t) = y(t) + \bar{y}. \quad (31)$$

式(31)说明:对每一个 t , 观测输入输出 $u^*(t)$ 和 $y^*(t)$ 分别都包含了常数均值 \bar{u} 和 \bar{y} . 相邻时刻数据观测值之差 $u^*(t) - u^*(t-1)$ 和 $y^*(t) - y^*(t-1)$ 就不包含均值了,且

$$u^*(t) - u^*(t-1) = u(t) - u(t-1),$$

$$y^*(t) - y^*(t-1) = y(t) - y(t-1).$$

因此,可用观测值之差进行辨识. 设 z^{-1} 是单位后移算子: $z^{-1}y(t) = y(t-1)$. 则有

$$(1 - z^{-1})y(t) = y(t) - y(t-1) =: \Delta y(t).$$

考虑差分方程描述的系统:

$$A(z)y(t) = B(z)u(t) + v(t). \quad (32)$$

辨识算法中所用的输入输出数据应该是经过零均值化处理、不含直流偏差的,但是实际观测到的输入输出含有偏差,所以方程两边同时乘以 $(1 - z^{-1})$ 得到

$$A(z)(1 - z^{-1})y(t) =$$

$$B(z)(1 - z^{-1})u(t) + (1 - z^{-1})v(t),$$

或

$$A(z)\Delta y(t) = B(z)\Delta u(t) + \Delta v(t). \quad (33)$$

其中差分 $\Delta u(t)$ 和 $\Delta y(t)$ 分别为

$$\begin{aligned}\Delta u(t) &= u(t) - u(t-1) = u^*(t) - u^*(t-1), \\ \Delta y(t) &= y(t) - y(t-1) = y^*(t) - y^*(t-1).\end{aligned}$$

那么可以直接利用 $\Delta u(t)$ 和 $\Delta y(t)$ 进行辨识, 它们不再含有直流分量. 这样实际上间接地对输入输出数据进行了零均值处理.

值得注意的是, 如果输入采用零均值随机噪声, 可以采用差分方法, 利用 $\Delta u(t)$ 和 $\Delta y(t)$ 进行辨识, 因为噪声之差还是随机噪声, 但是如果输入采用方波(如 M 序列), 那么 $\Delta u(t)$ 就变成了脉冲和一些零值, 可能会影响辨识效果. 采用差分后, 噪声模型的结构也发生了变化, 这种方法可以处理噪声为非零均值情形, 因为 $\Delta v(t) = v(t) - v(t-1)$ 的均值为零.

3.4.3 方程常数方法

通常用于辨识的输入输出数据需要进行零均值化. 如果用非零均值输入输出数据进行辨识, 可能导致参数估计误差很大, 这可以在待辨识模型中加入一未知常数项来解决. 下面用 CARMA 模型加以说明.

假设测量的输入数据和输出数据分别为 $\{u'(t)\}$ 和 $\{y'(t)\}$, 它们的均值为 $E[u'(t)] = \mu_u$ 和 $E[y'(t)] = \mu_y$ (未知的). 令

$$u(t) := u'(t) - \mu_u, \quad (34)$$

$$y(t) := y'(t) - \mu_y. \quad (35)$$

则 $\{u(t)\}$ 和 $\{y(t)\}$ 是零均值输入数据序列和零均值输出数据序列. 假设用于辨识的 CARMA 模型为

$$A(z)y(t) = B(z)u(t) + D(z)v(t), \quad (36)$$

$\{v(t)\}$ 可认为是零均值干扰随机噪声序列, $A(z)$, $B(z)$ 和 $D(z)$ 是单位后移算子 z^{-1} 的多项式 $[z^{-1}y(t) = y(t-1)]$:

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_{n_a}z^{-n_a},$$

$$B(z) = b_1z^{-1} + b_2z^{-2} + \cdots + b_{n_b}z^{-n_b},$$

$$D(z) = 1 + d_1z^{-1} + d_2z^{-2} + \cdots + d_{n_d}z^{-n_d},$$

其中 a_i , b_i 和 d_i 是要用输入输出数据 $\{u(t), y(t)\}$ 进行辨识的参数.

将式(34)和(35)代入式(36)得到

$$\begin{aligned}A(z)[y'(t) - \mu_y] &= \\ B(z)[u'(t) - \mu_u] &+ D(z)v(t),\end{aligned}$$

或

$$\begin{aligned}A(z)y'(t) &= \\ B(z)u'(t) &+ D(z)v(t) + A(1)\mu_y - B(1)\mu_u,\end{aligned} \quad (37)$$

其中

$$A(z)\mu_y = A(1)\mu_y = (1 + a_1 + a_2 + \cdots + a_{n_a})\mu_y,$$

$$B(z)\mu_u = B(1)\mu_u = (b_1 + b_2 + \cdots + b_{n_b})\mu_u.$$

令 $f := A(1)\mu_y - B(1)\mu_u$, 就得到一个带有静态未知扰动 f 的 CARMA 模型:

$$A(z)y'(t) = B(z)u'(t) + D(z)v(t) + f. \quad (38)$$

这个模型中, $\{y'(t)\}$ 和 $\{u'(t)\}$ 是非零均值输出输入数据, 把未知静态扰动(确定性扰动) f 看作一个参数进行辨识, 从而解决使用非零均值数据辨识的问题.

这个方法是在原模型中加一个未知常数, 故把这种方法称为方程常数方法.

3.4.4 数据预滤波方法

用一个线性滤波器对系统输入输出数据进行滤波, 不会改变线性系统的输入输出关系, 但可能改变系统干扰噪声的结构, 达到提高辨识精度的目的.

考虑如下系统:

$$y(t) = G(z)u(t) + H(z)v(t).$$

其中 $v(t)$ 为白噪声, $G(z)$ 是系统模型的传递函数, $H(z)$ 是噪声模型的传递函数. $w(t) := H(z)v(t)$ 是有色噪声(相关噪声).

为了去掉输入输出数据中的高频成份, 可采用低通滤波器 $L(z)$ 对输入输出数据进行滤波. 当然, 滤波器也可根据实际情况进行选择.

上述方程两边同乘以线性滤波器 $L(z)$ 得到

$$L(z)y(t) = G(z)L(z)u(t) + L(z)H(z)v(t).$$

令

$$y_f(t) := L(z)y(t), u_f(t) := L(z)u(t).$$

因此, 有

$$y_f(t) = G(z)u_f(t) + L(z)H(z)v(t).$$

如果取 $L(z) = \frac{1}{H(z)}$, 那么上式化为一个白噪声干扰的输出误差模型:

$$y_f(t) = G(z)u_f(t) + v(t).$$

然而噪声模型的传递函数 $H(z)$ 是未知, 在递推算法中, 通常采用其估计 $\hat{H}(z)$, 即取

$$L(z) = \frac{1}{\hat{H}(z)}$$

来对输入输出数据进行滤波, (参见文献[9]中的基于滤波式的递推辨识算法).

4 辨识方法的类别

辨识方法(估计方法)有很多类别, 按其计算方式可分为一次完成算法、递推估计方法(递推辨识方

法)和迭代估计方法(迭代辨识方法);按其实时性可分为在线估计方法(在线辨识方法)和离线估计方法(离线辨识方法);按其属性特征可分为最小二乘估计算法、最小均方估计算法、梯度估计算法、随机逼近估计算法、辅助模型辨识方法、多新息辨识方法、递阶辨识方法、极大似然辨识方法、贝叶斯辨识方法(Bayesian Identification Method)等,时不变参数估计方法和时变参数估计方法,以及随机参数估计方法等.下面简单介绍最小二乘辨识方法、梯度辨识方法、辅助模型辨识方法、多新息辨识方法、递阶辨识方法、耦合辨识方法,这些方法将在后继系列论文中做详细介绍.

4.1 最小二乘辨识方法

18世纪末,德国数学家高斯(C. F. Gauss, 30 April 1777—3 February 1855)首先提出了最小二乘法,他用最小二乘法计算天体运行的轨道.这一方法后来被广泛用于系统辨识领域.

20世纪60年代,随着电子计算机的普及,系统辨识方法(参数估计方法)有了飞速的发展.

为了选出使得模型输出与系统输出 $y(t)$ 尽可能接近的参数估计值,可用模型与系统输出的误差的平方和来度量接近程度,使误差平方和最小的参数值即为所求的估计值.

在系统辨识领域中,最小二乘法是一种基本的参数估计方法.它可用于动态系统辨识,也可用于静态系统参数拟合;可用于线性系统,也可用于非线性系统;可用于离线估计,也可用于在线估计;可用于递推算法,也可用于迭代算法.最小二乘法有离线辨识方法和在线辨识方法,递推辨识和迭代辨识方法.

针对不同的用途,对最小二乘法进行修正,就得到了各种相应的最小二乘类辨识算法.如用于辨识方程误差类模型的递推最小二乘类辨识方法,用于辨识输出误差类模型的辅助模型最小二乘类辨识方法,能够改善辨识精度的最小二乘迭代辨识算法和多新息最小二乘辨识算法,能够降低计算量的递阶最小二乘辨识算法和递阶最小二乘迭代辨识算法等^[10-12].

4.2 梯度辨识方法

梯度类辨识方法的特点是算法中没有协方差矩阵.与最小二乘法相比,梯度算法计算量小.

梯度辨识方法的准则函数可采用模型输出与系统输出的误差的平方,也可采用模型输出与系统输出的误差的平方和.而最小二乘辨识方法的准则函

数只能采用模型输出与系统输出的误差的平方和.

梯度辨识方法是沿着准则函数负梯度方向搜索参数估计(即通常所说的最速下降法).梯度辨识方法只能以递推或迭代算法形式实现.

梯度辨识方法派生出许多类型.如随机梯度辨识算法、遗忘梯度辨识算法、梯度迭代辨识算法、多新息随机梯度辨识算法、多新息遗忘梯度辨识算法,以及能够降低计算量的递阶随机梯度算法和递阶梯度迭代辨识算法等^[13-14].

4.3 辅助模型辨识方法

辅助模型辨识思想^[15]是本文作者在其硕士论文《多变量系统辨识》中提出的.辅助模型辨识方法主要是针对辨识模型信息向量中存在未知项的辨识难题提出的.辅助模型辨识(Auxiliary Model Identification)的基本思想是借助于一个辅助模型,用系统的可测信息建立一个辅助模型,用辅助模型的输出代替系统的不可测变量(即信息向量中的未知项)的一种辨识方法,并且通过选择辅助模型的参变量(参数),使辅助模型(Auxiliary Model)的输出逼近这些不可测变量,从而获得系统参数的一致估计.辅助模型辨识方法是估计输出误差类模型参数的一种有效方法.

最近,辅助模型辨识思想已经推广用于解决双率(多率)采样数据系统^[16]、损失数据系统^[17]以及输入非线性输出误差系统的辨识问题^[18]等,相关论文都发表在国际期刊上.

4.4 多新息辨识方法

多新息辨识理论^[19]是近年发展起来的新辨识理论.它是本文作者在其博士论文《时变参数系统辨识及其应用》中提出的.多新息辨识方法是受文献[20]算法间断迭代思想的启发,最初用类比方法,直接给出了变递推间隔多新息广义投影辨识算法的数学表达式^[21],后经过深入研究,从理论上详细推导了多新息投影辨识算法、多新息随机梯度算法、多新息最小二乘辨识算法、变递推间隔多新息最小二乘算法等.这使得多新息辨识算法有了严密的数学基础^[22-25].

一般单输入单输出系统递推辨识算法 t 时刻的参数估计 $\hat{\theta}(t) \in \mathbf{R}^n$ 等于前一时刻的参数估计 $\hat{\theta}(t-1)$ 加上增益向量 $\mathbf{L}(t) \in \mathbf{R}^n$ 与标量新息 $e(t) \in \mathbf{R}$ 的乘积,即

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mathbf{L}(t)e(t).$$

我们把这个标量新息 $e(t)$ 加以推广, 就导出了多新息辨识方法 (Multi-Innovation Identification Method) [21,23]. 对标量系统而言, 将算法中的标量新息 $e(t) \in \mathbf{R}^1$ 推广为新息向量 $\mathbf{E}(p, t) \in \mathbf{R}^p$, 即多新息 (Multi-Innovation) (这就是多新息辨识方法名称的由来), 为使矩阵乘法维数兼容, 增益向量 $\mathbf{L}(t) \in \mathbf{R}^n$ 须推广为增益矩阵 (Gain Matrix) $\Phi(p, t) \in \mathbf{R}^{n \times p}$, 那么多新息辨识算法可以写作

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \Phi(p, t)\mathbf{E}(p, t),$$

其中 $p \geq 1$ 为新息长度 (Innovation Length).

把基于多新息的辨识理论称为多新息辨识理论 (Multi-Innovation Identification Theory), 把基于多新息的辨识方法称为多新息辨识方法 [4,22-26].

4.5 递阶辨识方法

递阶辨识 (Hierarchical Identification) [27], 即分解辨识的概念是本文作者博士毕业不久, 1997 年给清华大学硕士、博士研究生讲授《大系统理论及应用》时, 受大系统递阶控制的分解协调原理 (Decomposition-Coordination Principle) 的启发提出的, 从而建立了递阶辨识原理. 递阶辨识原理不仅可用于解决维数大、变量数多的大系统辨识的计算量大问题 [27], 而且对所谓的小系统、具有复杂结构耦合多变量系统的辨识也非常有用. 如在多变量系统传递函数阵主模型 (Main Model) 的递阶辨识方面, 显示出其独到的特点.

递阶辨识原理分 3 步.

第 1 步: 辨识模型分解为一些子辨识模型;

第 2 步: 采用最小二乘原理分别辨识一些子模型;

第 3 步: 采用协调处理各子辨识算法间的关联项.

在递阶辨识领域, 基于递阶辨识原理, 本文作者等首次提出了传递函数主模型的递阶梯度迭代辨识方法和递阶随机梯度辨识方法 [13]、递阶最小二乘迭代辨识方法和递阶最小二乘辨识方法 [10], 提出了状态空间模型的递阶辨识方法 [28]、双率系统提升状态空间模型的递阶辨识方法 [29] 等, 相关论文都发表在国际权威期刊《Automatica》和《IEEE Transactions》上.

5 结语

本文简述了一些建模方法, 从辨识准则函数的优化方法角度, 丰富了之前辨识的 3 要素, 提出了更改为确切的辨识 4 要素, 阐述了辨识的新定义, 介绍

了一些系统模型与辨识模型, 说明了时变参数辨识应注意的问题, 简述了辨识的基本步骤与辨识目的, 包括辨识试验设计与数据预处理, 介绍了一些典型的辨识方法及其类别, 包括最小二乘辨识方法、梯度辨识方法、辅助模型辨识方法、多新息辨识方法、递阶辨识方法等.

这是系列综述类论文的第 1 篇, 后面将依次介绍有关辨识精度与辨识的基本问题以及新的辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识原理及其相应的辨识方法 [30]. 文献 [7-8, 31-33] 给出了一些经典的最小二乘辨识方法和梯度辨识方法, 其他辨识综述论文可参见文献 [34] 和 [35].

参考文献

References

- [1] 万百五. 控制论创立六十年 [J]. 控制理论与应用, 2008, 25(4): 597-602
WAN Baiwu. Sixty years of cybernetics since founding [J]. Control Theory and Applications, 2008, 25(4): 597-602
- [2] 廖晓昕. 漫谈 Lyapunov 稳定性的理论、方法和应用 [J]. 南京信息工程大学学报: 自然科学版, 2009, 1(1): 1-15
LIAO Xiaoxin. Talking on the theory, methods and applications of Lyapunov stability [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2009, 1(1): 1-15
- [3] 廖晓昕. 综合国力非线性扩散模型稳定性分析 [J]. 南京信息工程大学学报: 自然科学版, 2009, 1(3): 247-251
LIAO Xiaoxin. Stability analysis of nonlinear diffusion model describing comprehensive national power [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2009, 1(3): 247-251
- [4] 丁锋. 基于输出估计的多输入系统随机梯度估计算法 [J]. 南京信息工程大学学报: 自然科学版, 2010, 2(6): 481-488
DING Feng. Stochastic gradient estimation algorithm for multiple-input systems based on the output estimation [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2010, 2(6): 481-488
- [5] Zadeh L A. From circuit theory to system theory [J]. Proceedings of the IRE, 1962, 50(5): 856-865
- [6] Ljung L. Convergence analysis of parametric identification methods [J]. IEEE Transactions on Automatic Control, 1978, 23(5): 770-783
- [7] 方崇智, 萧德云. 过程辨识 [M]. 北京: 清华大学出版社, 1988
FANG Chongzhi, XIAO Deyun. Process identification

- [M]. Beijing: Tsinghua University Press, 1988
- [8] 谢新民, 丁锋. 自适应控制系统[M]. 北京: 清华大学出版社, 2002.
XIE Xinmin, DING Feng. Adaptive control systems[M]. Beijing: Tsinghua University Press, 2002
- [9] Wang D Q, Ding F. Input-output data filtering based recursive least squares parameter estimation for CARAR-MA systems[J]. Digital Signal Processing, 2010, 20(4): 991-999
- [10] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems[J]. IEEE Transactions on Automatic Control, 2005, 50(3): 397-402
- [11] Liu X G, Lu J. Least squares based iterative identification for a class of multirate systems[J]. Automatica, 2010, 46(3): 549-554
- [12] Liu Y J, Wang D Q, Ding F. Least-squares based iterative algorithms for identifying Box-Jenkins models with finite measurement data[J]. Digital Signal Processing, 2010, 20(5): 1458-1467
- [13] Ding F, Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems[J]. Automatica, 2005, 41(2): 315-325
- [14] Ding F, Liu PX, Liu G. Gradient based and least-squares based iterative identification methods for OE and OEMA systems[J]. Digital Signal Processing, 2010, 20(3): 664-677
- [15] 丁锋. 多变量系统辨识[D]. 北京: 清华大学自动化系, 1990
DING Feng. Identification of multivariable systems[D]. Beijing: Department of Automation, Tsinghua University, Beijing, 1990
- [16] Ding F, Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model[J]. Automatica, 2004, 40(10): 1739-1748
- [17] Ding F, Ding J. Least squares parameter estimation with irregularly missing data[J]. International Journal of Adaptive Control and Signal Processing, 2010, 24(7): 540-553
- [18] Ding F, Shi Y, Chen T. Auxiliary model based least-squares identification methods for Hammerstein output-error systems[J]. Systems & Control Letters, 2007, 56(5): 373-380
- [19] 丁锋. 时变参数系统辨识及其应用[D]. 北京: 清华大学自动化系, 1994
DING Feng. Time-varying parameter system identification and its applications[D]. Beijing: Department of Automation, Tsinghua University, Beijing, 1994
- [20] Moustafa K A F. Identification of stochastic time-varying systems[J]. IEE Proceeding, Part D: Control Theory and Applications, 1983, 130(4): 137-142
- [21] 丁锋, 谢新民, 方崇智. 时变系统辨识的多新息方法[J]. 自动化学报, 1996, 22(1): 85-91
DING Feng, XIE Xinmin, FANG Chongzhi. Multi-innovation identification methods for time-varying systems[J]. Acta Automatica Sinica, 1996, 22(1): 85-91
- [22] 丁锋, 萧德云, 丁韬. 多新息随机梯度辨识方法[J]. 控制理论与应用, 2003, 20(6): 870-874
DING Feng, XIAO Deyun, DING Tao. Multi-innovation stochastic gradient identification methods[J]. Control Theory and Application, 2003, 20(6): 870-874
- [23] Ding F, Chen T. Performance analysis of multi-innovation gradient type identification methods[J]. Automatica, 2007, 43(1): 1-14
- [24] Ding F. Several multi-innovation identification methods[J]. Digital Signal Processing, 2010, 20(4): 1027-1039
- [25] Ding F, Liu P X, Liu G. Multi-innovation least squares identification for linear and pseudo-linear regression models[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(3): 767-778
- [26] Ding F, Liu P X, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises[J]. Signal Processing, 2009, 89(10): 1883-1890
- [27] 丁锋, 杨家本. 大系统的递阶辨识[J]. 自动化学报, 1999, 25(5): 647-654
DING Feng, YANG Jiaben. Hierarchical identification of large scale systems[J]. Acta Automatica Sinica, 1999, 25(5): 647-654
- [28] 丁锋, 萧德云. 多变量系统状态空间模型的递阶辨识[J]. 控制与决策, 2005, 20(8): 848-853, 859
DING Feng, XIAO Deyun. Hierarchical identification of state space models for multivariable systems[J]. Control and Decision, 2005, 20(8): 848-853, 859
- [29] Ding F, Chen T. Hierarchical identification of lifted state-space models for general dual-rate systems[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2005, 52(6): 1179-1187.
- [30] 丁锋. 系统辨识理论与方法[M]. 北京: 电力出版社, 2011
DING Feng. System identification theory and methods[M]. Beijing: China Electric Power Press, 2011
- [31] Goodwin G C, Sin K S. Adaptive filtering, prediction and control[M]. Prentice-Hall: Englewood Cliffs, NJ, 1984
- [32] 冯纯伯, 史维. 自适应控制[M]. 北京: 电子工业出版社, 1986
FENG Chunbo, SHI Wei. Adaptive control[M]. Beijing: Electronics Industry Press, 1986
- [33] Ljung L. System identification: Theory for the user[M]. 2nd ed. Prentice-Hall: Englewood Cliffs, NJ, 1999
- [34] Ding F, Chen T. Modeling and identification for multirate systems[J]. Acta Automatica Sinica, 2005, 31(1): 105-122
- [35] 丁锋, 杨慧中, 纪志成. 时变系统辨识方法及其收敛定理[J]. 江南大学学报: 自然科学版, 2006, 5(1): 115-126
DING Feng, YANG Huizhong, JI Zhicheng. Time-varying system identification methods and convergence theorems[J]. Journal of Southern Yangtze University: Natural Science Edition, 2006, 5(1): 115-126

System identification. Part A: Introduction to the identification

DING Feng¹

1 School of Internet of Things Engineering, Jiangnan University, Wuxi 214122

Abstract System identification is the theory and methods of establishing mathematical models of systems. The mathematical modeling has a long research history, but the system identification discipline has only several tens of years. In this short decades, system identification has achieved great developments, new identification methods are born one after another, and the research results cover the theory and applications of natural science and social sciences, including physics, biology, earth science, meteorology, computer science, economics, psychology, political science and so on. In this context, we come back to ponder some basic problems of system identification, which is not without benefits for the development of system identification. This is a paper of an introduction to system identification which briefly introduces the definition of identification, system models and identification models, the basic steps and purposes of identification, including the experimental design of identification and data preprocessing, and the types of identification methods, including the least squares identification methods, gradient identification methods, auxiliary model based identification methods, and multi-innovation identification methods, and hierarchical identification methods, etc.

Key words system identification; recursive identification; parameter estimation; auxiliary model based identification; multi-innovation identification; hierarchical identification