

系统辨识(3): 辨识精度与辨识基本问题

丁锋^{1,2}

摘要

系统辨识是研究建立系统数学模型的理论与方法. 讨论系统辨识涉及的一些基本问题, 包括辨识精度、辨识方法的提出, 辨识输入信号的设计, 参数可辨识性与系统可辨识性, 开环可辨识性与闭环可辨识性, 可辨识性与能控性和能观测性的关系, 可辨识性与输入信号的关系, 以及与辨识方法收敛性相关的激励信号与激励条件, 辨识算法收敛分析的基本工具, 典型辨识算法的收敛结论等.

关键词

数学模型; 系统辨识; 参数估计; CAR 模型; CARMA 模型; CARAR 模型; CARARMA 模型; 输出误差模型; OEMA 模型; OEAR 模型

中图分类号 TP273
文献标志码 A

收稿日期 2011-06-08

资助项目 国家自然科学基金(60973043)

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. fding@jiangnan.edu.cn

0 引言

控制是许多科学问题的核心, 数学模型是一切控制问题的基础, 控制理论与自动化技术飞速发展铸造了高度的科学文明和人类的辉煌成就. 自动化和信息产品对我们的社会生活和个体生活, 乃至经济发展带来了巨大变革^[1-3]. 控制理论与自动化技术对现代各学科的发展和贡献达到极致^[4-6].

客观事物数学模型的建立, 客观事物运动规律的揭示, 为人与自然的和谐发展和统一奠定了基础. 在控制论高度发展的今天, 来回顾一下控制科学的基础——数学模型和建立数学模型的理论与方法——系统辨识的一些基本问题, 是别有一番情趣的.

公理是不需要条件的, 定理都是一定条件下的产物, 辨识算法的收敛性定理也不例外, 自然科学中的有关定理都是如此. 辨识算法的收敛性与持续激励信号密切相关, 直接使用持续激励输入信号研究辨识方法的性能十分不方便, 为此, 引入了与持续激励信号相关的持续激励条件. 本文系统总结和归纳出强持续激励条件, 广义强持续激励条件, 弱持续激励条件, 广义弱持续激励条件, 衰减激励条件, 期望激励条件, 条件期望激励条件等.

提出新的建模理论、辨识方法, 以及辨识方法在不同条件下, 特别是在更弱条件下的性能分析, 包括时不变系统参数估计的一致收敛性, 时变系统参数估计误差上界的估算, 如何选择算法中最佳参数(如遗忘因子递推最小二乘算法中的遗忘因子、有限数据窗最小二乘算法中数据窗长度等)获得最小估计误差上界等, 都是辨识研究的永恒主题^[7].

本文把线性参数系统辨识模型分为线性回归模型、伪线性回归模型 I、伪线性回归模型 II 和伪线性回归模型 III 4 类. 这 4 类模型都是线性参数模型, 它们可以是线性系统, 也可以是线性参数的非线性系统. 本文的伪线性回归模型 I 含盖了线性系统的 CARMA 模型, 伪线性回归模型 II 含盖了线性系统的输出误差模型, 伪线性回归模型 III 含盖了线性系统的 CARAR, CARARMA 模型, OEMA 模型, OEAR 模型, Box-Jenkins 模型.

辨识算法的收敛性分析是辨识领域的重要课题, 现存辨识算法的几个漂亮收敛定理只限定于几个特定的系统模型结构, 如线性回归模型、伪线性回归模型 I 和伪线性回归模型 II. 对于伪线性回归模

1 江南大学 物联网工程学院, 无锡 214122

2 江南大学 控制科学与工程研究中心, 无锡, 214122

型Ⅲ和非线性回归模型辨识方法的收敛分析,目前还没有简洁收敛结果和漂亮的证明,这是系统辨识研究者有待攻克的研究难题.

本文较长,为便于阅读,特将本文框架结构列示如下.

0 引言

1 辨识精度

1.1 物理量辨识

1.2 系统外特性辨识

2 辨识的基本问题

2.1 辨识方法的提出

2.2 辨识输入信号设计

2.3 可辨识性问题

2.4 辨识算法的收敛性能

3 激励信号与激励条件

3.1 持续激励信号

3.2 弱持续激励信号

3.3 衰减激励信号

3.4 持续激励信号的产生

3.5 基本激励条件

4 参数估计性质及分析工具

4.1 参数估计的统计性质

4.2 Cramér-Rao 不等式

4.3 实用有界收敛性

4.4 收敛性分析的基本工具

5 典型辨识算法的收敛定理

5.1 线性回归模型

5.2 伪线性回归模型 I

5.3 伪线性回归模型 II

5.4 伪线性回归模型 III

6 结语

1 辨识精度

辨识精度包括辨识模型精度、辨识模型阶次估计精度、辨识模型参数估计精度,或者辨识模型输出与实际系统输出的接近程度.辨识精度可以解释为物理量辨识精度和系统外特性辨识精度.

1.1 物理量辨识

物理系统模型中的参数一般都有明确的意义,对于物理量的辨识,目标是获得精确的模型参数,所使用的辨识模型结构应与物理系统一致.例如,对于如图1所示的R-L-C电路,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

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

这个2阶电路系统模型有2个参数, $a = \frac{R}{L}$ 是物理量电阻 R 与电感 L 的函数, $b = \frac{1}{LC}$ 是电感 L 与电容 C 的函数.对于这个电路系统,系统辨识的目标就是通过实验数据,精确地估计出 a 和 b 的值.物理量辨识的目标是确定 R 、 L 和 C 的值.显然,这里由估计出 a 和 b 的值,难以从这2个方程 $\frac{R}{L} = a$ 和 $\frac{1}{LC} = b$ 确定 R 、 L 和 C 的值(然而,可以通过其他测量手段,建立更多的方程来计算 R 、 L 和 C).对于给定的输入为电源电压 $u(t)$ 和输出为电容 C 上电压 $y(t)$ 的量测值,这个电路是系统可辨识的,但不是参数可辨识的.

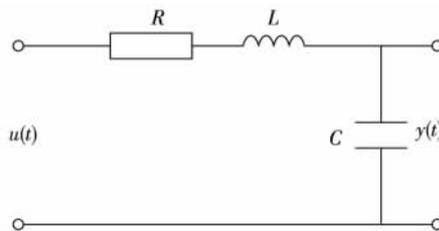


图1 电阻-电感-电容(R-L-C)电路

Fig. 1 The resistor-inductor-capacitor circuit

1.2 系统外特性辨识

系统辨识包括系统结构(或阶次)确定和系统参数估计.辨识精度可以理解为辨识出的系统模型结构的准确性和参数估计精度.对于物理系统,模型解析结构通常容易获得,剩余的问题就是参数辨识,就是物理量参数的估计,如万有引力定理中引力常量 G 的辨识^[5].

与物理量辨识相对应的是系统外特性辨识.对于一些实际工业过程系统,如锅炉燃烧系统、生物反应过程,其模型结构和参数往往呈现“不确定性”,且估计的模型结构或阶次也不一定准确(阶次估计只是一定准则意义下的近似).在这种情况下,把辨识精度约束为参数估计精度有失偏颇,一般把辨识精度解释为辨识模型输出与真实系统输出的接近程度.

例1 设对象的传递函数为

$$G(s) = \frac{1}{1 + 15s + 100s^2 + 10s^3}$$

辨识得到的传递函数为

$$\hat{G}(s) = \frac{1}{1 + 14.5s + 105s^2}$$

这个辨识模型的阶次与对象阶次不一致(当然可以选择一个与对象结构一致的辨识模型),因此比较参数估计精度意义不大.

2 个传递函数的单位阶跃响应曲线 $y(t)$ 和 $\hat{y}(t)$ 如图 2 所示,阶跃响应误差 $\hat{y}(t) - y(t)$ 随 t 变化曲线如图 3 所示,阶跃响应绝对误差 $|\hat{y}(t) - y(t)|$ 随 t 变化曲线如图 4 所示.对象 $G(s)$ 和估计 $\hat{G}(s)$ 的波德图如图 5 所示.这 2 个传递函数的增益都为 1,故它们阶跃响应的稳态值为 $y(\infty) = \hat{y}(\infty) = 1$.

通过下面的 Matlab 仿真,计算出阶跃响应最大相对误差为

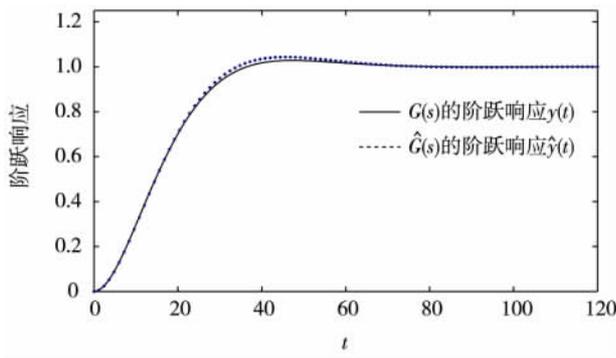


图 2 对象 $G(s)$ 和估计 $\hat{G}(s)$ 阶跃响应 $y(t)$ 和 $\hat{y}(t)$ 随 t 变化曲线

Fig. 2 The step responses of $G(s)$ and

$\hat{G}(s)$ ($y(t)$ and $\hat{y}(t)$ v. s. t)

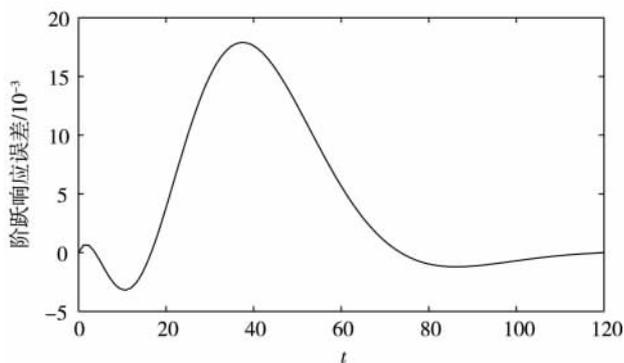


图 3 阶跃响应误差 $\hat{y}(t) - y(t)$ 随 t 变化曲线

Fig. 3 The step responses errors $\hat{y}(t) - y(t)$ v. s. t

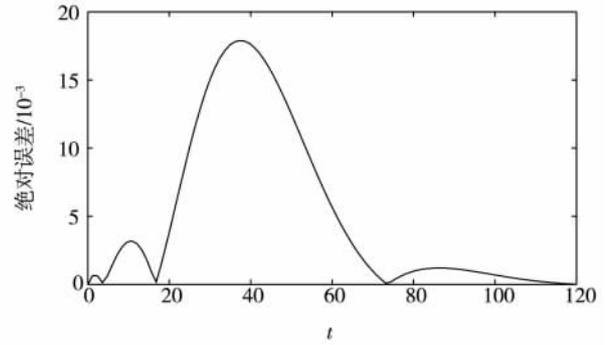


图 4 阶跃响应绝对误差 $|\hat{y}(t) - y(t)|$ 随 t 变化曲线

Fig. 4 The absolute errors $|\hat{y}(t) - y(t)|$ v. s. t

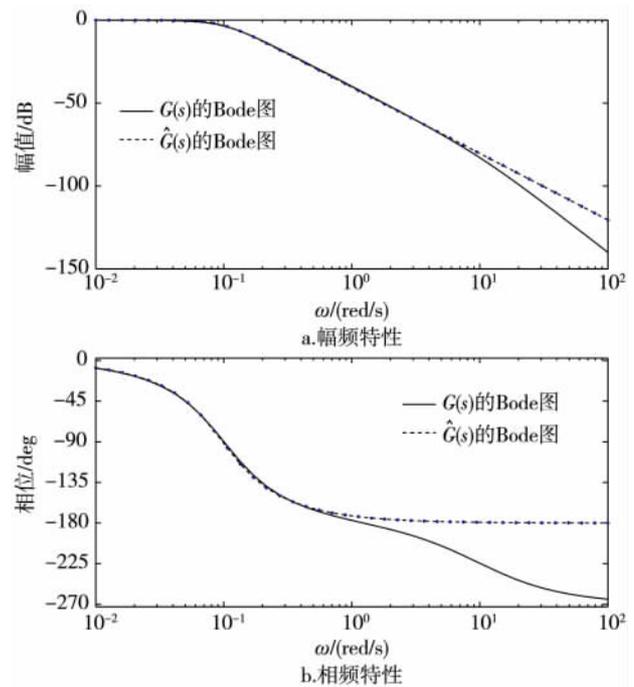


图 5 对象 $G(s)$ 和估计 $\hat{G}(s)$ 的波德图

Fig. 5 The Bode plots of $G(s)$ and $\hat{G}(s)$

$$\max |\hat{y}(t) - y(t)| / y(\infty) < 1.8\%$$

由此可见,尽管 2 个传递函数的阶次不同,但是阶跃响应还是很接近的,故辨识得到的 $\hat{G}(s)$ 能很好捕捉系统 $G(s)$ 的动态特性.从系统输出看,这个辨识精度还是令人满意的.

由频率特性波德图 5 可以看出:当频率 $\omega \leq 1$ rad/s 时,对象 $G(s)$ 和估计 $\hat{G}(s)$ 的复角相位是很接近的;当频率 $\omega \leq 10$ rad/s 时,对象 $G(s)$ 和估计 $\hat{G}(s)$ 的幅频特性是比较接近的.故当频率 ω 较小时, $\hat{G}(s)$ 是 $G(s)$ 的一个好的估计.

计算这个例子的阶跃响应 Matlab 源代码如下.

```

1 %-----
2 % Filename: ModelAcuracy2011_6_6.m
3 %           For computing the step responses and errors
4 %-----
5 clear; format short
6 b1 = [10 ,100 ,15 ,1 ]
7 tf1 = tf( 1 ,b1)
8
9 b2 = [105 ,14.5 ,1 ]
10 tf2 = tf( 1 ,b2)
11
12 L = 120;
13 [y1 ,t1] = step( tf1 ,L) ;
14 [y2 ,t2] = step( tf2 ,L) ;
15
16 figure( 1)
17 plot( t1 ,y1 , 'k' , t2 ,y2 , 'b. ' )
18 axis( [0 ,120 0 ,1.25 ] )
19 xlabel( ' \it          t' ) ;
20 ylabel( ' Step responses ' )
21
22 figure( 2)
23 plot( t1 ,y2 - y1 , 'k' )
24 axis( [0 ,120 0 ,0.005 0.02 ] )
25 xlabel( ' \it          t' ) ;
26 ylabel( ' Step response errors ' )
27
28 figure( 3)
29 plot( t1 ,abs( y2 - y1 ) , 'k' )
30 axis( [0 ,120 0 0.02 ] )
31 xlabel( ' \it          t' ) ;
32 ylabel( ' Absolute errors ' )
33
34 [t1 ,y1 ,t2 ,y2 ,y1 - y2 ,abs( y1 - y2 ) ]
35 s1 = [max( abs( y1 - y2 ) ) ,sum( abs( y1 - y2 ) ) ]

```

2 辨识的基本问题

特作申明: 本文中有的变量上下文使用了不同的定义, 只要一个变量重新定义了, 就采用新的定义. “ $A =: X$ ”或“ $X =: A$ ”表示“ A 记作(定义为) X ”之意(因为符号 \triangleq 没有左右之分, 含义模糊). 上标 T 表示向量或矩阵转置, 符号 I 表示适当维数单位阵, I_n 表示 n 阶单位阵. 矩阵 X 的范数定义为 $\|X\|^2 = \text{tr}[XX^T]$. $\lambda_{\min}[X]$ 表示对称矩阵 X 的最小特征值, $\lambda_{\max}[X]$ 表示对称矩阵 X 的最大特征值. 关系 $f(t) =$

$o(g(t))$ 表示 $g(t) > 0, \lim_{t \rightarrow \infty} \frac{f(t)}{g(t)} = 0; f(t) = O(g(t))$

表示 $g(t) \geq 0$, 存在常数 $\delta_1 > 0$ 满足 $|f(t)| \leq \delta_1 g(t)$.

辨识的基本问题包括: 1) 新辨识方法的提出; 2) 辨识方法的收敛性分析; 3) 辨识试验中输入信号的设计问题. 第 1 个问题是辨识的基石; 第 2 个是辨识中的理论问题, 涉及参数估计性质、收敛速度等, 是辨识性能分析中的较难课题, 受到了较多的关注; 第 3 个是辨识实践中输入信号的设计问题. 这些都是辨识领域中极具挑战性的研究课题^[8].

辨识方法是系统辨识的基础. 辨识输入信号的设计是为了保证辨识实验所获得的输入输出数据包含系统的本质特征, 也就是说输入信号应该激励系统的所有模态. 这样的输入输出数据包含了足够丰富的信息(或输入输出数据是充分丰富的), 或输入是持续激励信号, 或由输入输出数据构成的回归信息向量是持续激励的(persistently exciting), 以致于存在一个辨识方法能够从输入输出数据中确定系统参数.

有些辨识方法, 特别是时变参数辨识方法, 为了提高跟踪时变参数和克服数据饱和的能力, 有时在辨识方法中引入调节参量, 如遗忘因子最小二乘算法和遗忘梯度算法中的遗忘因子, 卡尔曼滤波算法或协方差修正算法中的修正项, 协方差复位最小二乘算法中的复位间隔, 多新息辨识方法中的迭代间隔和新息长度, 广义投影算法中记忆长度或数据窗长度, 有限数据窗最小二乘法中的数据窗长度等, 这些参量的引入使得其收敛性研究更为困难和复杂^[9].

在众多(时变) 参数辨识方法中, 谁的跟踪性能好, 谁的估计误差小, 谁优谁劣, 如何评判? 作者认为, 应在统一条件下(如持续激励条件、噪声方差相同等), 对各种(时变) 参数估计误差上界进行研究, 并寻求调节参量的最佳值, 以使均方参数估计误差上界最小. 通过比较各种方法的最小参数估计误差上界, 从而作出选择^[9]. 因此, 首要问题是研究各种(时变) 系统辨识方法的收敛性能和估计误差上界. 文献[9]研究了有限数据窗最小二乘法的收敛性及其最小估计误差上界.

与辨识方法性能(收敛性、参数估计收敛速率, 估计误差上界) 有关的还有参数可辨识性问题与系统可辨识性问题, 信号的持续激励性和持续激励条件等.

2.1 辨识方法的提出

一些辨识方法是根据一些基本定理或原理提出

的. 如对于一个线性参数估计问题, 最小二乘辨识算法是根据最小二乘原理, 极小化一个二次准则函数得到的算法; 随机梯度辨识方法和迭代辨识方法是通过梯度搜索或最小二乘搜索得到的算法^[10-12]; 辅助模型辨识方法是借助于辅助模型辨识思想, 用辅助模型输出代替系统的不可测变量的辨识方法^[13-18]; 多新息辨识方法是运用多新息辨识理论, 通过扩展新息长度得到的算法^[8, 19-25]; 递阶辨识方法是基于递阶辨识原理, 通过辨识模型分解和雅可比迭代得到的算法^[26-35]. 对于一个非线性参数估计问题, 可以采用梯度搜索方法、牛顿递推辨识方法和牛顿迭代辨识方法^[36].

有一些辨识方法是通过类比方法得到的. 如多新息辨识方法是本文作者受文献[37]算法间断迭代的启发, 最初用类比方法直接给出了多新息辨识算法的数学表达式^[19], 后来从理论上详细推导, 发展了一系列多新息辨识方法, 如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^[8, 20, 23, 25, 38].

2.2 辨识输入信号设计

开环系统辨识输入信号设计方法可参见文献[5]第3.3.1节. 开环系统辨识实验在对象输入端 $u(t)$ 上叠加一个持续激励信号(图6), 闭环系统辨识实验是在参考输入端 $r(t)$ 上叠加一个持续激励信号(图7). 闭环系统辨识输入信号设计方法还有待进一步研究.

2.3 可辨识性问题

2.3.1 参数可辨识性与系统可辨识性

可辨识性分为参数可辨识性和系统可辨识性. 参数可辨识性指系统数学模型中的每一个参数能根据系统的观测数据确定, 这样的系统称为参数是可辨识的. 系统可辨识性指系统的输入输出关系能够从观测数据确定, 这样的系统就称为系统是可辨识的. 参数可辨识肯定是系统可辨识的, 但系统可辨识不一定是参数可辨识的, 如上面的 R-L-C 电路系统.

考虑抛物线关系所描述的系统

$$y = ax^2 + bx + c. \quad (1)$$

这里 x 可认为是系统输入, y 是系统输出, a, b 和 c 是系统参数. 给定 L 个点 $(x_i, y_i) \quad i=1, 2, \dots, L$, 且至少有 3 个点不在同一直线上, 就可以唯一确定系统参数. 定义准则函数

$$J_1(a, b, c) := \sum_{i=1}^L (y_i - ax_i^2 - bx_i - c)^2 = \sum_{i=1}^L (y_i - \phi_i^T \theta)^2,$$

其中

$$\begin{cases} \phi_i := [x_i^2, x_i, 1]^T \in \mathbf{R}^3, \\ \theta := [a, b, c]^T \in \mathbf{R}^3. \end{cases}$$

极小化 $J_1(a, b, c)$ 给出 θ 的估计:

$$\hat{\theta} = \left[\sum_{i=1}^L \phi_i \phi_i^T \right]^{-1} \left[\sum_{i=1}^L \phi_i y_i \right].$$

当至少有 3 个点不在同一直线上, 矩阵 $\sum_{i=1}^L \phi_i \phi_i^T$ 就是可逆的, 参数估计向量 $\hat{\theta}$ 就有唯一解, 系统参数就是可辨识的. 因此模型(1)是参数可辨识的, 同时也是系统可辨识的.

在上述条件下, 考虑 4 个参数 (a, b, c, d) 的抛物线系统

$$y = ax^2 + bx + 3c + d. \quad (2)$$

如果令

$$\begin{cases} \phi_i := [x_i^2, x_i, 3, 1]^T \in \mathbf{R}^4, \\ \vartheta := [a, b, c, d]^T \in \mathbf{R}^4. \end{cases} \quad (3)$$

定义和极小化准则函数,

$$J_2(a, b, c, d) := \sum_{i=1}^L (y_i - ax_i^2 - bx_i - 3c - d)^2 = \sum_{i=1}^L (y_i - \phi_i^T \vartheta)^2,$$

给出 ϑ 的估计:

$$\hat{\vartheta} = \left[\sum_{i=1}^L \phi_i \phi_i^T \right]^{-1} \left[\sum_{i=1}^L \phi_i y_i \right].$$

即使有更多的点不在同一直线上, 矩阵 $\sum_{i=1}^L \phi_i \phi_i^T$ 也不是可逆的(当这个矩阵是可逆时, 就说 ϕ_i 是持续激励的, 或者说输入信号是持续激励信号), 参数估计向量 $\hat{\vartheta}$ 无解, 故系统参数不是可辨识的或系统参数是不可辨识的(只要有一个参数不可辨识, 这样的系统就是参数不可辨识的). 但这个系统是系统可辨识的, 即输入输出关系是可辨识的, 只需令

$$\begin{cases} \phi_i := [x_i^2, x_i, 1]^T \in \mathbf{R}^3, \\ \vartheta := [a, b, 3c + d]^T \in \mathbf{R}^3. \end{cases}$$

那么 ϑ 的估计为

$$\hat{\vartheta} = \left[\sum_{i=1}^L \phi_i \phi_i^T \right]^{-1} \left[\sum_{i=1}^L \phi_i y_i \right].$$

在这种情况下, 矩阵 $\sum_{i=1}^L \phi_i \phi_i^T$ 是可逆的(很简

单,读者试图证明之) 参数估计向量 $\hat{\theta} \in \mathbf{R}^3$ 有唯一解 故系统参数向量 $\theta = [a \ b \ 3c + d]^T$ 是可辨识的, 即 a b 和 $3c + d$ 是可辨识的, 但 c 和 d 是不可辨识的, 因此模型(2) 是系统可辨识的, 但不是参数可辨识的.

下面的例子是参数可辨识的和系统可辨识的,

$$y = ax^2 + b \ln x + c;$$

下面的例子不是参数可辨识的, 但是系统可辨识的,

$$y = ax^2 + b \ln x + \frac{c}{d+1},$$

$$y = ax^2 + b \ln x + c \ln x + \frac{1}{d+1}.$$

2.3.2 可辨识性与辨识算法的关系

从上述可辨识性定义看,可辨识性依赖于系统的观测数据,即用什么样的输入信号激励系统产生的输入输出观测数据.可辨识性不依赖参数辨识(估计)算法,辨识算法只是提供了一种检验可辨识性的判断依据,辨识算法不能作为判断系统不可辨识性的依据.也就是说,一个辨识算法能通过观测数据确定系统参数,系统就是参数可辨识的;一个辨识算法不能通过观测数据确定系统参数,不能说系统是参数不可辨识的.

可辨识性与系统的输入输出数据有很大关系.选择“良好”的输入信号,使得激励系统产生的输出信号也是“良好”的,我们期望这种“良好”的输入输出信号具有某种特征,使得系统是可辨识的.这种“良好”的输入信号定义为持续激励信号.由持续激励输入信号和系统输出信号构成的回归信息向量是持续激励的,那么系统是可辨识的.

在上面的系统(2)中,由于构建了式(3)中不恰当的信息向量 ϕ_i ,导致数据乘积矩阵 $\sum_{i=1}^L \phi_i \phi_i^T$ 不可逆,因为这个矩阵中有2行(或2列)是线性比例关系,故参数向量是不可辨识的.例如对于有限脉冲响应(FIR)模型,

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + b_3 u(t-3) = \varphi^T(t) \theta,$$

假设输入信号 $\{u(t) : t = 1, 2, \dots\}$ 是持续激励的,那么这个系统是参数可辨识的(即下式中参数向量 $\theta \in \mathbf{R}^3$ 是可辨识的),参数向量与信息向量分别为

$$\theta = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbf{R}^3, \quad \varphi(t) = \begin{bmatrix} u(t-1) \\ u(t-2) \\ u(t-3) \end{bmatrix} \in \mathbf{R}^3.$$

然而,如果重新定义参数向量和信息向量

$$\theta = \begin{bmatrix} b_1 \\ b_2/2 \\ b_2/2 \\ b_3 \end{bmatrix} \in \mathbf{R}^4, \quad \varphi(t) = \begin{bmatrix} u(t-1) \\ u(t-2) \\ u(t-2) \\ u(t-3) \end{bmatrix} \in \mathbf{R}^4,$$

那么这个维数为4的参数向量 $\theta \in \mathbf{R}^4$ 是不可辨识的.因此,如何参数化一个系统,得到一个可辨识的辨识模型是重要的.

只有在可辨识的前提下,估计算法给出的参数估计才是有效的.由于事先无法知道系统是否是可辨识的,因此只能在给定输入输出量测数据下,分析辨识算法参数估计是否收敛于真参数,来判断参数可辨识性和系统可辨识性,这就引出了辨识方法的收敛性分析或参数估计性能分析研究课题.

2.3.3 开环可辨识性与闭环可辨识性

为了使系统是可辨识的,输入信号必须满足一定的条件.一个最低的要求是在整个观测周期上系统的所有模态必须被输入信号持续激励(这引出了辨识输入信号设计课题).如果系统有模态没有被激励,系统将是不可辨识的.

1) 开环可辨识性.对于开环系统,对象的输入信号与系统输出无关,可以进行独立设计,这样的系统可辨识性称为开环系统可辨识性.对于图6所示的开环系统, $u(t)$ 和 $y(t)$ 为系统的输入和输出, $v(t)$ 为随机干扰噪声, $G(z)$ 是对象的传递函数.开环系统的输入可以影响输出,但输出不影响输入,故开环辨识输入信号可以独立设计.最简单的方法是采用零均值可测白噪声序列,或伪随机二进制(PRBS)序列.在Matlab中,可用 `rand`, `randn`, `idinput` 等函数产生输入信号或仿真用的干扰随机噪声.



图6 开环系统

Fig.6 The open-loop system

2) 开环辨识试验.即系统在正常运行情况下(稳定工作点),这时系统的所有变量认为是零,相当于零均值化,在系统输入端 $u(t)$ 上叠加一个不影响系统正常运行的输入信号(持续激励信号),或改变对象的输入,但改变幅度不影响系统正常运行,通过量测系统的输入和输出数据的改变量,测得实验数

据来研究其辨识问题.

3) 闭环可辨识性. 对于图 7 所示的闭环系统, $G_c(z)$ 是控制器的传递函数, $H(z)$ 是反馈通道的传递函数, $r(t)$ 是设定值(参考输入). 由于闭环系统存在反馈回路, 对象 $G(z)$ 的输入 $u(t)$ 与输出 $y(t)$ 通过反馈回路发生关系, 它们的关系为

$$u(t) = G_c(z) [r(t) - H(z)y(t)],$$

即输入信号与输出信号是相关的, 输入信号不可能独立设计(这引出了闭环系统辨识输入信号设计课题). 参考输入信号 $r(t)$ 通常是一恒定值(常数), 不具备持续激励的条件(不是一个“好”的输入信号), 如果忽略噪声 $v(t)$, 那么这样的闭环系统是不可辨识的.

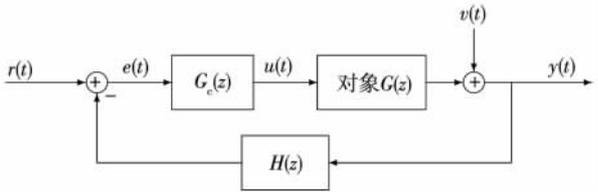


图 7 闭环系统

Fig. 7 The closed-loop system

4) 闭环辨识试验. 即系统在平稳运行情况下(系统稳定时, 系统各变量认为是零, 相当于零均值化), 在参考输入 $r(t)$ 上叠加一个不影响系统正常运行的激励信号, 或改变 $r(t)$, 通过量测对象的输入信号 $u(t)$ 和输出信号 $y(t)$, 测得实验数据来研究闭环系统的辨识问题. 就量测数据 $\{u(t), y(t)\}$ 来说, 对象 $G(z)$ 的参数辨识方法与开环辨识方法没有不同, 但关键是输入 $u(t)$ 与输出 $y(t)$ 是相关的, 使得闭环辨识方法的收敛性分析更加困难.

2.3.4 可辨识性与能控性和能观测性的关系

可辨识性与系统的能控性和能观测性有着密切的联系. 辨识是利用系统外部一些可测量来获取系统的数学模型, 它只能反映系统的外特性关系; 对系统内部一些不能观测或不能控状态的特性, 辨识结果是无法描述的. 换句话说, 输入与输出数据的关系是系统的外部特征, 描述系统的外特性关系, 在可辨识条件下, 辨识出的模型是能控能观测子系统(最小实现的状态空间模型或不可约传递函数模型). 系统的不能控模态和(或)不能观测模态是不可辨识的; 不能控和(或)不能观测系统是不可辨识的.

上述论断假设系统状态不可测, 如果系统内部状态是可测的, 情况就不同了. 因此, 讨论随机系统

状态空间模型的辨识问题, 要考虑状态可测和不可测 2 种情况.

考虑单输入单输出状态空间模型,

$$\begin{cases} x(t+1) = Ax(t) + bu(t), \\ y(t) = cx(t) + du(t). \end{cases} \quad (4)$$

其中: $x(t) \in \mathbf{R}^n$ 是 n 维状态变量, $u(t) \in \mathbf{R}$ 和 $y(t) \in \mathbf{R}$ 分别为系统的输入和输出(可得到的观测数据). $A \in \mathbf{R}^{n \times n}$, $b \in \mathbf{R}^n$, $c \in \mathbf{R}^{1 \times n}$ 和 $d \in \mathbf{R}$ 是待辨识参数矩阵或参数向量.

如果根据系统的输入输出或状态变量的信息可以唯一地确定未知参数 (A, b, c, d) , 则系统参数是可辨识的; 如果不能唯一地确定参数 (A, b, c, d) , 但可以确定输入输出间的传递关系, 则是系统可辨识的.

1) 状态可测情形. 下面讨论状态可测, 即状态已知情形参数矩阵的辨识问题. 假设输入信号 $u(t)$, 输出信号 $y(t)$ 和系统状态向量 $x(t)$ 都是可测的, 那么(4)就可写为系统参数 (A, b, c, d) 的线性回归模型:

$$Y(t) = \theta^T \varphi(t). \quad (5)$$

其中

$$Y(t) = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} \in \mathbf{R}^{n+1}, \quad \theta^T = \begin{bmatrix} A & b \\ c & d \end{bmatrix} \in \mathbf{R}^{(n+1) \times (n+1)},$$

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

如果 $\varphi(t)$ 是持续激励的, 故递推最小二乘算法可以用于估计(5)的未知参数矩阵 θ . 因此, 状态可测系统的状态空间模型的可辨识性, 与系统的能控性能观测性无关.

2) 状态不可测情形. 如果能控性矩阵 $Q_c = [b, Ab, A^2b, \dots, A^{n-1}b] \in \mathbf{R}^{n \times n}$ 满秩, 则系统是能控的, 或说 (A, b) 是能控的, 即可以找到一个控制 $u(t)$, 在有限的时间内, 使系统从任意给定的初始状态到达任意指定的终态; 如果能观测性矩阵

$$Q_o = \begin{bmatrix} c \\ cA \\ cA^2 \\ \vdots \\ cA^{n-1} \end{bmatrix} \in \mathbf{R}^{n \times n}$$

满秩, 则系统是能观测的, 或说 (c, A) 是能观测的, 即根据输入输出观测数据可以确定系统的所有状态. 简单地说, 能控就是控制输入 $u(t)$ 可以确定状态, 能观测就是输出能够确定状态. 因此, 输入输出

数据只能确定系统既能控又能观测子系统的外部特性;不能控和(或)不能观测的子系统是不可辨识的.如果能控性矩阵或能观测性矩阵不满秩,系统的外部描述仅依存于那些能控能观测的状态,所以 $(A \ b \ c)$ 中那些属于不能控或不能观测状态子空间的未知参数利用系统的外部可测信号是无法确定的.

3) 输入输出表达.系统(4)对应的输入输出表达的算子描述(差分方程模型)为

$$y(t) = \frac{b(z)}{a(z)}u(t) \text{ 或 } a(z)y(t) = b(z)u(t), \quad (6)$$

其中 $a(z)$ 和 $b(z)$ 均为单位后移算子 z^{-1} 的多项式 $[z^{-1}y(t) = y(t-1)]$,定义为

$$\begin{aligned} a(z) &:= z^{-n} \det [zI - A] = \\ &1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}, \\ b(z) &:= z^{-n} \text{cadj} [zI - A] b + da(z) = \\ &b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_n z^{-n}. \end{aligned}$$

如果 $a(z)$ 和 $b(z)$ 无公因子(对应于状态空间模型是最小实现)根据系统的输入输出数据可以唯一确定多项式 $a(z)$ 和 $b(z)$ 的未知参数 $(a_i \ b_i)$,系统(6)是参数可辨识的.换句话说,如果系统(4)既能控又能观测,那么系统(6)中多项式 $a(z)$ 和 $b(z)$ 不存在公因子(是互质的),系统(6)就是参数可辨识的, $2n+1$ 个参数都是可辨识的.

例2 考虑下列二阶系统

$$\begin{cases} \mathbf{x}(t+1) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \\ y(t) = [1 \ 0] \mathbf{x}(t), \end{cases}$$

系统的能控性矩阵为

$$Q_c = [b \ Ab] = \begin{bmatrix} 1 & a_{11} \\ 0 & a_{21} \end{bmatrix},$$

系统的能观测矩阵为

$$Q_o = \begin{bmatrix} c \\ cA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a_{11} & a_{12} \end{bmatrix}.$$

当 $a_{21} \neq 0$ $a_{12} \neq 0$ 时 Q_c 和 Q_o 是非奇异的,故系统既能控又能观测,但如果状态 $\mathbf{x}(t)$ 不可测,系统参数是不可辨识的.事实上,该系统的输入输出表达为

$$\begin{aligned} y(t) &= c(zI - A)^{-1} b u(t) = \\ [1 \ 0] &\begin{bmatrix} z - a_{11} & -a_{12} \\ -a_{21} & z - a_{22} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) = \\ [1 \ 0] &\frac{\begin{bmatrix} z - a_{22} & a_{12} \\ a_{21} & z - a_{11} \end{bmatrix}}{(z - a_{11})(z - a_{22}) - a_{12}a_{21}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) = \end{aligned}$$

$$\frac{z - a_{22}}{z^2 - (a_{11} + a_{22})z + a_{11}a_{22} - a_{12}a_{21}} u(t) = \frac{z^{-1} - a_{22}z^{-2}}{1 - (a_{11} + a_{22})z^{-1} + (a_{11}a_{22} - a_{12}a_{21})z^{-2}} u(t),$$

或

$$(1 - (a_{11} + a_{22})z^{-1} + (a_{11}a_{22} - a_{12}a_{21})z^{-2})y(t) = (z^{-1} - a_{22}z^{-2})u(t).$$

写成差分方程为

$$y(t) - (a_{11} + a_{22})y(t-1) + (a_{11}a_{22} - a_{12}a_{21})y(t-2) = u(t-1) - a_{22}u(t-2).$$

对于这样的线性系统,最小二乘方法可以估计出3个参数 $-(a_{11} + a_{22})$ 、 $(a_{11}a_{22} - a_{12}a_{21})$ 和 $-a_{22}$ (注: $u(t-1)$ 的系数1是已知的,不用估计).只有参数 a_{22} 和 a_{11} 是可辨识的,参数 a_{12} 和 a_{21} 是不可辨识的,但系统是可辨识的.

上述分析表明,能控性能观测性并不意味着系统参数是可辨识的.对于状态空间模型,当状态不可测时,通常采用规范状态模型进行辨识.

读者可以研究下列一般二阶系统的能控性能观测性与可辨识性的关系,

$$\begin{cases} \mathbf{x}(t+1) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} u(t), \\ y(t) = [c_1 \ c_2] \mathbf{x}(t) + v(t). \end{cases}$$

对于单输入单输出系统,状态空间描述(4)中有 $n^2 + 2n + 1 = (n+1)^2$ 个未知参数(A 中 n^2 个参数, b 中 n 个参数, c 中 n 个参数, d 中1个参数).假设系统既能控又能观测,如果系统状态未知,那么状态空间模型(4)是系统可辨识的,但不是参数可辨识的,而能控能观测状态空间模型(4)对应的能控性或能观测性规范型中只有 $2n+1$ 个参数,是参数可辨识的.能控能观测规范状态空间模型具有与输入输出表达相同的参数数目,故是参数可辨识的(对多变量系统也是如此),而对于状态不可测的非规范型的状态空间系统,不管系统是否可控和可观测,系统不是参数可辨识的.

2.3.5 可辨识性与输入信号之间的关系

可辨识性是系统的一种特征.对于零初值状态空间模型,不可控和(或)不可观测子系统是不可辨识的.可辨识性与系统的输入密切相关.对于零初值状态空间模型,如果输入信号选择得不好,可控可观测状态空间系统也可能是不可辨识的.下面讨论可辨识性与输入信号之间的关系.

例3 考虑如下自回归系统

$$x(t+1) = Ax(t), \quad (7)$$

根据 n 维状态向量 $x(t) \in \mathbf{R}^n$ 的测量值可以唯一确定矩阵 A 则系统是可辨识的(参数矩阵 $A \in \mathbf{R}^{n \times n}$ 也是可辨识的). 由式(7)可得

$$\begin{cases} x(1) = Ax(0), \\ x(2) = Ax(1) = A^2x(0), \\ \vdots \\ x(n) = Ax(n-1) = A^n x(0). \end{cases}$$

可以等价写为

$$\begin{bmatrix} x(1) & x(2) & \cdots & x(n) \end{bmatrix} = A \begin{bmatrix} x(0) & x(1) & \cdots & x(n-1) \end{bmatrix}, \quad (8)$$

或

$$\begin{bmatrix} x(1) & x(2) & \cdots & x(n) \end{bmatrix} = A \begin{bmatrix} x(0) & Ax(0) & \cdots & A^{n-1}x(0) \end{bmatrix}. \quad (9)$$

从式(8)可以看出,如果量测的状态向量 $x(t)$ 构成的矩阵 $[x(0) \ x(1) \ \cdots \ x(n-1)] \in \mathbf{R}^{n \times n}$ 满秩,那么参数矩阵 A 就是可辨识的,且有

$$A = [x(1) \ x(2) \ \cdots \ x(n)] [x(0) \ x(1) \ \cdots \ x(n-1)]^{-1} = [x(1) \ x(2) \ \cdots \ x(n)] [x(0) \ Ax(0) \ \cdots \ A^{n-1}x(0)]^{-1}. \quad (10)$$

我们要找出系统的可辨识条件,就是要搞清楚“ $[x(0) \ x(1) \ \cdots \ x(n-1)] \in \mathbf{R}^{n \times n}$ 满秩”意味着什么?注意到 $[x(0) \ x(1) \ \cdots \ x(n-1)] = [x(0) \ Ax(0) \ \cdots \ A^{n-1}x(0)]$. 从上式看,矩阵 A 可辨识的条件是 A 可辨识性矩阵 $[x(0) \ Ax(0) \ \cdots \ A^{n-1}x(0)] \in \mathbf{R}^{n \times n}$ 非奇异(满秩),其中 $x(0)$ 为自回归系统的初始值,相当于激励系统的输入信号. 如果存在向量 $b \in \mathbf{R}^n$ 作为初值 $x(0)$,且 $(A \ b)$ 是可控的,那么有 $[b \ Ab \ \cdots \ A^{n-1}b] = [x(0) \ Ax(0) \ \cdots \ A^{n-1}x(0)]$ 是满秩的,系统是可辨识的.

系统可辨识就是要输入信号激励系统的所有模态,可辨识性矩阵 $[x(0) \ Ax(0) \ \cdots \ A^{n-1}x(0)]$ 必须非奇异. 如果初始状态 $x(0) = 0$,相当于激励系统的输入信号为零,则系统是不可辨识的,说明静止的系统是不可辨识的.

假设矩阵 A 有 n 个不同的非零特征值 $\lambda_i \ i=1, 2, \cdots, n$, 对应 n 个独立的特征向量 $r_1 \ r_2 \ \cdots \ r_n$, 矩阵 $R := [r_1 \ r_2 \ \cdots \ r_n] \in \mathbf{R}^{n \times n}$ 非奇异. 如果 $x(0)$ 取矩阵 A 的某个特征值 λ_j 对应的特征向量 r_j , 即 $x(0) = r_j$ (或取 $x(0) = \alpha r_j$, 情形是一样的,其中 α 为任意非零常数),根据特征值与特征向量的关系 $Ar_j = \lambda_j r_j$, 有

$$Ar_j = \lambda_j r_j, \ A^2 r_j = A \lambda_j r_j = \lambda_j^2 r_j, \ \cdots \ A^{n-1} r_j = \lambda_j^{n-1} r_j.$$

或

$$Ax(0) = \lambda_j x(0),$$

$$A^2 x(0) = \lambda_j^2 x(0), \ \cdots \ A^{n-1} x(0) = \lambda_j^{n-1} x(0).$$

则可辨识性矩阵

$$\begin{bmatrix} x(0) & Ax(0) & \cdots & A^{n-1}x(0) \end{bmatrix} = \begin{bmatrix} x(0) & \lambda_j x(0) & \cdots & \lambda_j^{n-1} x(0) \end{bmatrix}$$

是奇异矩阵(其各列是线性相关的),因此系统是不可辨识的. 这意味着 $x(0) = r_j$, $x(0)$ 只激励系统的模态 λ_j^i (因为状态解中包括此项),没有激励系统的其他模态 $\lambda_i^i (i \neq j)$,使系统成为不可辨识的(注:在连续系统中模态为 $\exp(\lambda_j t)$, λ_j 为连续系统的特征值).

如果取 $x(0)$ 为所有特征向量的线性组合,即 $x(0) = \alpha_1 r_1 + \alpha_2 r_2 + \cdots + \alpha_n r_n$, α_i 均为任意非零常数,则有

$$A^i x(0) = \alpha_1 A^i r_1 + \alpha_2 A^i r_2 + \cdots + \alpha_n A^i r_n = \alpha_1 \lambda_1^i r_1 + \alpha_2 \lambda_2^i r_2 + \cdots + \alpha_n \lambda_n^i r_n.$$

因此,可辨识性矩阵

$$\begin{bmatrix} x(0) & Ax(0) & \cdots & A^{n-1}x(0) \end{bmatrix} = \begin{bmatrix} \alpha_1 r_1 + \alpha_2 r_2 + \cdots + \alpha_n r_n & \alpha_1 \lambda_1 r_1 + \alpha_2 \lambda_2 r_2 + \cdots + \alpha_n \lambda_n r_n & \cdots & \alpha_1 \lambda_1^{n-1} r_1 + \alpha_2 \lambda_2^{n-1} r_2 + \cdots + \alpha_n \lambda_n^{n-1} r_n \end{bmatrix} =$$

$$[r_1 \ r_2 \ \cdots \ r_n] \left\{ \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, \begin{bmatrix} \alpha_1 \lambda_1 \\ \alpha_2 \lambda_2 \\ \vdots \\ \alpha_n \lambda_n \end{bmatrix}, \cdots, \begin{bmatrix} \alpha_1 \lambda_1^{n-1} \\ \alpha_2 \lambda_2^{n-1} \\ \vdots \\ \alpha_n \lambda_n^{n-1} \end{bmatrix} \right\} = R \begin{bmatrix} \alpha_1 & \alpha_1 \lambda_1 & \alpha_1 \lambda_1^2 & \cdots & \alpha_1 \lambda_1^{n-1} \\ \alpha_2 & \alpha_2 \lambda_2 & \alpha_2 \lambda_2^2 & \cdots & \alpha_2 \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \alpha_n \lambda_n & \alpha_n \lambda_n^2 & \cdots & \alpha_n \lambda_n^{n-1} \end{bmatrix}$$

是非奇异的,因为由独立特征向量构成的矩阵 R 是非奇异的,非零常数 α_i 构成的对角阵是非奇异的,上式最后一个由不同特征值构成的范德蒙矩阵(Vandermonde matrix)也是非奇异的. 在取 $x(0)$ 为所有特征向量的线性组合的条件下,且线性组合的系数都非零,可辨识矩阵是非奇异的,参数矩阵 A 是可辨识的.

简单取 $x(0) = r_1 + r_2 + \cdots + r_n$, 系统参数矩阵 A 也是可辨识的;相反,取 $x(0) = \alpha_1 r_1 + \alpha_2 r_2 + \cdots + \alpha_{n-1} r_{n-1}$, 则参数矩阵 A 是不可辨识的. 当系统有重特征值时 A 是可辨识的条件值得进一步研究.

这个例子说明,系统可辨识的条件是系统的所有模态都必须被输入信号所激励. 这就引出持续激

励输入信号的概念. 对上述 $2n + 1$ 个参数的单输入单输出系统 (6) 来说, 可辨识性的充分条件是输入信号至少为 $2n + 1$ 阶持续激励条件. 下节讨论输入信号与持续激励条件的关系.

2.4 辨识算法的收敛性能

考虑观测变量 y, x_1, x_2, \dots, x_n . 如果 y 是 x_1, x_2, \dots, x_n 的线性组合, 那么存在一组参数 $\theta_1, \theta_2, \dots, \theta_n$ 使得下式成立,

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n.$$

这是确定性情形, 即每个测量变量都是准确的. 通常测量是有误差的, 为了平衡方程, 在上式右边加入噪声项 v 得到

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n + v.$$

对于 $t = 1, 2, \dots, t$ 时, 可得到观测 y 和 x_i , 记作 $y(t)$ 和 $x_i(t)$, 对应的误差记作 $v(t)$, 因此可得到 t 个方程式 (一般观测数据长度 t 应远大于参数数目, 即拟合阶次 $n: t \gg n$),

$$\begin{cases} y(1) = \theta_1 x_1(1) + \theta_2 x_2(1) + \dots + \theta_n x_n(1) + v(1), \\ y(2) = \theta_1 x_1(2) + \theta_2 x_2(2) + \dots + \theta_n x_n(2) + v(2), \\ \vdots \\ y(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \dots + \theta_n x_n(t) + v(t). \end{cases} \quad (11)$$

这些方程也可以简单写作

$$y(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \dots + \theta_n x_n(t) + v(t), \quad t = 1, 2, 3, \dots, t. \quad (12)$$

$\{v(t)\}$ 可看作零均值方差为 σ^2 的随机白噪声序列, n 是拟合模型阶次或模型参数数目 (假设已知), 且当 $t \leq 0$ 时 $y(t) = 0, x_i(t) = 0, v(t) = 0$.

定义参数向量 (parameter vector)

$$\theta = [\theta_1, \theta_2, \dots, \theta_n]^T \in \mathbf{R}^n$$

和信息向量 (information vector)

$$\varphi(t) := [x_1(t), x_2(t), \dots, x_n(t)]^T \in \mathbf{R}^n.$$

参照 θ 和 $\varphi(t)$ 的定义, 式 (12) 可以写为

$$y(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \dots + \theta_n x_n(t) + v(t) =$$

$$[x_1(t), x_2(t), \dots, x_n(t)] \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} + v(t) =$$

$$\varphi^T(t) \theta + v(t). \quad (13)$$

式 (13) 是一个线性回归模型, 在系统辨识中称为辨识模型 (identification model) 或辨识表达式 (identification representation or expression).

式 (11) 可以写为

$$\begin{cases} y(1) = \varphi^T(1) \theta + v(1), \\ y(2) = \varphi^T(2) \theta + v(2), \\ \vdots \\ y(t) = \varphi^T(t) \theta + v(t). \end{cases} \quad (14)$$

当然, 式 (14) 也可在式 (13) 中令 $t = 1, 2, \dots, t$ 得到. 式 (14) 写成矩阵形式为

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(t) \end{bmatrix} = \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(t) \end{bmatrix} \theta + \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(t) \end{bmatrix},$$

也可简单写作

$$Y_t = H_t \theta + V_t. \quad (15)$$

式 (15) 中

$$Y_t := \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(t) \end{bmatrix} \in \mathbf{R}^t, \quad H_t := \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(t) \end{bmatrix} \in \mathbf{R}^{t \times n},$$

$$V_t := \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(t) \end{bmatrix} \in \mathbf{R}^t.$$

最小二乘辨识原理 (least squares identification principle) 就是利用系统的观测数据 $\{y(t), \varphi(t)\}$, 极小下列二次准则函数 (quadratic criterion function):

$$J(\theta) := \sum_{i=1}^t v^2(i) = \sum_{i=1}^t [y(i) - \varphi^T(i) \theta]^2 = V_t^T V_t = (Y_t - H_t \theta)^T (Y_t - H_t \theta),$$

得到参数向量 θ 的估计. 设 $\theta = \hat{\theta}_{LS}$ 时, $\min J(\theta) = J(\hat{\theta}_{LS})$. 令 $J(\theta)$ 对 θ 的偏导数为零, 得到

$$\frac{\partial J(\theta)}{\partial \theta} \Big|_{\theta = \hat{\theta}_{LS}} = -2H_t^T (Y_t - H_t \theta) \Big|_{\theta = \hat{\theta}_{LS}} = 0,$$

或

$$(H_t^T H_t) \hat{\theta}_{LS} = H_t^T Y_t. \quad (16)$$

这个方程称为规范方程 (normal equation) 或正则方程. 在持续激励条件假设下, 数据长度 t 足够大, $(H_t^T H_t)$ 为正定矩阵时, 由式 (16) 可以求得

$$\hat{\theta}_{LS} = (H_t^T H_t)^{-1} H_t^T Y_t. \quad (17)$$

式 (17) 给出的估计 $\hat{\theta}_{LS}$ 称为最小二乘估计 (Least Squares Estimate, LSE), 简称 LS 估计. 由于它与数据长度 t 有关, 故可等价写作 $\hat{\theta}_{LS}(t)$, 在递推算法中记作 $\hat{\theta}(t)$. 将 H_t 和 Y_t 的定义式代入式 (17) 展开可得

$$\hat{\theta}(t) = \hat{\theta}_{LS}(t) = (H_t^T H_t)^{-1} H_t^T Y_t =$$

$$\left[\sum_{i=1}^t \varphi(i) \varphi^T(i) \right]^{-1} \left[\sum_{i=1}^t \varphi(i) y(i) \right]. \quad (18)$$

式(18)就是一次完成最小二乘算法,也称离线最小二乘辨识算法,以区别后面的在线辨识算法。

关于离线算法(17)给出的LS估计 $\hat{\theta}_{LS}$ 的统计性质,有如下定理^[39-40]。

定理1(无偏性定理) 对于系统(15),设噪声向量 V_t 的均值为零,且 V_t 和 H_t 是统计独立的(statistically independent),则最小二乘参数估计 $\hat{\theta}_{LS}$ 是 θ 的无偏估计(unbiased estimate),即 $E[\hat{\theta}_{LS}] = \theta$ 。

无偏性是衡量参数估计是否围绕真参数波动的一个统计性质。无偏性只要求 H_t 与 V_t 统计独立,并不要求 V_t 是不相关白噪声向量(uncorrelated white noise vector)。

定理2(估计误差协方差定理) 对于系统(15),设噪声向量 V_t 的均值为零,协方差阵 $\text{cov}[V_t] = \Sigma_v$,且 V_t 与 H_t 是统计独立的,则参数估计偏差 $\tilde{\theta}_{LS} = \hat{\theta}_{LS} - \theta$ 的协方差阵为

$$\text{cov}[\tilde{\theta}_{LS}] = E[(H_t^T H_t)^{-1} H_t^T \Sigma_v H_t (H_t^T H_t)^{-1}].$$

如果 $\{v(t)\}$ 是零均值方差 σ^2 的白噪声序列,即 $E[v(t)] = 0$, $E[v^2(t)] = \sigma^2$,就有 $\text{cov}[V_t] = \Sigma_v = \sigma^2 I_t$, I_t 为 t 阶单位阵,参数估计偏差(估计误差)协方差阵简化为

$$\text{cov}[\tilde{\theta}_{LS}] = E[(H_t^T H_t)^{-1} H_t^T \sigma^2 H_t (H_t^T H_t)^{-1}] = \sigma^2 E[(H_t^T H_t)^{-1}],$$

两边取迹得到

$$E[\|\hat{\theta}_{LS} - \theta\|^2] = \text{tr}\{\text{cov}[\tilde{\theta}_{LS}]\} = \sigma^2 \text{tr}\{E[(H_t^T H_t)^{-1}]\}.$$

$\tilde{\theta}_{LS}$ 的协方差阵 $\text{cov}[\hat{\theta}_{LS}]$ 主对角线各元是 $\tilde{\theta}_{LS}$ 各分量的方差,它可以衡量参数估计的散度;而非对角线上各元可以用来衡量参数估计 $\tilde{\theta}_{LS}$ 各分量相互影响的程度或相关性的大小。从上式可知,随着数据长度 t 增加,如果信息向量的乘积矩阵 $(H_t^T H_t)$ 的最小特征趋于无穷,那么均方参数估计误差收敛于零。如果 $(H_t^T H_t) \geq C I_n \ln \ln t$ (对于常数 $C > 0$),就有

$$\lim_{t \rightarrow \infty} E[\|\hat{\theta}_{LS} - \theta\|^2] = \lim_{t \rightarrow \infty} \sigma^2 \text{tr}\{E[(H_t^T H_t)^{-1}]\} \leq \lim_{t \rightarrow \infty} \sigma^2 \text{tr}\{E[C I_n \ln \ln t]^{-1}\} = \lim_{t \rightarrow \infty} \frac{n \sigma^2}{C \ln \ln t} = 0. \quad (19)$$

定理3(一致收敛定理) 对于系统(15),设 V_t 是与 H_t 统计独立的白噪声向量,即

$$\begin{cases} E[v(t)] = 0, & E[v(t)v(i)] = 0, \quad t \neq i, \\ E[v^2(t)] = \sigma^2, & \text{cov}[V_t] = \sigma^2 I_t, \end{cases}$$

且存在常数 $0 < \alpha \leq \beta$ 使得下列弱持续激励条件(Weak Persistent Excitation condition, WPE)成立,

$$(WPE) \quad \alpha I \leq \frac{1}{t} (H_t^T H_t) \leq \beta I \quad \text{a. s.} \quad \text{对充分大 } t,$$

那么式(17)给出的最小二乘估计 $\tilde{\theta}_{LS}$ 是一致收敛的,即 $\tilde{\theta}_{LS}$ 依概率1(wp1)或几乎必然(almost surely, a. s.)收敛于真参数向量 θ ,用关系式可表达为

$$\lim_{t \rightarrow \infty} \hat{\theta}_{LS} = \theta \quad \text{a. s.}$$

证明 由于 V_t 是不相关白噪声向量,利用定理2的结论,有

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow \infty} \text{cov}[\tilde{\theta}_{LS}] = \lim_{t \rightarrow \infty} \sigma^2 E[(H_t^T H_t)^{-1}] = \\ &\lim_{t \rightarrow \infty} \frac{\sigma^2}{t} E\left[\left(\frac{1}{t} H_t^T H_t\right)^{-1}\right] \leq \lim_{t \rightarrow \infty} \frac{\sigma^2}{t} \frac{I}{\alpha} = 0, \\ \lim_{t \rightarrow \infty} \text{cov}[\tilde{\theta}_{LS}] &\geq \lim_{t \rightarrow \infty} \frac{\sigma^2}{t} \frac{I}{\beta} = 0. \end{aligned}$$

又由定理1知 $E[\tilde{\theta}_{LS}] = 0$,故 $\lim_{t \rightarrow \infty} \hat{\theta}_{LS} = \theta$, a. s. 定理证毕。

定理4(噪声方差估计定理) 对于系统(15),设 V_t 是与 H_t 统计独立的零均值白噪声向量,即 $v(t)$ 为零均值随机白噪声:

$$\begin{cases} E[v(t)] = 0, & E[v(t)v(i)] = 0, \quad t \neq i, \\ E[v^2(t)] = \sigma^2, & \text{cov}[V_t] = \sigma^2 I_t. \end{cases}$$

参数向量 θ 的最小二乘估计 $\tilde{\theta}_{LS}$ 取式(17),即

$$\hat{\theta}_{LS}(t) = (H_t^T H_t)^{-1} H_t^T Y_t. \quad (20)$$

对应的准则函数值为

$$J[\hat{\theta}_{LS}(t)] = [Y_t - H_t \hat{\theta}_{LS}(t)]^T [Y_t - H_t \hat{\theta}_{LS}(t)],$$

则平稳随机噪声 $v(t)$ 的方差 σ^2 的估计为

$$\hat{\sigma}^2 = \frac{J[\hat{\theta}_{LS}(t)]}{t-n}, \quad t \text{ 充分大时.}$$

式中 $n = \dim \theta = \theta$ 的维数。

证明 定义输出残差 $\varepsilon(t)$ 和残差向量 ε_t 分别为

$$\varepsilon(i) = y(i) - \varphi^T(i) \hat{\theta}_{LS}(t) \quad i=1, 2, \dots, t,$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \vdots \\ \varepsilon(t) \end{bmatrix} = Y_t - H_t \hat{\theta}_{LS}(t).$$

将式(20)代入上式,并利用式(15)可得

$$\varepsilon_t = Y_t - H_t (H_t^T H_t)^{-1} H_t^T Y_t =$$

$$\begin{aligned} [I_t - H_t(H_t^T H_t)^{-1} H_t^T] Y_t &= \\ [I_t - H_t(H_t^T H_t)^{-1} H_t^T] [H_t \theta + V_t] &= \\ [I_t - H_t(H_t^T H_t)^{-1} H_t^T] V_t &=: QV_t, \end{aligned}$$

式中 $Q = I_t - H_t(H_t^T H_t)^{-1} H_t^T$. 因为 $Q^2 = Q$ $Q^T = Q$, 即矩阵 Q 为对称幂等矩阵, 所以

$$E[\varepsilon_t^T \varepsilon_t] = E[V_t^T Q Q V_t] = E[V_t^T Q V_t].$$

利用公式

$$\begin{cases} x^T A x = \text{tr}[A x x^T], \\ E\{\text{tr}[A] \} = \text{tr}\{E[A]\}, \\ \text{tr}[A^T] = \text{tr}[A], \end{cases}$$

由假设 V_t 为不相关白噪声向量, 且与 H_t 统计独立, 则

$$\begin{aligned} E\{J[\hat{\theta}_{LS}(t)]\} &= E\{\varepsilon_t^T \varepsilon_t\} = E\{\text{tr}[Q V_t^T V_t]\} = \\ \text{tr}\{E[Q V_t^T V_t]\} &= \sigma^2 \text{tr}[E(Q)] = \\ \sigma^2 E\{\text{tr}[I_t - H_t(H_t^T H_t)^{-1} H_t^T]\} &= \\ \sigma^2 E\{\text{tr}[I_t] - \text{tr}[H_t(H_t^T H_t)^{-1} H_t^T]\} &= \\ \sigma^2(t - \dim \theta). \end{aligned}$$

于是

$$\sigma^2 = \frac{E[\varepsilon_t^T \varepsilon_t]}{t - \dim \theta} = \frac{E\{J[\hat{\theta}_{LS}(t)]\}}{t - \dim \theta}.$$

上式表明: 可按式

$$\hat{\sigma}^2 = \frac{J[\hat{\theta}_{LS}(t)]}{t - \dim \theta}$$

计算噪声方差的估计值, 这个噪声方差估计是 σ^2 的无偏估计量. 该定理提供了一种计算噪声方差的估计值的方法.

3 激励信号与激励条件

就系统参数辨识而言, 期望利用系统的输入输出数据能够确定系统参数, 也就是期望输入信号能够激发出系统的所有特征或所有模态, 使得输出数据包括系统模型的全部信息. 这样的信号是充分丰富的(sufficiently rich) 或称为持续激励信号(persistent excitation signal). 在系统辨识算法的收敛性分析中, 直接使用输入信号持续激励假设并不方便, 而是使用持续激励条件, 它是指由输入输出数据构成的信息向量满足的不等式. 持续激励条件有强持续激励条件、弱持续激励条件和衰减激励条件等^[7].

定理 3 说明, 在弱持续激励条件下, 参数估计误差收敛于零. 参数估计性质依赖于持续激励条件, 本节介绍激励信号与激励条件.

3.1 持续激励信号

系统的可辨识性依赖于系统的输入信号, 辨识

算法的收敛性依赖于持续激励条件, 下面介绍持续激励信号与持续激励条件.

定义 1 (持续激励信号定义) 对于信号 $u_0(t) \in \mathbf{R}$, 定义 n 维信号列向量:

$$\varphi(t) := [u_0(t) \ u_0(t-1) \ \cdots \ u_0(t-n+1)]^T \in \mathbf{R}^n,$$

若存在正常数 α 和整数 $N \geq n$ 使得下式成立:

$$(A1) \quad \frac{1}{N} \sum_{i=1}^N \varphi(t-i+1) \varphi^T(t-i+1) \geq \alpha I, \\ t \geq N+n-2,$$

则称 $u_0(t)$ 为 n 阶持续激励信号(persistent excitation signal). 式中 I 为适当维数单位阵. 条件(A1)表示对每一个不小于 $N+n-2$ 的 t 都成立 $t \geq N+n-2$ 约束上式求和的信息向量 $\varphi(t-i+1)$ 中的所有输入项时都大于零, 因为通常假设 t 小于零时, 所有输入输出变量都设为零. 因此, 条件(A1)称为强持续激励条件(Strong Persistent Excitation condition, SPE). 条件(A1)可以等价写为

$$(A1) \quad \frac{1}{N} \sum_{i=0}^{N-1} \varphi(t-i) \varphi^T(t-i) \geq \alpha I \quad t \geq N+n-2$$

或

$$(A1'') \quad \frac{1}{N} \sum_{i=0}^{N-1} \varphi(t+i) \varphi^T(t+i) \geq \alpha I \quad t \geq n-1.$$

或

$$(A1''') \quad \frac{1}{N} \sum_{j=t}^{t+N-1} \varphi(j) \varphi^T(j) \geq \alpha I \quad t \geq n-1.$$

注: 存在常数 α , 表示在不同的地方, 其值可能是不一样的. 一般来说, 系统有 n 个参数, 就要求至少 n 阶激励信号.

1) 有限脉冲响应模型

持续激励信号直接与系统参数可辨识性相关. 下面通过一个简单的有限脉冲响应模型(Finite Impulse Response model, FIR)加以说明. 设输入为 $u(t)$ 的 n 个参数 $(b_0 \ b_1 \ \cdots \ b_{n-1})$ 的 FIR 模型为

$$y(t) = b_0 u(t) + b_1 u(t-1) + \cdots + b_{n-1} u(t-n+1). \quad (21)$$

其中 $y(t)$ 为系统输出. 定义信息向量 $\varphi(t)$ 和参数向量 θ 如下:

$$\varphi(t) := \begin{bmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-n+1) \end{bmatrix} \in \mathbf{R}^n; \quad \theta := \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-1} \end{bmatrix} \in \mathbf{R}^n. \quad (22)$$

式(21)可以写成最小二乘格式,

$$y(t) = \varphi^T(t) \theta. \quad (23)$$

考虑 t 到 $t+p-1$ 共 p ($p \geq n$) 组数据, 定义堆积输出向量 $Y(t)$ 和堆积信息矩阵 $H(t)$ 如下:

$$Y(t) := \begin{bmatrix} y(t) \\ y(t+1) \\ \vdots \\ y(t+p-1) \end{bmatrix} \in \mathbf{R}^p;$$

$$H(t) := \begin{bmatrix} \varphi^T(t) \\ \varphi^T(t+1) \\ \vdots \\ \varphi^T(t+p-1) \end{bmatrix} \in \mathbf{R}^{p \times n}. \quad (24)$$

由式(23)可得

$$Y(t) = H(t) \theta.$$

上式两边左乘 $H^T(t)$ 可得

$$H^T(t) H(t) \theta = H^T(t) Y(t). \quad (25)$$

如果数据乘积矩阵 (data product moment matrix)

$$S(t) := H^T(t) H(t) = \sum_{i=0}^{p-1} \varphi(t+i) \varphi^T(t+i)$$

是非奇异的, 那么从式(25)可以得到 θ 的最小二乘估计:

$$\hat{\theta} = [H^T(t) H(t)]^{-1} H^T(t) Y(t) = S^{-1}(t) H^T(t) Y(t). \quad (26)$$

$S(t)$ 是对称非负定阵, 它非奇异的条件是 $S(t) > 0$. 这就要求 $u(t)$ 为满足条件(A1)的至少 n 阶持续激励信号. 数据长度 p 必须大于或等于 n , 因为矩阵乘积的秩等于 2 个乘积矩阵列秩和行秩中最小的一个.

从这个 FIR 模型的最小二乘估计式(26)可以看出: 估计 $\hat{\theta}$ 有解, 要求式(26)的非负定矩阵 $S(t)$ 可逆, 即下列持续激励条件成立:

$$S(t) = \sum_{i=0}^{p-1} \varphi(t+i) \varphi^T(t+i) \geq \alpha I. \quad (27)$$

如果输入 $u(t)$ 采用至少 n 阶持续激励信号 (因为参数向量维数为 n), 那么式(27)与条件(A1')是等价的. 可以说持续激励信号是从这个简单的 FIR 模型参数估计有解导出的.

细心的读者可能会看出, 对于这个确定性 FIR 模型(21), 似乎并不要求持续激励条件(27)对每一个 t 都成立, 只需对某一个 t 成立即可. 这对于确定性系统确实如此 (对于测量变量含有误差 (简称噪声) 的随机系统, 情况不一样). 对于 n 个参数的 FIR 模型(21), 只需要 n 个方程就可以唯一确定参数向量 $\theta = [b_0 \ b_1 \ \dots \ b_{n-1}]^T \in \mathbf{R}^n$, 只需在式(21)中任取至少 n 个不同的 $t = t_s, s \in \{0, 1, 2, \dots\}$ 就可得到这些方程. 这时持续激励条件可修改为

$$\sum_{s=0}^{p-1} \varphi(t_s) \varphi^T(t_s) \geq \alpha I. \quad (28)$$

这时的参数估计为

$$\hat{\theta} = [H^T(t_s) H(t_s)]^{-1} H^T(t_s) Y(t_s) = S^{-1}(t_s) H^T(t_s) Y(t_s).$$

注: 由于 t 是一个时标变量, 是动态变化的, 式(24)修改为从 $t-p+1$ 到 t 的 p 组数据构成的堆积输出向量和堆积信息矩阵:

$$Y(t) := \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix} \in \mathbf{R}^p;$$

$$H(t) := \begin{bmatrix} \varphi^T(t) \\ \varphi^T(t-1) \\ \vdots \\ \varphi^T(t-p+1) \end{bmatrix} \in \mathbf{R}^{p \times n},$$

上述最小二乘估计结论(26)也是不变的.

2) 受控自回归模型

对于有限脉冲响应模型, 持续激励条件的信息向量 $\varphi(t)$ 只包含了系统输入信号 $u(t-i)$ (参见式(21)中 $\varphi(t)$ 的定义). 然而, 对于受控自回归模型 (Controlled AutoRegressive model, CAR), 情况就不一样了, 因为信息向量不仅包含系统的输入信号 $u(t-i)$, 而且包含系统的输出信号 $y(t-i)$.

考虑下列 CAR 模型,

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

其中: $y(t) \in \mathbf{R}$ 是系统输出, $u(t) \in \mathbf{R}$ 是系统输入, $v(t) \in \mathbf{R}$ 为零均值白噪声, $A(z)$ 和 $B(z)$ 是单位后移算子 z^{-1} 的多项式 ($z^{-1} y(t) = y(t-1)$):

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

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

这个系统有 $n := n_a + n_b$ 个参数 (a_i, b_i), 其差分方程形式为

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

定义参数向量 θ 和信息向量 $\varphi(t)$ 如下:

$$\theta := [a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b}]^T \in \mathbf{R}^n,$$

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

则 CAR 模型可以写为

$$y(t) = \varphi^T(t) \theta + v(t). \quad (29)$$

对于这个 CAR 系统的辨识模型(29), 系统输出 $y(t)$ 可以写为参数向量 $\theta \in \mathbf{R}^n$ 的线性函数, 这个系

统的可辨识性仍要求输入信号 $u(t)$ 是持续激励信号, 这个 CAR 模型的持续激励条件定义式仍如 (A1'') 或式 (28) 的形式.

不管是线性系统还是非线性系统, 只要能写成式 (29) 的形式, 那么持续激励条件是一样的. 因为持续激励条件是关于信息向量 $\varphi(t)$ 的, 而不管 $\varphi(t)$ 的内部构成如何. 例如, 对于下列非线性系统

$$y(t) = a_1 y(t-1) + a_2 y(t-2) y(t-3) + b_1 u(t) + b_2 u^2(t-1) + c \cos(\sqrt{t}) + v(t), \quad (30)$$

定义信息向量和参数向量如下:

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

那么这个非线性系统的辨识模型为

$$y(t) = \varphi^T(t) \theta + v(t). \quad (31)$$

故这是一个线性参数的非线性系统(即线性参数系统). 这个非线性系统辨识模型与式 (29) 的线性 CAR 系统辨识模型是一样的. 因此, 不管是线性系统还是非线性系统, 也不管是否能写成式 (29) 的形式, 对于可辨识性而言, 一般都要求输入信号是持续激励的.

3.2 弱持续激励信号

对于所有的量测数据, 持续激励信号要求长度为 N , 从任意起点 t 到终点 $t+N-1$ 数据窗内的数据都要满足不等式 (A1'). 显然, 对每一个 t 要求不等式 (A1'') 成立是很严格的, 这也是 (A1'') 称为强持续激励条件的原因. 由此引出弱持续激励信号和弱持续激励条件.

定义 2(弱持续激励信号的定义) 对于信号 $u_0(t) \in \mathbf{R}$, 定义 n 维信号列向量:

$$\varphi(t) := [u_0(t) \ u_0(t-1) \ \dots \ u_0(t-n+1)]^T \in \mathbf{R}^n,$$

若存在正常数 α 和充分大 t 使得下式成立:

$$(A2) \quad \frac{1}{t} \sum_{j=1}^t \varphi(j) \varphi^T(j) \geq \alpha I,$$

或等价写为

$$(A2') \quad \frac{1}{L} \sum_{t=1}^L \varphi(t) \varphi^T(t) \geq \alpha I,$$

则称 $u_0(t)$ 为 n 阶弱持续激励信号(weak persistent excitation signal), 其中 L 为观测数据长度, $L \gg n$. 条件 (A2) 或 (A2') 称为弱持续激励条件.

由强持续激励条件可以推导出弱持续激励条件成立(读者可证明之). 弱持续激励条件就可保证 FIR 模型和 CAR 模型的最小二乘参数估计的收

敛性.

3.3 衰减激励信号

辨识算法的收敛性往往是考虑时间 $t \rightarrow \infty$ 时参数估计的行为, 然而在实际中, 不容许长时间加入一持续激励信号对系统进行扰动试验, 而衰减激励信号对系统的影响只是暂时的(短时间的), 实际中更有意义. 因此, 定义衰减激励信号以及研究衰减指数应满足的条件, 保证辨识算法参数估计的收敛性, 对于提高辨识算法的应用效果具有重要意义^[7, 41].

定义 3(衰减激励信号定义 1) 设

$$(U1) \quad u(t) = \frac{u_0(t)}{t^\varepsilon} + u_1(t) \quad \varepsilon > 0,$$

$u_0(t)$ 为满足 (A1) 的持续激励信号, $u_1(t)$ 为非持续激励信号(最特殊的情形是 $u_1(t) \equiv 0$), 则 $u(t)$ 是一个衰减激励信号. 这里 $\varepsilon > 0$ 称为衰减指数(attenuating index). 当 $u_1(t) \equiv 0$ 时, 信号 $u(t)$ 随着 t 增大而趋于零, 故称为衰减激励信号.

由式 (19) 可知, 参数估计误差收敛于零, 衰减激励信号 (U1) 构成的信息向量, 其数据乘积矩阵 $(H_t^T H_t)$ 的最小特征值必须趋于无穷大, 所以衰减指数必须满足 $\varepsilon \leq \frac{1}{2}$.

衰减激励信号的定义有多种形式, 下面给出对数形式的衰减激励信号.

定义 4(衰减激励信号定义 2) 设

$$(U2) \quad u(t) = \frac{u_0(t)}{\ln t} + u_1(t) \quad t \geq t_0,$$

$u_0(t)$ 为满足 (A1) 的持续激励信号, $u_1(t)$ 为非持续激励信号(最特殊的情形是 $u_1(t) \equiv 0$), 则 $u(t)$ 也是一个衰减激励信号. 这个定义没有引入衰减指数, 因为这个衰减激励信号比 (U1) 衰减得更慢, 可以保证 $(H_t^T H_t)$ 的最小特征值仍然趋于无穷大. 当然引入衰减指数 $\varepsilon > 0$ 也是可以的, 只需将 (U2) 的分母修改为 $(\ln t)^\varepsilon$, 其中 $\varepsilon > 0$.

定义 5(衰减激励信号定义 3) 衰减更慢的重对数衰减激励信号定义为

$$(U3) \quad u(t) = \frac{u_0(t)}{\ln \ln t} + u_1(t) \quad t \geq t_0.$$

$$(U4) \quad u(t) = \frac{u_0(t)}{\ln \ln \ln t} + u_1(t) \quad t \geq t_0.$$

其他形式的衰减激励信号如下:

$$(U5) \quad u(t) = \frac{u_0(t)}{t^\varepsilon (\ln t)^c + 1} + u_1(t),$$

$$t \geq t_0, \rho > 0, \rho \leq \varepsilon < \frac{1}{2}. \quad (32)$$

$$(U6) \quad u(t) = \frac{u_0(t)}{t^{\varepsilon} (\ln t)^c + \cos \pi t + 2} + u_1(t),$$

$$t \geq t_0, \rho > 0, \rho \leq \varepsilon < \frac{1}{2}. \quad (33)$$

不同形式的衰减激励信号对应的衰减激励条件也有多种形式. 衰减激励信号 (U2) — (U4) 对应的衰减激励条件的表达式较为简单, 在工程上也最容易实现.

下面我们不加证明地给出 2 个定理.

定理 5 设 $\{u(t)\}$ 为 n 阶持续激励信号 (persistent excitation signal), 则由模型

$$y(t) = (b_1 z^{-1} + b_2 z^{-2} + \dots + b_p z^{-p}) u(t)$$

产生的输出 $\{y(t)\}$ 也为 n 阶持续激励信号, 其中 b_i 不全为零.

定理 6 设 $\{u(t)\}$ 为 n 阶持续激励信号, 则由模型

$$\begin{cases} A(z) y(t) = B(z) u(t), \\ A(z) : = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a}, \\ B(z) : = b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b} \end{cases}$$

产生的输出 $\{y(t)\}$ 为 n 阶持续激励信号的充要条件是 $A(z)$ 与 $B(z)$ 无公因子.

这 2 个定理目前还没有简洁的证明, 如果读者有兴趣深入研究, 研究成果可以公开发表.

3.4 持续激励信号的产生

产生持续激励信号的最简单方法是采用零均值可测白噪声序列, 或伪随机二进制 (PRBS) 序列. 在 Matlab 中, 可用 rand, randn, jinput 等函数产生持续激励输入信号或仿真用的干扰随机噪声. 白噪声是无穷阶持续激励信号, 它可以用作激励任何系统的输入信号. 周期为 n 的 PRBS 码或 M 序列也是一种很好的 n 阶持续激励信号, 它们也可以作为激励系统的输入信号^[39].

3.4.1 白噪声激励信号

白噪声序列是一种最简单的随机序列. 如果随机序列 $\{v(t)\}$ 均值为零, 方差为 σ^2 , 且两两不相关, 则对应的相关函数为

$$R_v(k) := E[v(t)v(t+k)] = \begin{cases} \sigma^2, & k=0, \\ 0, & k \neq 0. \end{cases}$$

则这种随机序列 $\{v(t)\}$ 称为白噪声序列 (white noise sequence). 当 $k \neq 0$ 时, 白噪声的相关函数 $R_v(k)$ 为零, 即

$$R_v(0) = \sigma^2, R_v(1) = R_v(2) = \dots = R_v(\infty) = 0.$$

如果不是白噪声, 也不是周期随机序列, 即相关噪声 (有色噪声) 就有

$$R_v(0) > R_v(1) > R_v(2) > R_v(3) > \dots$$

对每个固定的 t , 白噪声 (white noise) $v(t)$ 是零均值, 方差为常数的独立随机变量. 白噪声序列是平稳各态遍历的.

白噪声序列 $\{v(t)\}$ 的谱密度函数为

$$f_v(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} R_v(k) e^{-jk\omega} = \frac{\sigma^2}{2\pi}, -\pi \leq \omega \leq \pi.$$

即白噪声的谱密度函数 (或功率谱密度) 为常数 $\frac{\sigma^2}{2\pi}$,

所以白噪声序列的功率在 $-\pi$ 到 $+\pi$ 的全频段内均匀分布. 基于这个特点, 人们借用光学中“白色光”一词, 称这种噪声为“白”噪声. 与白噪声对应的是有色噪声 (colored noise) 或相关噪声 (correlated noise).

同样, 对白噪声向量序列 (white noise vector) $\{v(t)\}$, 有

$$\begin{cases} E[v(t)] = \mathbf{0}, \\ R_v(k) = E[v(t)v^T(t+k)] = \begin{cases} \mathbf{R}, & k=0, \\ \mathbf{0}, & k \neq 0, \end{cases} \end{cases}$$

其中 \mathbf{R} 为正定矩阵 (positive definite matrix).

3.4.2 白噪声的产生方法

在辨识仿真研究中, 经常要用到各种不同分布的白噪声序列. 为方便起见, 把各种不同分布的白噪声序列统称为随机数 (random number), 这里只讨论 (0, 1) 均匀分布随机数 (记为 $M(0, 1)$) 和正态分布随机数的产生方法及其性质.

产生 (0, 1) 均匀分布随机数的一种方法是, 把已有的 $M(0, 1)$ 均匀分布的随机数 (如 Rand 的百万随机数) 存放在计算机内存中, 使用时访问内存逐个读取随机数. 例如在 FORTRAN 语言中, 直接调用函数 RAN(1) 就可产生 $M(0, 1)$ 均匀分布的随机数; 在 Turbo C 语言中, 函数 random(x) 产生均匀分布不超过 x 的整数, 则 $\text{random}(x)/x$ 为 $M(0, 1)$ 均匀分布的随机数; 在 Matlab 中, rand 和 randn 产生均匀分布和正态分布的随机数. 下面的定理阐述了正态分布随机数的产生方法.

定理 7^[39] 设 ξ_1 和 ξ_2 是 2 个互为独立的 $M(0, 1)$ 均匀分布 (uniform distribution) 的随机变量, 则

$$\begin{cases} \eta_1 = \sqrt{-2 \ln \xi_1} \cos(2\pi \xi_2), \\ \eta_2 = \sqrt{-2 \ln \xi_1} \sin(2\pi \xi_2) \end{cases}$$

是相互独立、服从 $N(0, 1)$ 正态分布的随机变量。

证明 根据随机变量概率密度的变换关系, 有

$$p(\eta_1, \eta_2) = |\det[\mathbf{J}]|^{-1} p(\xi_1, \xi_2),$$

其中 $p(*, *)$ 为概率密度函数, $|a|$ 为 a 的绝对值, \mathbf{J} 为 Jacobi 矩阵, 即

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \eta_1}{\partial \xi_1} & \frac{\partial \eta_1}{\partial \xi_2} \\ \frac{\partial \eta_2}{\partial \xi_1} & \frac{\partial \eta_2}{\partial \xi_2} \end{bmatrix} =$$

$$\begin{bmatrix} -\frac{1}{\xi_1} \frac{1}{\sqrt{-2\ln \xi_1}} \cos(2\pi\xi_2) & -2\pi \sqrt{-2\ln \xi_1} \sin(2\pi\xi_2) \\ -\frac{1}{\xi_1} \frac{1}{\sqrt{-2\ln \xi_1}} \sin(2\pi\xi_2) & 2\pi \sqrt{-2\ln \xi_1} \cos(2\pi\xi_2) \end{bmatrix}.$$

则

$$|\det[\mathbf{J}]| = \frac{2\pi}{\xi_1}.$$

注意到

$$\eta_1^2 + \eta_2^2 =$$

$$\left[\sqrt{-2\ln \xi_1} \cos(2\pi\xi_2) \right]^2 + \left[\sqrt{-2\ln \xi_1} \sin(2\pi\xi_2) \right]^2 = -2\ln \xi_1.$$

解得

$$\xi_1 = \exp\left(-\frac{\eta_1^2 + \eta_2^2}{2}\right),$$

及

$$p(\xi_1, \xi_2) = p(\xi_1) p(\xi_2) = 1.$$

那么

$$\begin{aligned} p(\eta_1, \eta_2) &= |\det[\mathbf{J}]|^{-1} p(\xi_1, \xi_2) = \\ &= \frac{\xi_1}{2\pi} = \frac{1}{2\pi} \exp\left(-\frac{\eta_1^2 + \eta_2^2}{2}\right) = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\eta_1^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\eta_2^2}{2}\right) = \\ &= p(\eta_1) p(\eta_2). \end{aligned}$$

可见 η_1 和 η_2 是互为独立、服从 $N(0, 1)$ 正态分布的随机变量。

Matlab 程序产生正态分布随机数. 下面的 Matlab 函数 f_rm.m 产生 $N(0, 1)$ 正态分布的长度为 length 的 2 个随机数序列. 用法: eta = f_rm(300) 产生为长度为 300 的 2 列正态分布的随机数.

```
1 function eta = f_rm(length)
2 % Generate the norm distributed random number N(0,1)
3 % rand('state', 1); randn('state', 1);
4 % Filename: f_rm.m
5 % Usage: eta = f_rm(300)
```

```
6 % produces two random variable sequences with
   length 300
7 % -----
8 for i = 1: length
9     r1 = rand(1);
10    while r1 == 0
11        r1 = rand(1);
12    end
13    r2 = rand(1);
14    if i == 1
15        eta = [sqrt(-2*log(r1))*cos(2*pi*r2),
16              sqrt(-2*log(r1))*sin(2*pi*r2)];
17    else
18        eta = [eta; sqrt(-2*log(r1))*cos(2*pi*r2),
19              sqrt(-2*log(r1))*sin(2*pi*r2)];
20    end
21 return
```

假设 $\{u(t)\}$ 是一个零均值, 方差为 σ^2 的白噪声序列, 则其均值和相关函数 $R_u(k)$ 有如下关系,

$$E[u(t)] = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{t=1}^L u(t) = 0,$$

$$R_u(k) = E[u(t)u(t+k)] =$$

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{t=1}^L u(t)u(t+k) = R_u(-k).$$

假设 $u(t) \in \mathbf{R}$ 为输入信号, 定义 n 维输入信号向量,

$\boldsymbol{\varphi}(t) = [u(t) \ u(t-1) \ \cdots \ u(t-n+1)]^T \in \mathbf{R}^n$, 将其代入条件(A1'')的左边得到

$$\begin{aligned} & \frac{1}{N} \sum_{j=t}^{t+N-1} \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j) = \\ & \frac{1}{N} \sum_{j=t}^{t+N-1} \begin{bmatrix} u(j) \\ u(j-1) \\ \vdots \\ u(j-n+1) \end{bmatrix} [u(j) \ u(j-1) \ \cdots \ u(j-n+1)] = \\ & \frac{1}{N} \sum_{j=t}^{t+N-1} \begin{bmatrix} u^2(j) & \cdots & u(j)u(j-n+1) \\ u(j-1)u(j) & \cdots & u(j-1)u(j-n+1) \\ u(j-2)u(j) & \cdots & u(j-2)u(j-n+1) \\ \vdots & & \vdots \\ u(j-n+1)u(j) & \cdots & u^2(j-n+1) \end{bmatrix}. \end{aligned}$$

上式取极限 $N \rightarrow \infty$ 利用相关函数的定义可得

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=t}^{t+N-1} \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j) =$$

$$\begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n-2) \\ \vdots & \vdots & & \vdots \\ R_u(n-2) & R_u(n-3) & \cdots & R_u(1) \\ R_u(n-1) & R_u(n-2) & \cdots & R_u(0) \end{bmatrix}$$

根据白噪声的性质:

$$R_u(0) = \sigma^2,$$

$$R_u(1) = R_u(2) = \cdots = R_u(n) = \cdots = R_u(\infty) = 0,$$

则有

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=t}^{t+N-1} \varphi(j) \varphi^T(j) = \sigma^2 \mathbf{I}_n.$$

因此,白噪声是一个 n 阶持续激励信号,也是一个无穷阶持续激励信号. 因为白噪声的谱密度函数在任何频率都大于零,故可以用作激励任何系统的输入信号. 如果有某些频率使得 $f_u(\omega) = 0$, 则 $u(t)$ 是有限阶持续激励信号.

3.5 基本激励条件

辨识算法的收敛性一般要求输入为持续激励信号,或要求由输入输出数据构成的回归信息向量是充分丰富的(持续激励的),等价于一些不同形式的激励条件成立.

辨识算法的收敛性研究都是分析不同激励条件下参数估计的性质. 创新性的成果是研究现存辨识方法在更弱激励条件下的性能,或者研究新提出辨识方法在不同激励条件下的收敛性能,建立相关收敛性定理(参数估计收敛速度、参数估计误差上界等). 下面讨论本文作者总结和提出的一些基本激励条件. 广义强持续激励条件、广义弱持续激励条件、衰减持续激励条件都是本文作者提出的较弱激励条件.

3.5.1 强持续激励条件

激励条件是关于输入输出数据构成的回归信息向量 $\varphi(t) \in \mathbf{R}^n$ 的一些不等式. 在不同的激励条件中,存在正常数 α, β, γ 和整数 $N \geq \dim \varphi(t) = n$ 一般是不同的. 下面给出几个不同形式的强持续激励条件(Strong Persistent Excitation condition, SPE), 一个比一个条件弱. 如果不作特别说明,下面持续激励条件中的 t 取值表示 $t = 1, 2, 3, \dots, s$ 取值表示 $s = 1, 2, 3, \dots$

1) 强持续激励条件 1

$$(SPE1) \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \varphi(t+i) \varphi^T(t+i) \leq \beta \mathbf{I} \text{ a. s.}$$

它有下列等价形式:

$$(SPE1') \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{j=t}^{t+N-1} \varphi(j) \varphi^T(j) \leq \beta \mathbf{I} \text{ a. s.}$$

研究表明: 保证辨识算法参数估计收敛性只需要强持续激励条件的下界 $\alpha \mathbf{I}$, 下界约束可以推导出一些参数估计算法(如递推最小二乘算法)的参数估计误差上界(对于有限数据长度); 辨识算法收敛性不需要强持续激励条件的上界 $\beta \mathbf{I}$, 上界约束可以推导出一些参数估计算法(如递推最小二乘算法)的参数估计误差下界.

(SPE1) 要求对每一个 $t (t = 1, 2, 3, \dots)$ 都成立, 即对于窗口长度为 N 的滑动数据窗的信息向量 $\varphi(t)$ 都成立. 这可以解释为对于长度为 $L = Nk$ 的数据量 (SPE1) 包含了 $Nk - N + 1$ 个不等式.

规范化强持续激励条件(Normalized Strong Persistent Excitation condition, NSPE) 如下.

$$(NSPE) \quad \frac{1}{N} \sum_{i=0}^{N-1} \frac{\varphi(t+i) \varphi^T(t+i)}{\|\varphi(t+i)\|^2} \geq \alpha \mathbf{I} \text{ a. s.}$$

或

$$(NSPE) \quad \frac{1}{N} \sum_{i=0}^{N-1} \frac{\varphi(t+i) \varphi^T(t+i)}{1 + \|\varphi(t+i)\|^2} \geq \alpha \mathbf{I} \text{ a. s.}$$

NSPE 不需要对上界作限制, 因为其上是单位阵. 上式分母中加上的 1 是为了避免分母为零.

注: 其他激励条件也可进行规范化, 本文不一一讨论.

2) 强持续激励条件 2

$$(SPE2) \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \varphi(Nt+i) \varphi^T(Nt+i) \leq \beta \mathbf{I} \text{ a. s.}$$

或等价表示为

$$(SPE2') \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{j=Nt}^{Nt+N-1} \varphi(j) \varphi^T(j) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(SPE2'') \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \varphi(t+i) \varphi^T(t+i) \leq \beta \mathbf{I},$$

a. s. $t = Nk, k = 1, 2, \dots$

(SPE1) 要求对每一个 t 都成立, 弱化了 (SPE2) 只要求对稀少规则的 t 成立(稀少周期的 t 成立). 换句话说, (SPE2) 只要求对每一个 $t = Nk (k = 0, 1, 2, \dots)$ 都成立, 即对于窗口长度为 N 的连续数据窗的信息向量 $\varphi(t)$ 都成立. 这可以解释为在长度为 $L = Nk$ 的数据窗内, (SPE2) 包含了 k 个不等式.

下面是动态数据窗长度的强持续激励条件, 数据窗长度由 N 变为 t_s^* .

3) 强持续激励条件 3

定义一整数序列 $\{t_s; s = 0, 1, 2, \dots\}$ 满足

$$0 = t_0 < t_1 < t_2 < t_3 < \dots < t_{s-1} < t_s < \dots,$$

且 $t_s^* := t_s - t_{s-1} \geq n$. 强持续激励条件 3 如下^[42].

$$(SPE3) \quad \alpha \mathbf{I} \leq \frac{1}{t_{s+1}^*} \sum_{i=0}^{t_{s+1}^*-1} \boldsymbol{\varphi}(t_s + i) \boldsymbol{\varphi}^T(t_s + i) \leq \beta \mathbf{I} \text{ a. s.}$$

或等价表示为

$$(SPE3) \quad \alpha \mathbf{I} \leq \frac{1}{t_{s+1}^* - t_s} \sum_{j=t_s}^{t_{s+1}^*-1} \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j) \leq \beta \mathbf{I} \text{ a. s.}$$

弱化了了的(SPE3)的数据窗长度变为动态的 t_s^* , 它只要求对稀少不规则的 $t = t_s (s = 1, 2, 3, \dots)$ 成立.

强持续激励条件的上界是有界的, 如果把上界进一步放大, 就得到新的持续激励条件, 我们称为广义强持续激励条件(Generalized Strong Persistent Excitation condition, GSPE). 由上述 3 个强持续激励条件, 可以得到 3 个对应的广义强持续激励条件.

4) 广义强持续激励条件 1

$$(GSPE1) \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \leq \beta t^\gamma \mathbf{I} \text{ a. s.}$$

5) 广义强持续激励条件 2

$$(GSPE2) \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(Nt+i) \boldsymbol{\varphi}^T(Nt+i) \leq \beta t^\gamma \mathbf{I} \text{ a. s.}$$

或

$$(GSPE2') \quad \alpha \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(Nt+i) \boldsymbol{\varphi}^T(Nt+i) \leq \beta (Nt)^\gamma \mathbf{I} \text{ a. s.}$$

6) 广义强持续激励条件 3

$$(GSPE3) \quad \alpha \mathbf{I} \leq \frac{1}{t_{s+1}^*} \sum_{i=0}^{t_{s+1}^*-1} \boldsymbol{\varphi}(t_s + i) \boldsymbol{\varphi}^T(t_s + i) \leq \beta t_s^\gamma \mathbf{I} \text{ a. s.}$$

3.5.2 弱持续激励条件

在所考虑量测数据集中, 强持续激励条件中信息向量要满足许多不等式, 而弱持续激励条件(Weak Persistent Excitation condition, WPE)只有一个不等式.

1) 弱持续激励条件 1

$$(WPE1) \quad \alpha \mathbf{I} \leq \frac{1}{t} \sum_{j=1}^t \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j) \leq \beta \mathbf{I} \text{ a. s. 对充分大 } t.$$

充分大 t .

这个条件使用了所有观测信息向量 $\boldsymbol{\varphi}(j) (j = 1, 2, \dots, t)$. 如果观测信息向量 $\boldsymbol{\varphi}(t)$ 是各态遍历的, 则有

$$(WPE1') \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \boldsymbol{\varphi}(i) \boldsymbol{\varphi}^T(i) = \mathbf{R} > 0 \text{ a. s.}$$

2) 弱持续激励条件 2

$$(WPE2) \quad \alpha \mathbf{I} \leq \frac{1}{t} \sum_{j=1}^t \boldsymbol{\varphi}(jN) \boldsymbol{\varphi}^T(jN) \leq \beta \mathbf{I} \text{ a. s. ,}$$

对充分大 t .

这个条件只使用了采样周期为 N 的信息向量 $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}(jN) (j = 1, 2, \dots, t)$.

3) 弱持续激励条件 3

$$(WPE3) \quad \alpha \mathbf{I} \leq \frac{1}{s} \sum_{j=1}^s \boldsymbol{\varphi}(t_j) \boldsymbol{\varphi}^T(t_j) \leq \beta \mathbf{I} \text{ a. s. ,}$$

对充分大 s .

这个条件使用了非均匀采样间隔 t_j^* 的信息向量 $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}(t_j) (j = 1, 2, \dots, s)$.

相应地, 我们有广义弱持续激励条件.

4) 广义弱持续激励条件 1

本文作者提出的广义弱持续激励条件(Generalized Weak Persistent Excitation condition, GWPE)如下^[15, 43]:

$$(GWPE) \quad \alpha \mathbf{I} \leq \frac{1}{t} \sum_{i=1}^t \boldsymbol{\varphi}(i) \boldsymbol{\varphi}^T(i) \leq \beta t^\gamma \mathbf{I} \text{ a. s. 对充分大 } t.$$

5) 广义弱持续激励条件 2

$$(GWPE2) \quad \alpha \mathbf{I} \leq \frac{1}{t} \sum_{j=1}^t \boldsymbol{\varphi}(jN) \boldsymbol{\varphi}^T(jN) \leq \beta t^\gamma \mathbf{I} \text{ a. s. 对充分大 } t.$$

6) 广义弱持续激励条件 3

$$(GWPE3) \quad \alpha \mathbf{I} \leq \frac{1}{s} \sum_{j=1}^s \boldsymbol{\varphi}(t_j) \boldsymbol{\varphi}^T(t_j) \leq \beta s^\gamma \mathbf{I} \text{ a. s. 对充分大 } s.$$

这里引入条件数的概念. 矩阵 X 的条件数(condition number)定义为 X 的最大奇异值与最小奇异值之比(ratio of the maximal and minimal singular values of X). 因此条件数总是不小于 1. 对于对称正定矩阵 X , 其条件数等于其最大特征值(maximum eigenvalue)与最小特征值(minimum eigenvalue)之比.

从条件数的定义可知: 当 $t \rightarrow \infty$ 时, 强持续激励条件与弱持续激励条件中的矩阵的条件数是有界的; 而广义强持续激励条件与广义弱持续激励条件中的矩阵的条件数是无界的.

3.5.3 衰减激励条件

1) 衰减激励条件 1

这里只讨论衰减激励信号(U1) — (U3) 所对应的衰减激励条件(Attenuating Excitation condition, AE)^[41, 44, 46] 定义为

$$(AE1) \quad \frac{\alpha}{(t+N-1)^{2\epsilon}} \mathbf{I} \leq \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \leq$$

$$\beta \mathbf{I} \text{ a. s.}$$

$$(AE1') \quad \frac{\alpha}{[\ln(t+N-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(AE1'') \quad \frac{\alpha}{[\ln \ln(t+N-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \leq \beta \mathbf{I} \text{ a. s.}$$

当衰减指数 $\varepsilon = 0$ 时, 条件(AE1) 等同(SPE1), 故强持续激励条件是衰减激励条件的一个特例.

2) 衰减激励条件 2

$$(AE2) \quad \frac{\alpha}{(Nt+N-1)^{2\varepsilon}} \mathbf{I} \leq$$

$$\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(Nt+i) \boldsymbol{\varphi}^T(Nt+i) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(AE2') \quad \frac{\alpha}{[\ln(Nt+N-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(Nt+i) \boldsymbol{\varphi}^T(Nt+i) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(AE2'') \quad \frac{\alpha}{[\ln \ln(Nt+N-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(Nt+i) \boldsymbol{\varphi}^T(Nt+i) \leq \beta \mathbf{I} \text{ a. s.}$$

3) 衰减激励条件 3

$$(AE3) \quad \frac{\alpha}{(t_{s+1}-1)^{2\varepsilon}} \mathbf{I} \leq$$

$$\frac{1}{t_{s+1}^*} \sum_{i=0}^{t_{s+1}^*-1} \boldsymbol{\varphi}(t_s+i) \boldsymbol{\varphi}^T(t_s+i) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(AE3') \quad \frac{\alpha}{[\ln(t_{s+1}-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{t_{s+1}^*} \sum_{i=0}^{t_{s+1}^*-1} \boldsymbol{\varphi}(t_s+i) \boldsymbol{\varphi}^T(t_s+i) \leq \beta \mathbf{I} \text{ a. s.}$$

$$(AE3'') \quad \frac{\alpha}{[\ln \ln(t_{s+1}-1)]^2} \mathbf{I} \leq$$

$$\frac{1}{t_{s+1}^*} \sum_{i=0}^{t_{s+1}^*-1} \boldsymbol{\varphi}(t_s+i) \boldsymbol{\varphi}^T(t_s+i) \leq \beta \mathbf{I} \text{ a. s.}$$

目前, 衰减激励条件下, 辨识算法的收敛性研究主要集中在时不变确定性系统或时不变随机系统. 作者在衰减激励条件下, 研究了一些辨识算法的一致收敛性, 例如, 文献[44-45]针对确定性系统和随机系统多新息辨识算法, 文献[41, 46-47]针对随机系统递推最小二乘、递阶最小二乘和最小均方算法.

3.5.4 其他激励条件

上面讨论的激励条件不等式都是假设依概率 1

(wp1) 或几乎必然(a. s.) 成立. 事实上, 激励条件还可以以数学期望的形式或条件数学期望的形式给出, 这里简单加以介绍.

我们以期望形式给出的强持续激励条件 1 (Expected SPE condition (ESPE) 为

$$(ESPE) \quad \alpha \mathbf{I} \leq E \left[\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \right] \leq \beta \mathbf{I}.$$

文献[48]以条件期望给出的强持续激励条件 1 (CESPE: Conditionally Expectation SPE condition) 为

$$(CESPE) \quad \alpha \mathbf{I} \leq E \left[\frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i) \mid \mathcal{F}_{t-1} \right] \leq \beta \mathbf{I} \text{ a. s.}$$

或

$$(NCESPE) \quad E \left[\frac{1}{N} \sum_{i=0}^{N-1} \frac{\boldsymbol{\varphi}(t+i) \boldsymbol{\varphi}^T(t+i)}{1 + \|\boldsymbol{\varphi}(t+i)\|^2} \mid \mathcal{F}_{t-1} \right] \geq \alpha \mathbf{I} \text{ a. s.}$$

其中 \mathcal{F}_t 是由直到 t 时刻的观测数据生成的 σ 代数.

其他激励条件也可以数学期望形式或条件数学期望形式给出, 这里不一一讨论.

4 参数估计性质及分析工具

对于递推辨识算法和迭代辨识算法, 期望随着递推变量 t 或迭代变量 k 的增加, 参数估计不断接近系统的真实参数. 参数估计的这种称为收敛性. 参数估计的收敛性也指参数辨识方法的收敛性, 或参数估计算法的收敛性, 简称估计算法收敛性或辨识算法的收敛性.

4.1 参数估计的统计性质

所谓参数估计的收敛性 (convergence) 是指: 当数据量趋于无穷 (即 $t \rightarrow \infty$) 时参数估计的行为, 或参数估计 $\hat{\boldsymbol{\theta}}(t)$ 是否趋于真参数, 或接近真参数向量 $\boldsymbol{\theta}$ 的程度. 设 $\boldsymbol{\theta}$ 是未知参数向量 (矩阵), $\hat{\boldsymbol{\theta}}(t)$ 是 $\boldsymbol{\theta}$ 在 t 时刻的估计, $\tilde{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}$ 是参数估计误差 (parameter estimation error) 向量 (矩阵). 参数估计接近真参数的程度可用无偏性、渐近无偏性、一致收敛性、有界收敛性等来描述.

1) 无偏性 (unbiasness)

对所有的 t , 如果参数估计 $\{\hat{\boldsymbol{\theta}}(t)\}$ 的期望都等于真参数 $\boldsymbol{\theta}$, 即

$$E[\hat{\boldsymbol{\theta}}(t)] = \boldsymbol{\theta} \text{ 或 } E[\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}] = 0 \quad t=1, 2, 3, \dots$$

就称参数估计是无偏估计 (unbiased estimate), 或称参数估计是无偏的 (unbiased).

2) 渐近无偏性 (asymptotical unbias)

当 $t \rightarrow \infty$ 时, 如果参数估计 $\hat{\boldsymbol{\theta}}(t)$ 的期望都等于真

参数 θ 即

$$\lim_{t \rightarrow \infty} E[\hat{\theta}(t)] = \theta,$$

就说参数估计是渐近无偏的 (asymptotically unbiased).

3) 一致性 (consistency)

对于任意给 $\varepsilon > 0$ 当 $t \rightarrow \infty$ 时 如果 $\|\hat{\theta}(t) - \theta\| < \varepsilon$ 的概率 P 趋于 1 即

$$\lim_{t \rightarrow \infty} P(\|\hat{\theta}(t) - \theta\| < \varepsilon) = 1,$$

就称 $\hat{\theta}(t)$ 是一致估计 (consistent estimate), 或称 $\hat{\theta}(t)$ 依概率 1 (wp1) 或几乎必然 (a. s.) 收敛于真参数 θ 记作

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta, \text{wp1},$$

或

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta \text{ a. s.}$$

上式等价于

$$\begin{cases} \lim_{t \rightarrow \infty} E[\hat{\theta}(t)] = \theta, \\ \lim_{t \rightarrow \infty} \text{var}[\hat{\theta}(t) - \theta] = 0. \end{cases}$$

4) 均方收敛性 (mean square convergence)

当 $t \rightarrow \infty$ 时, 如果参数估计参数估计 $\{\hat{\theta}(t)\}$ 满足

$$\lim_{t \rightarrow \infty} E\{[\hat{\theta}(t) - \theta]^T [\hat{\theta}(t) - \theta]\} = 0,$$

或

$$\lim_{t \rightarrow \infty} E\{\|\hat{\theta}(t) - \theta\|^2\} = 0,$$

就称 $\hat{\theta}(t)$ 均方收敛于真参数 θ , 或称 $\hat{\theta}(t)$ 是均方 (mean square m. s.) 收敛的, 记作

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta \text{ m. s.}$$

概率论中均方收敛不一定依概率 1 收敛; 依概率 1 收敛也不一定均方收敛.

5) 有效性

设 $\{\hat{\theta}(t)\}$ 是无偏随机估计序列, 且有 $E\{[\hat{\theta}(t) - \theta][\hat{\theta}(t) - \theta]^T\}$ 各元素达到最小值, 则 $\hat{\theta}(t)$ 称为有效估计值. 更详细地说, 有效估计值的偏差 $\tilde{\theta}(t) := \hat{\theta}(t) - \theta$ 的协方差阵达到 Cramér-Rao 不等式下界.

这个有效性的定义不是很合理, 严格地说任何估计都不会达到 Cramér-Rao 不等式的下界, 都是无效的. 本文作者认为, 参数估计误差上界小于可以容忍的值, 就认为参数估计是有效的. 因为真参数 θ 是未知的, 是要估计的, 即使得到 θ 的估计 $\hat{\theta}(t)$, 也难以得到参数估计误差 $\tilde{\theta}(t) := \hat{\theta}(t) - \theta$ 的值 (参数估计误差 $\tilde{\theta}(t)$ 只是在仿真中使用的一个评价指标, 实际系统的参数估计误差不可能知道的), 所以通常使用参数估计误差上界来衡量估计精度.

收敛性是参数估计理论的一个重要内容. 当估计值的数学期望等于参数真值时, 参数估计就是无偏估计. 当估计值是数据的线性函数时, 参数估计就是线性估计; 当估计值的均方差最小时, 参数估计为一致最小均方误差估计; 若线性估计又是一致最小均方误差估计, 则称为最优线性无偏估计; 若无偏估计值的方差达到 Cramér-Rao 不等式的下界, 则称为有效估计值; 若以概率 1 收敛于真参数, 就称为一致性估计值. 在一定条件下, 最小二乘估计是最优线性无偏估计, 它的估计值是有效估计, 而且是一致性估计.

4.2 Cramér-Rao 不等式

Cramér-Rao 不等式给出了概率意义下和有限数据长度时, 时不变系统参数估计误差协方差的下界公式. 随着数据长度增加, 这个下界趋于零. 有效估计值要求 $E[\hat{\theta}(t)] = \theta$, 且协方差矩阵 $\text{var}[\hat{\theta}(t) - \theta]$ 各元素为最小. Cramér-Rao 给出了这个下界公式.

定理 8^[49] (Cramér-Rao 不等式) 考虑一个随机向量 y , 它在参数 θ 条件下的条件概率密度函数记作 $p(y|\theta)$. 在一定的正则条件下, 参数 θ 的任何无偏估计值 $\hat{\theta}$ 都将满足下列不等式,

$$\text{cov}[\tilde{\theta}] = E\{[\hat{\theta}(t) - \theta][\hat{\theta}(t) - \theta]^T\} \geq M^{-1},$$

其中 M 为 Fisher 信息矩阵, 定义为

$$M = E_{y|\theta} \left\{ \left[\frac{\partial \ln p(y|\theta)}{\partial \theta} \right] \left[\frac{\partial \ln p(y|\theta)}{\partial \theta} \right]^T \right\},$$

定理的证明参考文献 [49]. 这个定理即为著名的 Cramér-Rao 不等式. 它表明估计值偏差的方差不能小于 Fisher 信息矩阵逆的对角元素.

Cramér-Rao 不等式给出了时不变系统参数估计误差协方差阵下界的估算公式^[50]:

$$E[\tilde{\theta}(t) \tilde{\theta}^T(t)] \geq M^{-1}.$$

不幸的是, 人们无法通过计算得到估计误差协方差阵的下界, 因为计算 Fisher 信息矩阵 (Fisher information matrix) M 需要系统真参数的知识, 所以 Cramér-Rao 不等式只有抽象的理论意义. 分析说明: 对于时不变随机系统的最小二乘算法, 当 $t \rightarrow \infty$ 时, 估计误差 (协方差阵) (下界) 趋于零, 即 $M \rightarrow \infty$, 所以 Cramér-Rao 的估计误差协方差阵下界的实际用处不大. 参数估计误差上界比下界重要, 因为上界可以衡量参数估计精度.

对于时变系统, 如果参数变化规律未知, 任何辨识算法给出的参数估计都不可能收敛于真参数, 即

时变参数估计不存在一致收敛性. 因此, 为了评价时变参数估计精度, 作者引入了有界收敛性这一重要概念. 有界收敛性强调参数估计误差上界的估算.

4.3 实用有界收敛性

当参数估计几乎必然收敛于真参数就说参数估计是一致收敛的. 如果参数估计不能收敛到真参数, 即当 $t \rightarrow \infty$ 时, 参数估计误差 $\hat{\theta}(t) := \hat{\theta}(t) - \theta$ 不为零. 算法增益不趋于零的(时变)参数估计算法的参数估计不可能收敛于真参数. 如果参数估计误差不趋于零, 我们期望这个误差越小越好, 因为参数估计误差是衡量参数估计精度的重要指标. 但是, 由于真参数向量 θ 是未知的, 所以在实际中, 即使获得参数估计 $\hat{\theta}(t)$, 也无法通过计算得到参数估计误差 $\hat{\theta}(t) = \hat{\theta}(t) - \theta$. 为此, 我们引入了参数估计误差上界(parameter estimation error upper bound)的重要概念, 并且期望这个误差上界越小越好. 参数估计误差上界可以间接说明参数估计精度.

研究参数估计算法的性能时, 须区别下列 3 种收敛性^[51-52]:

1) 一致收敛性(consistent convergence)

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta \text{ a. s. (m. s.)}$$

2) 有界收敛性(bounded convergence)

$$E[\|\hat{\theta}(t) - \theta\|^2] \leq f(t) \leq \varepsilon < \infty, \text{ 或 } \|\hat{\theta}(t) - \theta\|^2 \leq f(t) \leq \varepsilon < \infty \text{ a. s. (工程上)}$$

3) 一般收敛性(general convergence)

$$\limsup_{t \rightarrow \infty} E[\|\hat{\theta}(t) - \theta\|^2] < \infty \text{ 或 } \|\hat{\theta}(t) - \theta\|^2 < \infty \text{ a. s. (理论上)}$$

一致收敛性固然重要, 但它只描述了参数估计误差的极限行为(limit behavior). 对于有限数据点(实际中只能得到有限数据), 一致收敛性不能给出有关估计误差 $\hat{\theta}(t)$ 的任何动态信息. 例如, 假设参数估计误差 $\hat{\theta}(t)$ 以 $1/\sqrt{t}$ 速率收敛于零, 即

$$\|\tilde{\theta}(t)\|^2 = O(1/t) \rightarrow 0 \text{ a. s. , 当 } t \rightarrow \infty.$$

这里 $f(t) = O(g(t))$ 表示存在未知常数 C 和充分大 t 使得 $f(t) \leq Cg(t)$. 于是有

$$\|\tilde{\theta}(t)\|^2 \leq \frac{C}{t} \rightarrow 0.$$

虽然知道这个参数估计误差趋近于零, 但对于某个有限的 t 不知道估计误差有多大, 因为 C 是未知的, 也许 C 很大, 如 $C = 20\ 110\ 607\ 012\ 036$, 对于数据长度 $t = 2\ 011$, 这个误差上界还是令人难以接受的. 所以人们更感兴趣知道参数估计误差的上界,

即参数估计的有界收敛性; 或者人们在证明参数估计一致收敛性的同时, 应力求找到误差上界, 即应找出或估计出上式中的 C 值, 这样对于给定的数据长度 $t = L$, 就可以计算出误差上界.

对于有界收敛性, 如果一个算法的误差上界 ε 值较小(如 $\varepsilon = 10^{-3}$), 那么这个算法参数估计精度较高, 算法也有较大应用价值; 反之, ε 值较大(如 $\varepsilon = 10^3$) 就没有太大的应用价值. 有界收敛性是针对实际应用提出的, 具有较大的工程意义. 假如证明了某个算法的参数估计误差上界满足

$$E[\|\tilde{\theta}(t)\|^2] \leq f(t) = \frac{1\ 963.030\ 9}{t^2} + \frac{1\ 983.021\ 4}{t}. \quad (34)$$

这是一个有界收敛性结果, 它也是一个一致收敛性结果, 因为当 $t \rightarrow \infty$ 时 $f(t) \rightarrow 0$. 式(34)提供了一个计算参数估计误差上界的公式, 是十分有用的, 它同时提供了下面这个有趣问题的答案: 对于给定的小数 $\varepsilon > 0$ 和一个辨识算法, 需要多大的数据量(即数据长度 $t = L$ 为多大时), 才能保证参数估计误差(上界)小于 ε ?

如果估计误差收敛于零, 期望找出类似式(34)参数估计误差上界. 研究参数估计的有界性或收敛性, 不仅要找到误差上界, 而且要研究如何减少这个误差上界.

在理论分析中, 如果难以得到有界收敛性的结论, 就只能降格以求, 来分析算法的一般收敛性. 一般收敛性表明: 存在常数 C 使得

$$\|\tilde{\theta}(t)\|^2 \leq C < \infty.$$

通常这个 C 是未知的. 如果能找到这个 C 值, 就变成了有界收敛性问题.

4.4 收敛性分析的基本工具

辨识方法收敛性分析一直是控制领域的重要研究课题. 长期以来, 人们以能证明辨识方法的收敛性和找到辨识算法收敛的最简洁、最弱条件引以为自豪和骄傲. 早期的辨识算法收敛性分析假设系统输入和输出信号具有有限非零功率, 噪声是独立同分布随机序列且 4 阶矩存在^[53], 或过程噪声与系统输入是平稳的和各态遍历的^[54]. 在这些理想条件下, 都没有能够得到参数估计误差收敛的简洁结果.

关于辨识算法收敛性, 1982 年华人统计学家 Lai 等^[55] 在标准递推最小二乘算法的收敛性研究方面取得了重要突破: 假设系统噪声 $\{v(t)\}$ 方差 $E[\|v(t)\|^2] = \sigma^2$ 和高阶矩 $E[\|v(t)\|^\gamma] < \infty$ 存在 ($\gamma > 2$), 获得了参数估计误差收敛速率. 后来, 最小二乘

和随机梯度类算法的收敛性研究都是基于这样所谓的“方差和高阶矩存在”弱假设,如文献[56-60].

作者认为:研究参数估计 2 范数(或 2 阶矩)的收敛性(如均方估计误差),假设过程噪声平稳各态遍历,且高阶矩存在是不必要的.最近,运用鞅收敛定理和鞅超收敛定理等的一系列辨识方法和自适应控制方法的收敛性研究成果表明:参数估计误差或输出跟踪误差 2 范数的收敛性只依赖于噪声 2 阶矩统计特性^[9, 15, 20, 27].这一贡献既保证了参数估计的收敛性,又极大地减弱了对噪声方差和高阶矩存在的假设.文献[9, 61]在强持续激励条件和噪声方差有界条件下,获得了有限数据窗递推最小二乘算法和遗忘梯度算法的参数估计误差界.文献[15]研究了双率系统辅助模型递推最小二乘算法参数估计的收敛性能.文献[27-28]在同一条件下证明了递阶最小二乘和递阶随机梯度算法是一致收敛的,这个证明并没有假设噪声方差和高阶矩存在,也就是说噪声方差可以是时变的,即所考虑的系统可能是非平稳的.

随机过程理论和随机鞅理论是研究递推辨识方法收敛性的主要分析工具.鞅收敛定理(martingale convergence theorem)相当于李雅普诺夫稳定性定理在随机系统中的推广,它主要用来研究时不变参数系统辨识算法参数估计的一致收敛性,适用于一些增益趋于零的算法,例如递推最小二乘算法、随机梯度算法、递推增广最小二乘算法、增广随机梯度算法、多新息随机梯度算法、多新息最小二乘算法、辅助模型最小二乘算法和辅助模型随机梯度算法、递阶随机梯度辨识方法等,不适用于增益向量不趋于零的遗忘因子最小二乘算法、有限数据窗最小二乘算法、投影算法等.由于时变参数估计算法不存在一致收敛性,故鞅收敛定理不适用时变系统参数估计算法.为此,本文作者建立了研究时变系统参数估计方法有界收敛性的鞅超收敛定理(martingale hyperconvergence theorem),为时变系统自适应辨识和控制算法的稳定性和收敛性分析开辟了新思路.

定理 9(Frechet 定理) 设 $\{x(t)\}$ 是依概率 1 收敛于 x_0 的随机向量序列, $f(\cdot)$ 为连续标量函数.则下列概率极限成立,

$$f(x(t)) \rightarrow f(x_0) \text{ w.p.1, 当 } t \rightarrow \infty,$$

或写成

$$\lim_{t \rightarrow \infty} f(x(t)) = f(x_0).$$

定理 10(极限定理) 设矩阵 A_t 和 B_t 存在概率极限,且其维数不随 t 的增加而变化,应用 Frechet 定

理,可得

$$\lim_{t \rightarrow \infty} (A_t B_t) = (\lim_{t \rightarrow \infty} A_t) (\lim_{t \rightarrow \infty} B_t),$$

$$\lim_{t \rightarrow \infty} (A_t^{-1}) = (\lim_{t \rightarrow \infty} A_t)^{-1}.$$

定理 11(鞅收敛定理)^[62] 设 $\{T(t)\}, \{f(t)\}, \{g(t)\}$ 均为非负随机变量序列,它们关于 σ 代数 \mathcal{F}_t 是可测的(或适应的),且使得下式成立,

$$E[T(t+1) | \mathcal{F}_t] \leq T(t) - f(t) + g(t).$$

若 $\sum_{t=1}^{\infty} g(t) < \infty$ a. s., 则 $T(t)$ 几乎确定地收敛于有限的随机变量 T_0 , 即 $T(t) \rightarrow T_0 < \infty$ a. s., 并且 $\sum_{t=1}^{\infty} f(t) < \infty$ a. s.

在辨识算法的收敛性研究中,通常 $T(t)$ 是参数估计误差的非负定函数,通过 $T(t)$ 和 $f(t)$ 的收敛性和持续激励条件等来确定参数估计误差是否收敛于零; $f(t)$ 和 $g(t)$ 一般是参数估计误差和系统输入输出数据的函数; \mathcal{F}_t 是直到 t 时刻系统输入输出数据或是由直到 t 时刻输入输出数据生成的 σ 代数.利用鞅收敛定理研究时不变系统参数辨识算法的收敛性,可参见文献[15, 27-28, 40, 63-65].

定理 12(鞅超收敛定理)^[51, 63, 66] 考虑非负定函数 $T(t) := T[x(t)]$ 和集

$$R_t := [x(t) : g[x(t)] \leq \eta_t < \infty \text{ a. s.}],$$

若对于 $x(t) \in R_t^c$ 有下式成立(R_t^c 是 R_t 的补集),

$$E[T(t) | \mathcal{F}_{t-1}] - T(t-1) =: \Delta T(t) \leq -b(t) \text{ a. s. 对于 } x(t) \in R_t^c, \quad (35)$$

其中 $g(x) = (a^T x)^2$ 称为收敛变量, a 是一个非零的时变或时不变向量, $\eta_t \geq 0$ 是一个非降有界随机变量(即 $R_t \subset R_{t+1}$), $b(t)$ 是一个随机变量, $(x(t), \mathcal{F}_t)$ 是一个适应序列.如果 $x(t) \in R_t^c$ 时, $\sum_{t=t_0}^{\infty} b(t) = \infty$ a. s. ($t_0 < \infty$), 则对于充分大 t , 有 $x(t) \in R_t$ a. s. 成立, $\lim_{t \rightarrow \infty} x(t) \in R_{\infty}$ a. s.

对于时变参数估计算法,首先构造一个参数估计误差 $x(t)$ 的非负定函数 $T(t)$, 然后导出不等式(35), 找出变量 $b(t)$ 满足条件时的集 R_t , 最后利用强持续激励条件和一些基本不等式关系, 推导出参数估计误差 $x(t)$ 的显示上界.由于鞅超收敛定理可以给出参数估计误差的上界,这对评价参数估计精度,提高算法的实际应用效果具有重要意义.

鞅超收敛定理是研究时变随机系统辨识算法和自适应控制算法收敛性的有效工具.鞅超收敛定理的提出和运用鞅超收敛定理对时变参数估计误差有

界性的一系列研究成果,建立了时变系统参数估计误差界分析理论体系^[9, 51, 63, 66].

5 典型辨识算法的收敛定理

为方便起见,设 $\{u(t)\}$ 为系统输入序列, $\{y(t)\}$ 为系统观测输出序列, $\{v(t)\}$ 是零均值随机白噪声序列, z^{-1} 为单位后移算子 $z^{-1}y(t) = y(t-1)$ 或 $zy(t) = y(t+1)$, $A(z)$, $B(z)$, $C(z)$ 和 $D(z)$ 是算子 z^{-1} 的常系数时不变多项式,定义如下:

$$A(z) : = 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a} \in \mathbf{R};$$

$$B(z) : = b_1z^{-1} + b_2z^{-2} + \dots + b_{n_b}z^{-n_b} \in \mathbf{R};$$

$$C(z) : = 1 + c_1z^{-1} + c_2z^{-2} + \dots + c_{n_c}z^{-n_c} \in \mathbf{R};$$

$$D(z) : = 1 + d_1z^{-1} + d_2z^{-2} + \dots + d_{n_d}z^{-n_d} \in \mathbf{R}.$$

多项式系数 a_i , b_i , c_i 和 d_i 为模型参数. 根据移位算子的性质,有

$$A(z)y(t) = (1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a})y(t) = y(t) + a_1y(t-1) + a_2y(t-2) + \dots + a_{n_a}y(t-n_a),$$

$$B(z)u(t) = (b_1z^{-1} + b_2z^{-2} + \dots + b_{n_b}z^{-n_b})u(t) = b_1u(t-1) + b_2u(t-2) + \dots + b_{n_b}u(t-n_b),$$

$$D(z)v(t) = (1 + d_1z^{-1} + d_2z^{-2} + \dots + d_{n_d}z^{-n_d})v(t) = v(t) + d_1v(t-1) + d_2v(t-2) + \dots + d_{n_d}v(t-n_d) \text{ 等.}$$

本文把线性参数系统辨识模型分为 4 类: 1) 线性回归模型; 2) 伪线性回归模型 I; 3) 伪线性回归模型 II; 4) 伪线性回归模型 III. 这 4 类模型都是线性参数模型,它们可以是线性系统,也可以是线性参数的非线性系统.

5.1 线性回归模型

考虑线性回归模型(linear regression model),

$$y(t) = \varphi^T(t)\theta + v(t),$$

其中 $\{y(t)\}$ 是观测输出序列, $\{v(t)\}$ 是零均值不相关随机噪声序列, $\varphi(t) \in \mathbf{R}^n$ 是观测数据构成的可测信息向量, $\theta \in \mathbf{R}^n$ 是要估计的参数向量.

线性回归模型与伪线性回归模型的区别在于: 信息向量 $\varphi(t)$ 是可测的,就是线性回归模型; 信息向量 $\varphi(t)$ 是不可测的(即信息向量有的元是未知的),就是伪线性回归模型.

值得指出的是: 线性回归模型和伪线性回归模型是指系统输出或其他变量是参数的线性函数,与系统是否是线性无关. 因此,线性回归模型和伪线性回归模型描述的系统可以线性控制系统,也可以是非线性控制系统. 例如,非线性系统(30)可以写为线性回归模型(31),因此下面讨论的递推最小二乘辨识算法和随机梯度辨识算法的收敛结论,对非线性系

统(30)也成立. 动态系统的 CAR 模型 $[A(z)y(t) = B(z)u(t) + v(t)]$, AR 模型 $[A(z)y(t) = v(t)]$ 和 FIR 模型 $[y(t) = B(z)u(t) + v(t)]$ 可以写为线性回归模型,但线性回归模型不限于这些模型.

5.1.1 递推最小二乘算法

1) 基本算法

参考文献[39, 52],辨识系统(36)参数向量 θ 的递推最小二乘算法(Recursive Least Squares algorithm, RLS)为

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)], \quad (37) \\ P^{-1}(t) &= P^{-1}(t-1) + \varphi(t)\varphi^T(t) \quad P(0) = p_0I > 0. \quad (38) \end{aligned}$$

在分析递推最小二乘算法的收敛性时,使用式(37)~(38)的形式更为方便. 在计算参数估计 $\hat{\theta}(t)$ 时,为避免参数估计误差协方差阵(covariance matrix) $P(t) \in \mathbf{R}^{n \times n}$ 的求逆运算,通常运用矩阵求逆公式和引入增益向量 $L(t) := P(t)\varphi(t) \in \mathbf{R}^{n \times n}$,将算法(37)~(38)化为一种等价形式.

将矩阵求逆公式

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}. \quad (39)$$

应用式(38),可得

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{1 + \varphi^T(t)P(t-1)\varphi(t)}. \quad (40)$$

因此,递推最小二乘算法(37)~(38)可等价表示为

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)], \quad (41) \\ P(t) &= P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{1 + \varphi^T(t)P(t-1)\varphi(t)}, \\ P(0) &= p_0I. \quad (42) \end{aligned}$$

式(40)两边右乘向量 $\varphi(t)$ 可得

$$\begin{aligned} P(t)\varphi(t) &= P(t-1)\varphi(t) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)\varphi(t)}{1 + \varphi^T(t)P(t-1)\varphi(t)} = \\ P(t-1)\varphi(t) &\left[1 - \frac{\varphi^T(t)P(t-1)\varphi(t)}{1 + \varphi^T(t)P(t-1)\varphi(t)}\right] = \\ \frac{P(t-1)\varphi(t)}{1 + \varphi^T(t)P(t-1)\varphi(t)} &= L(t). \quad (43) \end{aligned}$$

借助于增益向量 $L(t)$,式(40)可以写为

$$\begin{aligned} P(t) &= P(t-1) - L(t)\varphi^T(t)P(t-1) = \\ [I - L(t)\varphi^T(t)]P(t-1) &= P(0). \end{aligned}$$

因此,递推最小二乘算法可表达为

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + L(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)], \quad (44) \\ L(t) &= P(t-1)\varphi(t)[1 + \varphi^T(t)P(t-1)\varphi(t)]^{-1}, \quad (45) \end{aligned}$$

$$P(t) = [I - L(t) \varphi^T(t)] P(t-1) \quad P(0) = p_0 I. \quad (46)$$

$\hat{\theta}(t)$ 为 t 时刻 θ 的估计, $L(t)$ 称为增益向量 (gain vector) $P(t)$ 称为协方差矩阵或协方差阵 (covariance matrix) $e(t) := y(t) - \varphi^T(t) \hat{\theta}(t-1)$ 称为新息 (innovation).

定义残差 (residual) $\varepsilon(t) := y(t) - \varphi^T(t) \hat{\theta}(t)$. 新息与残差的区别在于: 假设当前时刻是 t , 当前观测是 $y(t)$ 和 $\varphi(t)$, 残差表达式是使用当前观测数据计算之后的参数估计 $\hat{\theta}(t)$, 而新息是使用之前的参数估计 $\hat{\theta}(t-1)$. 也就是说, 新息使用刷新前的参数估计, 残差使用刷新后的参数估计. 新息与残差有下列关系,

$$\begin{cases} e(t) = [1 + \varphi^T(t) P(t-1) \varphi(t)] \varepsilon(t), \\ \varepsilon(t) = [1 - \varphi^T(t) P(t) \varphi(t)] e(t). \end{cases}$$

事实上

$$\begin{aligned} e(t) &= y(t) - \varphi^T(t) \hat{\theta}(t) = \\ &= y(t) - \varphi^T(t) [\hat{\theta}(t-1) + L(t) e(t)] = \\ &= y(t) - \varphi^T(t) \hat{\theta}(t-1) - \varphi^T(t) L(t) e(t) = \\ &= e(t) - \varphi^T(t) L(t) e(t) = \\ &= [1 - \varphi^T(t) L(t)] e(t) = \\ &= [1 - \varphi^T(t) P(t) \varphi(t)] e(t) = \\ &= \left[1 - \varphi^T(t) \frac{P(t-1) \varphi(t)}{1 + \varphi^T(t) P(t-1) \varphi(t)} \right] e(t) = \\ &= \frac{e(t)}{1 + \varphi^T(t) P(t-1) \varphi(t)}. \end{aligned}$$

协方差阵有下列几种表示,

$$\begin{aligned} P(t) &= \\ P(t-1) - \frac{P(t-1) \varphi(t) \varphi^T(t) P(t-1)}{1 + \varphi^T(t) P(t-1) \varphi(t)} &= \\ P(t-1) - P(t-1) \varphi(t) [1 + \varphi^T(t) P(t-1) \varphi(t)]^{-1} \varphi^T(t) P(t-1) &= \end{aligned}$$

$$\begin{aligned} P(t-1) - L(t) [1 + \varphi^T(t) P(t-1) \varphi(t)] L^T(t) &= \\ [I - L(t) \varphi^T(t)] P(t-1) &= \\ P(t-1) [I - \varphi(t) L^T(t)] &= \\ P(t-1) - L(t) [P(t-1) \varphi(t)]^T. \end{aligned}$$

这里利用了 $P(t)$ 是对称阵的性质. 用计算机实现 RLS 算法, 计算量最小的 RLS 算法表达如下:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \varphi^T(t) \hat{\theta}(t-1)]; \quad (47)$$

$$L(t) = \frac{[P(t-1) \varphi(t)]}{1 + \varphi^T(t) [P(t-1) \varphi(t)]}; \quad (48)$$

$$\begin{aligned} P(t) &= P(t-1) - L(t) [P(t-1) \varphi(t)]^T, \\ P(0) &= p_0 I. \end{aligned} \quad (49)$$

在递推算法中, 初值通常选择为 $P(0) =$ 很大对称正定阵, 如 $P(0) = p_0 I$ $p_0 = 10^6 \gg 1$; $\hat{\theta}(0) =$ 很小实向量, 如 $\hat{\theta}(0) = \mathbf{1}_n / p_0$. 其中 $\mathbf{1}_n$ 是元均为 1 的 n 维列向量, 即 $\mathbf{1}_n = [1, 1, \dots, 1]^T \in \mathbf{R}^n$. $P(0)$ 取一个大对称正定阵是根据 $P^{-1}(t)$ 的定义确定的^[52].

2) 递推最小二乘算法的计算量

评价辨识算法的计算量大小, 一般用算法的乘法次数 (包括除法次数) 和加法次数 (包括减法次数) 来衡量. 一次乘法次数或一次加法次数称为一个 flop^[67]. 乘法次数与加法次数之和的 flop 数就是算法计算量. 如果同为递推算法, 可以比较每递推计算一步的 flop 数, 来判断算法计算量的大小. 表 1 列出了 RLS 算法 (47) — (49) 每一步递推计算中的乘法次数、加法次数和 flop 数.

3) 递推最小二乘算法计算步骤

递推最小二乘算法 (47) — (49) 的计算步骤如下:

① 令 $t = 1$, 置参数估计初值 $\hat{\theta}(0) = \mathbf{1}_n / p_0$, 协方差阵初值 $P(0) = p_0 I$ $p_0 = 10^6$;

表 1 RLS 算法每步的计算量

Table1 Computational efficiency of the RLS algorithm

| | 计算次序 | flop 数 | |
|-------------------|---|-------------|-------------|
| | | 乘法次数 | 加法次数 |
| $\hat{\theta}(t)$ | $\alpha := y(t) - \varphi^T(t) \hat{\theta}(t-1)$ | n | n |
| | $\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) \alpha$ | n | n |
| 变量 | $Q := P(t-1) \varphi(t)$ | n^2 | $(n-1)n$ |
| | $L(t) = Q / [1 + \varphi^T(t) Q]$ | $2n$ | n |
| $P(t)$ | $S := L(t) [Q]^T$ | n^2 | |
| | $P(t) = P(t-1) - S$ | | n^2 |
| 总次数 | | $2n^2 + 4n$ | $2n^2 + 2n$ |
| 总 flop 数 | | $4n^2 + 6n$ | |

- ② 采集观测数据 构成 $y(t)$ 和信息向量 $\varphi(t)$;
 - ③ 用式(48) 计算增益向量 $L(t)$,用式(49) 计算协方差阵 $P(t)$;
 - ④ 通过式(47) 刷新参数估计向量 $\hat{\theta}(t)$;
 - ⑤ t 增 1 转到第 2 步 继续进行递推计算.
- RLS 算法计算参数估计 $\hat{\theta}(t)$ 流程如图 8 所示.

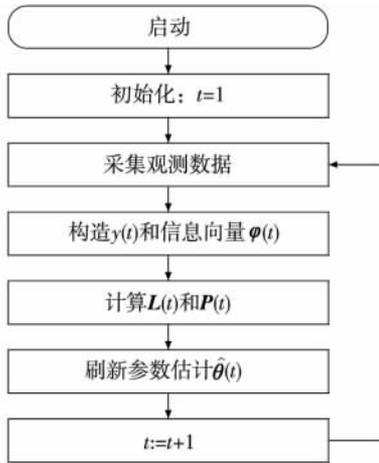


图 8 RLS 算法计算参数估计 $\hat{\theta}(t)$ 的流程

Fig.8 The flowchart for computing the parameter estimate $\hat{\theta}(t)$

4) 收敛定理

设 $\{v(t), \mathcal{F}_t\}$ 是鞅差序列 (martingale difference sequence) \mathcal{F}_t 是由直到 t 时刻的观测生成的 σ 代数序列, 即 $\mathcal{F}_t = \sigma(y(t), \varphi(t), y(t-1), \varphi(t-1), \dots, y(0), \varphi(0))$. 序列 $\{v(t)\}$ 满足噪声假设:

- (C1) $E[v(t) | \mathcal{F}_{t-1}] = 0$ a. s.
- (C2) $E[v^2(t) | \mathcal{F}_{t-1}] = \sigma^2(t) \leq \sigma^2 < \infty$ a. s.
- (C3) $E[v^2(t) | \mathcal{F}_{t-1}] = \sigma^2(t) \leq \sigma^2 r^\varepsilon(t)$ a. s. ,

$0 \leq \varepsilon < 1$.

即 $\{v(t)\}$ 相当于是零均值, 时变方差 $\sigma^2(t)$ 的不相关非平稳随机噪声序列. 上式中 $r(t)$ 在不同算法的收敛性定理中有不同的定义.

由式(38) 可得

$$\begin{aligned}
 P^{-1}(t) &= P^{-1}(t-1) + \varphi(t) \varphi^T(t) = \\
 P^{-1}(t-2) &+ \varphi(t-1) \varphi^T(t-1) + \varphi(t) \varphi^T(t) = \\
 \sum_{j=1}^t \varphi(j) \varphi^T(j) &+ P^{-1}(0) = \sum_{j=1}^t \varphi(j) \varphi^T(j) + \frac{I}{p_0}. \quad (50)
 \end{aligned}$$

对于 RLS 算法 $r(t)$ 定义为协方差阵 $P(t)$ 的逆的迹 (trace):

$$r(t) := \text{tr}[P^{-1}(t)] = \sum_{j=1}^t \|\varphi(j)\|^2 + n/p_0. \quad (51)$$

定理 13 对于系统(36) 和 RLS 算法(37) — (38) 假设 (C1) 和 (C2) 成立 那么对任意 $c > 1$, RLS

算法给出的参数估计误差满足^[43, 68]

- 1) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{[\ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right)$ a. s.
- 2) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln r(t) [\ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right)$ a. s.
- 3) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln r(t) \ln \ln r(t) [\ln \ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right)$ a. s.
- 4) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln r(t) \ln \ln r(t) \ln \ln \ln r(t) [\ln \ln \ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right)$ a. s.

由非负定矩阵的迹 $\text{tr}[X]$, 最大特征值 $\lambda_{\max}[X]$, 行列式 $\det[X] := |X|$ 之间的关系, 可知

$r(t) = O(\lambda_{\max}[P^{-1}(t)])$, $\ln r(t) = O(\ln |P^{-1}(t)|)$, 因此定理 13 中 $r(t)$ 分别用 $\lambda_{\max}[P^{-1}(t)]$ 和 $|P^{-1}(t)|$ 代替 结论依然成立.

因为 p_0 很大, 所以式(50) 右边最后一项可以忽略. 从式(51) $r(t)$ 和式(50) $P^{-1}(t)$ 的表达式可知: 定理 13 中 RLS 参数估计(误差) 只依赖于观测数据 $\varphi(t)$. 如果下列广义弱持续激励条件成立^[15, 43, 68],

$$\text{(GWPE)} \quad \alpha I \leq \frac{1}{t} \sum_{i=1}^t \varphi(i) \varphi^T(i) \leq \beta t^\gamma I \text{ a. s. ,}$$

对充分大 t . 则定理 13 的结论可以表示为

- 1) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{[\ln t]^c}{t}\right) \rightarrow 0$ a. s.
- 2) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln t [\ln \ln t]^c}{t}\right) \rightarrow 0$ a. s.
- 3) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln t \ln \ln t [\ln \ln \ln t]^c}{t}\right) \rightarrow 0$ a. s.
- 4) $\|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln t \ln \ln t \ln \ln \ln t [\ln \ln \ln \ln t]^c}{t}\right) \rightarrow 0$ a. s.

定理 13 说明: 当噪声 $\{v(t)\}$ 的方差有界时, 最小二乘参数估计的收敛速率是 $P^{-1}(t)$ 的最大特征值 (greatest eigenvalue) 或迹的对数的 c 次方与 $P^{-1}(t)$ 的最小特征值 (smallest eigenvalue) 之比. 下面的定理 14 说明当噪声方差无界时, 最小二乘参数估计误差仍然收敛于零.

定理 14 对于系统(36) 和 RLS 算法(37) — (38) 假设 (C1) 和 (C3) 成立 那么 RLS 算法给出的参数估计误差满足

$$\|\hat{\theta}(t) - \theta\|^2 = o\left(\frac{r(t)}{\lambda_{\min}[P^{-1}(t)]}\right) =$$

$$o\left(\frac{\lambda_{\max}[\mathbf{P}^{-1}(t)]}{\lambda_{\min}[\mathbf{P}^{-1}(t)]}\right) \text{ a. s.}$$

如果下列弱持续激励条件 1 成立,

$$(WPE1) \quad \alpha \mathbf{I} \leq \frac{1}{t} \sum_{j=1}^t \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j) \leq \beta \mathbf{I} \text{ a. s. 对}$$

充分大 t . 则有 $\lim_{t \rightarrow \infty} \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = 0$ a. s., 即参数估计误差收敛于零.

5.1.2 随机梯度辨识算法

线性回归模型(36)重写为

$$y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + v(t) \quad \boldsymbol{\varphi}(t) \in \mathbf{R}^n \quad \boldsymbol{\theta} \in \mathbf{R}^n. \quad (52)$$

估计参数向量 $\boldsymbol{\theta}$ 的随机梯度算法(Stochastic Gradient algorithm, SG)为

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r(t)} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (53)$$

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2, \\ r(0) = 1 \text{ 或 } r(0) = n/p_0. \quad (54)$$

定理 15^[69-70] 对于系统(53)和 SG 算法(53)–(54), 定义

$$\mathbf{R}(t) := \sum_{j=1}^t \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^T(j), \quad r(t) := \text{tr}[\mathbf{R}(t)],$$

假设(C1)和(C3)成立, $r(t) \rightarrow \infty$ 和 $r(t) = O(\lambda_{\min}[\mathbf{R}(t)])$, 那么参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{r(t)}{\lambda_{\min}[\mathbf{R}(t)]}\right) \rightarrow 0 \text{ a. s.}$$

即参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 一致收敛于真参数向量 $\boldsymbol{\theta}$. SG 算法中没有协方差阵 $\mathbf{P}(t)$, 但上述 $\mathbf{R}(t)$ 定义类似于 $\mathbf{P}^{-1}(t)$.

矩阵 $\mathbf{R}(t)$ 是由输入输出构成的数据乘积矩阵. 因为 $\mathbf{R}(t)$ 是非负定对称矩阵, 它是随数据增加而不断递增, 又 $r(t) \leq \lambda_{\max}[\mathbf{R}(t)]$ 和 $\lambda_{\max}[\mathbf{R}(t)] \leq r(t)$ 成立, 所以 $r(t) = O(\lambda_{\min}[\mathbf{R}(t)])$ 意味着 $\mathbf{R}(t)$ 的条件数有界, 相当于关于信息向量 $\boldsymbol{\varphi}(t)$ 的弱持续激励条件(WPE1)成立.

值得指出的是: 定理 15 的条件(C1)和(C3)是迄今为止, 我们找到的随机梯度算法参数估计误差二阶矩收敛的最弱条件^[69-70]. 与文献[54-60]不同的是, 这里并没有假设噪声 $\{v(t)\}$ 的 2 阶矩有限和高阶矩存在 $E[v^\gamma(t) | \mathcal{F}_{t-1}] < \infty$ a. s. 或 $E[v^\gamma(t)] < \infty$, 对于某 $\gamma > 2$. 事实上, 证明参数估计误差 2 阶矩的收敛性要求噪声高阶矩存在的条件是苛刻的.

尽管 SG 参数估计误差收敛于零, 当收敛速度很慢. 为了加快 SG 算法的收敛速度, 可在式(54)中引入遗忘因子 λ , 得到遗忘因子随机梯度算法(FF-SG), 简称遗忘梯度算法(Forgetting Gradient algo-

gorithm, FG)^[61, 69-70]:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r(t)} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (55)$$

$$r(t) = \lambda r(t-1) + \|\boldsymbol{\varphi}(t)\|^2 \quad 0 \leq \lambda < 1, \\ r(0) = 1 \text{ 或 } r(0) = n/p_0. \quad (56)$$

遗忘因子可以加快算法开始工作阶段的参数估计收敛速度, 但会增大参数估计误差的方差, 参数估计误差并不趋于零(对随机系统而言). 在随机梯度辨识算法中引入收敛指数(convergence index) ε , 就得到修正随机梯度算法(Modification Stochastic Gradient algorithm, M-SG), 或称为 Epsilon 随机梯度算法(Epsilon Stochastic Gradient algorithm, ε -SG)^[71-72]:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r^\varepsilon(t)} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \\ \frac{1}{2} < \varepsilon \leq 1, \quad (57)$$

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2 \quad r(0) = 1. \quad (58)$$

这里 $1/r^\varepsilon(t)$ 是收敛因子或步长. 修正随机梯度算法是本文作者最近提出的, 它比随机梯度算法具有更高的收敛速度, 其性能优于遗忘梯度算法^[71-72].

5.2 伪线性回归模型 I

一种简单的伪线性回归模型(pseudo-linear regression model)为

$$y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + v(t). \quad (59)$$

其中: $\{y(t)\}$ 是观测输出序列, $\{v(t)\}$ 是零均值不相关随机噪声序列, $\boldsymbol{\varphi}(t) \in \mathbf{R}^n$ 是观测数据以及不可测噪声构成的信息向量(也就是说 $\boldsymbol{\varphi}(t)$ 中有些元是不可测的), $\boldsymbol{\theta} \in \mathbf{R}^n$ 是要估计的参数向量.

伪线性回归模型 I 中的信息向量可以写作

$$\boldsymbol{\varphi}(t) = [\boldsymbol{\phi}^T(t) \quad v(t-1) \quad v(t-2) \quad \cdots \quad v(t-n)]^T \in \mathbf{R}^n.$$

其中: $\boldsymbol{\phi}(t) \in \mathbf{R}^n$ 是观测数据构成的可测信息向量, $v(t-i)$ 是不可测的未知量噪声项, 在辨识算法中一般用其估计值 $\hat{v}(t-i)$ 代替. 伪线性回归模型 I 的一种等价形式是

$$y(t) = \boldsymbol{\phi}^T(t) \boldsymbol{\theta}_1 + D(z) v(t). \quad (60)$$

由于子信息向量 $\boldsymbol{\phi}(t)$ 可能包含系统输入输出的非线性项, 所以伪线性回归模型并不限于线性系统.

本文的伪线性回归模型 I 中信息向量 $\boldsymbol{\varphi}(t) \in \mathbf{R}^n$ 只包含白噪声 $v(t)$ 的回归项 $v(t-i)$ (如果(还)包括了其他相关噪声的回归项, 就是伪线性回归模型 II). 动态系统的 CARMA 模型($A(z)y(t) = B(z)u(t) + D(z)v(t)$), ARMA 模型($A(z)y(t) = D(z)v(t)$), MA 模型($y(t) = D(z)v(t)$) 和 C-MA 模型($y(t) = B(z)u(t) + D(z)v(t)$) 可以写为伪线性回归

模型 I 的形式,但伪线性回归模型 I 不限于这些模型.

定义增广参数向量 θ 和信息向量 $\varphi_0(t)$ 分别为

$$\theta = [\theta_1^T \ d_1 \ d_2 \ ; \ \dots \ d_{n_d}]^T \in \mathbf{R}^n \quad n = n_a + n_b + n_d,$$

$$\theta_1 = [a_1 \ a_2 \ ; \ \dots \ a_{n_a} \ b_1 \ b_2 \ ; \ \dots \ b_{n_b}]^T \in \mathbf{R}^{n_a+n_b},$$

$$\varphi_0(t) := [\phi^T(t) \ v(t-1) \ v(t-2) \ ; \ \dots \ v(t-n_d)]^T \in \mathbf{R}^n,$$

$$\phi(t) := [-y(t-1) \ , -y(t-2) \ ; \ \dots \ , -y(t-n_a) \ ,$$

$$u(t-1) \ \mu(t-2) \ ; \ \dots \ \mu(t-n_b)]^T \in \mathbf{R}^{n_a+n_b}.$$

则 CARMA 系统 $A(z)y(t) = B(z)u(t) + D(z)v(t)$ 可以写为下列伪线性回归模型形式,

$$y(t) = \varphi_0^T(t) \theta + v(t). \quad (61)$$

下面基于伪线性模型(61)的递推增广最小二乘辨识算法和增广随机梯度辨识算法的收敛结论,对伪线性系统(59)也成立.

5.2.1 递推增广最小二乘算法

估计 CARMA 模型(61)参数向量 θ 的递推增广最小二乘算法(Recursive Extended Least Squares algorithm,RELS)可表述为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t) \hat{\varphi}(t) [y(t) - \hat{\varphi}^T(t) \hat{\theta}(t-1)] \quad (62)$$

$$P^{-1}(t) = P^{-1}(t-1) + \hat{\varphi}(t) \hat{\varphi}^T(t) \quad P(0) = p_0 I, \quad (63)$$

$$\hat{\varphi}(t) = [\phi^T(t) \ \hat{v}(t-1) \ \hat{v}(t-2) \ ; \ \dots \ \hat{v}(t-n_d)]^T, \quad (64)$$

$$\phi(t) = [-y(t-1) \ , -y(t-2) \ ; \ \dots \ , -y(t-n_a) \ ,$$

$$u(t-1) \ \mu(t-2) \ ; \ \dots \ \mu(t-n_b)]^T, \quad (65)$$

$$\hat{v}(t) = y(t) - \hat{\varphi}^T(t) \hat{\theta}(t). \quad (66)$$

在上述 RELS 算法中,没有使用式(61)中不可测向量 $\varphi_0(t)$,而是使用可测向量 $\hat{\varphi}(t)$,它是用估计残差 $\hat{v}(t-i)$ 代替 $\varphi_0(t)$ 中的不可测噪声 $v(t-i)$ 得到的,故这个 RELS 算法可以实现.

由式(63)可得

$$P^{-1}(t) = \sum_{j=1}^t \hat{\varphi}(j) \hat{\varphi}^T(j) + \frac{I}{p_0} \quad (67)$$

定义

$$r(t) := \text{tr}[P^{-1}(t)] = \sum_{j=1}^t \|\hat{\varphi}(j)\|^2 + n/p_0, \quad (68)$$

$$P_0^{-1}(t) := \sum_{j=1}^t \varphi_0(j) \varphi_0^T(j) + \frac{I}{p_0}, \quad (69)$$

$$r_0(t) := \text{tr}[P_0^{-1}(t)] = \sum_{j=1}^t \|\varphi_0(j)\|^2 + n/p_0. \quad (70)$$

与线性回归模型相比,伪线性回归模型 I 包含了相关噪声项 $D(z)v(t)$,所以在其收敛性定理中多了对 $D(z)$ 的约束.关于 RELS 算法的性能有下列引理和定理.

定理 16^[65,73] 对于系统(61)和 RELS 算法(62)——(66),假设(C1)和(C2)成立 $D^{-1}(z) - \frac{1}{2}$ 是

严格正实函数,那么对任意 $c > 1$,RELS 算法给出的参数估计误差满足

$$1) \quad \|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{[\ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right) \text{ a. s.}$$

$$2) \quad \|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln r(t) [\ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right) \text{ a. s.}$$

$$3) \quad \|\hat{\theta}(t) - \theta\|^2 =$$

$$O\left(\frac{\ln r(t) \ln \ln r(t) [\ln \ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right) \text{ a. s.}$$

$$4) \quad \|\hat{\theta}(t) - \theta\|^2 =$$

$$O\left(\frac{\ln r(t) \ln \ln r(t) \ln \ln \ln r(t) [\ln \ln \ln \ln r(t)]^c}{\lambda_{\min}[P^{-1}(t)]}\right) \text{ a. s.}$$

RELS 算法定理 16 的结论与 RLS 算法定理 13 的结论很类似,但有本质区别.因为 RLS 算法中 $P^{-1}(t)$ [参见式(50)]和 $r(t)$ [参见式(51)]只与观测信息向量 $\varphi(t)$ 有关,所以定理 13 中 RLS 参数估计(误差)只依赖于观测信息向量 $\varphi(t)$.然而,RELS 算法中 $P^{-1}(t)$ [参见式(67)]和 $r(t)$ [参见式(68)]与算法算出的向量 $\hat{\varphi}(t)$ 有关, $\hat{\varphi}(t)$ 包含了估计的噪声项 $\hat{v}(t-i)$,它反过来是根据参数估计 $\hat{\theta}(t)$ 计算的 [参见式(66)].也就是说,定理 16 结论中参数估计误差收敛速度通过式(67)、(68)和式(66)依赖于自己,依赖算法本身,这是不期望的.一种办法是增加条件,使 RELS 算法参数估计误差依赖于 $P_0(t)$ 和 $r_0(t)$,不依赖于 $P(t)$ 和 $r(t)$.

假设 $[\ln r_0(t)]^c = o(\lambda_{\min}[P_0^{-1}(t)])$,可以证明 $r(t) = O(r_0(t))$, $\lambda_{\min}[P_0^{-1}(t)] = O(\lambda_{\min}[P^{-1}(t)])$,那么定理 16 的结论可以表示为

$$1) \quad \|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{[\ln r_0(t)]^c}{\lambda_{\min}[P_0^{-1}(t)]}\right) \text{ a. s.}$$

$$2) \quad \|\hat{\theta}(t) - \theta\|^2 = O\left(\frac{\ln r_0(t) [\ln \ln r_0(t)]^c}{\lambda_{\min}[P_0^{-1}(t)]}\right) \text{ a. s.}$$

$$3) \quad \|\hat{\theta}(t) - \theta\|^2 =$$

$$O\left(\frac{\ln r_0(t) \ln \ln r_0(t) [\ln \ln \ln r_0(t)]^c}{\lambda_{\min}[P_0^{-1}(t)]}\right) \text{ a. s.}$$

$$4) \quad \|\hat{\theta}(t) - \theta\|^2 =$$

$$O\left(\frac{\ln r_0(t) \ln \ln r_0(t) \ln \ln \ln r_0(t) [\ln \ln \ln \ln r_0(t)]^c}{\lambda_{\min}[P_0^{-1}(t)]}\right) \text{ a. s.}$$

定理 17^[65] 对于系统(61)和 RELS 算法(62)——(66),假设(C1)和(C3)成立 $D^{-1}(z) - \frac{1}{2}$ 是严格正实函数,那么 RELS 算法给出的参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{\lambda_{\max}[\mathbf{P}^{-1}(t)]}{\lambda_{\min}[\mathbf{P}^{-1}(t)]}\right) \text{ a. s.}$$

当 $[\ln r_0(t)]^c = o(\lambda_{\min}[\mathbf{P}_0^{-1}(t)])$ 时, 有

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{\lambda_{\max}[\mathbf{P}_0^{-1}(t)]}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

将矩阵求逆公式(39)应用于(63), 可以得到避免求协方差矩阵逆的 RELS 算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}(t) [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (71)$$

$$\mathbf{L}(t) = \mathbf{P}(t-1) \hat{\boldsymbol{\varphi}}(t) [1 + \hat{\boldsymbol{\varphi}}^T(t) \mathbf{P}(t-1) \hat{\boldsymbol{\varphi}}(t)]^{-1}, \quad (72)$$

$$\mathbf{P}(t) = [\mathbf{I} - \mathbf{L}(t) \hat{\boldsymbol{\varphi}}^T(t)] \mathbf{P}(t-1) \quad \mathbf{P}(0) = p_0 \mathbf{I}, \quad (73)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\phi}^T(t) \hat{y}(t-1) \hat{y}(t-2) \cdots \hat{y}(t-n_d)]^T, \quad (74)$$

$$\boldsymbol{\phi}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), \\ u(t-1) \mu(t-2), \cdots, \mu(t-n_b)]^T, \quad (75)$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (76)$$

这个算法中也可以用新息 $e(t) := y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)$ 代替式(74) $\hat{\boldsymbol{\varphi}}(t)$ 中的残差 $\hat{v}(t)$.

5.2.2 增广随机梯度辨识算法

伪线性回归模型(59)重写如下:

$$y(t) = \boldsymbol{\varphi}_0^T(t) \boldsymbol{\theta} + v(t) \quad \boldsymbol{\theta} \in \mathbf{R}^n;$$

$$\boldsymbol{\varphi}_0(t) := [\boldsymbol{\phi}^T(t) \quad p(t-1) \quad p(t-2) \cdots, \\ v(t-n_d)]^T \in \mathbf{R}^n. \quad (77)$$

基于信息向量 $\boldsymbol{\varphi}_0(t)$ 中不可测噪声 $v(t-i)$ 用其估计值代替的思想, 可以得到估计 $\boldsymbol{\theta}$ 的增广随机梯度算法 (Extended Stochastic Gradient algorithm, ESG). 如果信息向量中的不可测噪声 $v(t)$ 用其估计残差 $\hat{v}(t) := y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t)$ 代替, 就得到基于残差的 ESG 算法; 如果 $v(t)$ 用新息 $e(t) := y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)$ 代替, 就得到基于新息的 ESG 算法. 下面分别研究这 2 个算法的收敛性.

1) 基于残差的 ESG 算法

基于残差的 ESG 算法 (Residual based ESG algorithm, R-ESG) 可表示为

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (78)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad r(0) = 1, \quad (79)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\phi}^T(t) \hat{y}(t-1) \hat{y}(t-2) \cdots \hat{y}(t-n_d)]^T, \quad (80)$$

$$\boldsymbol{\phi}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), \\ u(t-1) \mu(t-2), \cdots, \mu(t-n_b)]^T, \quad (81)$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (82)$$

定理 18 对于系统(77)和 R-ESG 算法(78) — (82) 定义

$$\mathbf{R}(t) := \sum_{j=1}^t \hat{\boldsymbol{\varphi}}(j) \hat{\boldsymbol{\varphi}}^T(j),$$

$\mathbf{R}_0(t) := \sum_{j=1}^t \boldsymbol{\varphi}_0(j) \boldsymbol{\varphi}_0^T(j)$, $r_0(t) := \text{tr}[\mathbf{R}_0(t)]$.
假设 (C1) 和 (C2) 成立, $D(z)$ 严格正实, 且 $r(t) \rightarrow \infty$,

$$(C4) \quad \limsup_{t \rightarrow \infty} \frac{r(t)}{\lambda_{\min}[\mathbf{R}(t)]} < \infty \text{ a. s.}$$

那么参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{r(t)}{\lambda_{\min}[\mathbf{R}(t)]}\right) \rightarrow \infty \text{ a. s.}$$

如果 $[\ln r_0(t)]^c = o(\lambda_{\min}[\mathbf{P}_0^{-1}(t)])$ 且 $r_0 \rightarrow \infty$,

$$(C5) \quad \limsup_{t \rightarrow \infty} \frac{r_0(t)}{\lambda_{\min}[\mathbf{R}_0(t)]} < \infty \text{ a. s.}$$

则有

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{r_0(t)}{\lambda_{\min}[\mathbf{R}_0(t)]}\right) < \infty \text{ a. s.}$$

即参数估计误差一致收敛于真参数向量: $[\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}] \rightarrow 0$ a. s.

同样, 我们可以得到基于残差的遗忘因子增广随机梯度算法 (FF-R-ESG):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (83)$$

$$r(t) = \lambda r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad \rho < \lambda < 1 \quad r(0) = 1, \quad (84)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\phi}^T(t) \hat{y}(t-1) \hat{y}(t-2) \cdots \hat{y}(t-n_d)]^T, \quad (85)$$

$$\boldsymbol{\phi}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), \\ u(t-1) \mu(t-2), \cdots, \mu(t-n_b)]^T, \quad (86)$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (87)$$

基于残差的修正增广随机梯度算法 (Modified Extended Stochastic Gradient algorithm, M-ESG), 或称为 Epsilon 增广随机梯度算法 (Epsilon Extended Stochastic Gradient algorithm, ε -ESG):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r^\varepsilon(t)} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \\ \frac{1}{2} < \varepsilon \leq 1, \quad (88)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad r(0) = 1, \quad (89)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\phi}^T(t) \hat{y}(t-1) \hat{y}(t-2) \cdots \hat{y}(t-n_d)]^T, \quad (90)$$

$$\boldsymbol{\phi}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), \\ u(t-1) \mu(t-2), \cdots, \mu(t-n_b)]^T, \quad (91)$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (92)$$

FF-R-ESG 算法和 M-ESG 算法的参数估计误差收敛性分析是很困难的, 仍然是一个辨识研究课题.

2) 基于新息的 ESG 算法

基于新息的 ESG 算法 (Innovation based ESG algorithm, I-ESG) 可表示为

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} e(t), \quad (93)$$

$$e(t) = y(t) - \hat{\varphi}^T(t) \hat{\theta}(t-1), \quad (94)$$

$$r(t) = r(t-1) + \|\hat{\varphi}(t)\|^2 r(0) = 1, \quad (95)$$

$$\hat{\varphi}(t) = [\phi^T(t) \ e(t-1) \ e(t-2) \ \dots \ e(t-n_d)]^T, \quad (96)$$

$$\phi(t) = [-y(t-1) \ , -y(t-2) \ ; \dots \ , -y(t-n_a) \ , \ u(t-1) \ \mu(t-2) \ ; \dots \ \mu(t-n_b)]^T. \quad (99)$$

I-ESG 算法与 R-ESG 算法收敛条件略有不同, 只需将定理 18 中“ $D(z)$ 严格正实”改为“ $D(z) - 1/2$ 严格正实”就得到 I-ESG 算法的收敛结论.

同样, 我们可以得到基于新息的遗忘因子增广随机梯度算法 (FF-I-ESG):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\hat{\varphi}(t)}{r(t)} e(t), \quad (98)$$

$$e(t) = y(t) - \hat{\varphi}^T(t) \hat{\theta}(t-1), \quad (99)$$

$$r(t) = \lambda r(t-1) + \|\hat{\varphi}(t)\|^2 \quad 0 < \lambda < 1 \quad r(0) = 1, \quad (100)$$

$$\hat{\varphi}(t) = [\phi^T(t) \ e(t-1) \ e(t-2) \ \dots \ e(t-n_d)]^T, \quad (101)$$

$$\phi(t) = [-y(t-1) \ , -y(t-2) \ ; \dots \ , -y(t-n_a) \ , \ u(t-1) \ \mu(t-2) \ ; \dots \ \mu(t-n_b)]^T. \quad (102)$$

基于新息的修正增广随机梯度算法 (M-I-ESG), 或称为 Epsilon 增广随机梯度算法 (ε -ESG):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\hat{\varphi}(t)}{r^\varepsilon(t)} e(t) \quad \frac{1}{2} < \varepsilon \leq 1, \quad (103)$$

$$e(t) = y(t) - \hat{\varphi}^T(t) \hat{\theta}(t-1), \quad (104)$$

$$r(t) = r(t-1) + \|\hat{\varphi}(t)\|^2 \quad r(0) = 1, \quad (105)$$

$$\hat{\varphi}(t) = [\phi^T(t) \ e(t-1) \ e(t-2) \ \dots \ e(t-n_d)]^T, \quad (106)$$

$$\phi(t) = [-y(t-1) \ , -y(t-2) \ ; \dots \ , -y(t-n_a) \ , \ u(t-1) \ \mu(t-2) \ ; \dots \ \mu(t-n_b)]^T. \quad (107)$$

FF-R-ESG 算法和 M-I-ESG 算法的参数估计误差收敛性分析是很困难的, 仍然是一个辨识研究课题. 这个算法的参数估计误差界分析也是很困难的, 仍然是一个辨识研究课题.

5.3 伪线性回归模型 II

伪线性回归模型 I 中信息向量包含了不可测白噪声项 $v(t-i)$, 伪线性回归模型 II 的特征是信息向量包含了不可测输出项: 真实输出或无噪输出 $x(t-i)$. 本文的伪线性回归模型 II 就是指一类白噪声干扰的标准输出误差模型, 它具有下列形式,

$$y(t) = \frac{f(z) \mu(t)}{A(z)} + v(t),$$

其中 $f(z) \mu(t)$ 是参数的线性函数. 伪线性回归模型 II 可以是线性系统, 也可以是非线性系统. 当 $f(z) \mu(t) = B(z) u(t)$ 时, 我们就得到线性系统的标准输出误差模型 (Output Error model, OE),

$$y(t) = \frac{B(z)}{A(z)} u(t) + v(t), \quad (108)$$

其中 $\{u(t)\}$ 和 $\{y(t)\}$ 分别是系统的输入和输出序列, $\{v(t)\}$ 是零均值方差为 σ^2 的随机白噪声序列, $G(z) = \frac{B(z)}{A(z)}$ 为系统的传递函数, $A(z)$ 和 $B(z)$ 是单位后移算子 z^{-1} 的多项式, 如前所定义.

输出误差模型示意如图 9 所示, 其中 $x(t) := \frac{B(z)}{A(z)} u(t)$ 为系统的真实输出 (true output) 或无噪输出 (noise-free output) (不可得到), $y(t)$ 是 $x(t)$ 的含噪量测.

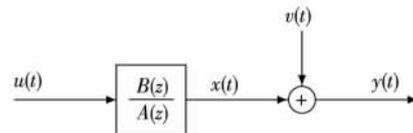


图 9 输出误差系统

Fig. 9 The output error system

定义参数向量 θ 和信息向量 $\varphi_0(t)$ 如下:

$$\theta = [a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b}]^T \in \mathbf{R}^n \quad n = n_a + n_b;$$

$$\varphi_0(t) = [-x(t-1) \ , -x(t-2) \ ; \dots \ , -x(t-n_a) \ , \ u(t-1) \ \mu(t-2) \ ; \dots \ \mu(t-n_b)]^T \in \mathbf{R}^n.$$

则可得输出误差系统的伪线性回归模型 II,

$$x(t) = \varphi_0^T(t) \theta,$$

$$y(t) = x(t) + v(t) = \varphi_0^T(t) \theta + v(t). \quad (109)$$

这个伪线性回归模型中信息向量 $\varphi_0(t)$ 包含了系统的不可测真实输出 $x(t-i)$.

5.3.1 辅助模型递推最小二乘法

输出误差模型的辨识是采用辅助模型辨识思想: 信息向量中不可测真实输出 $x(t-i)$ 用辅助模型的输出代替或估计的输出 $\hat{x}(t-i)$ 代替. 估计伪线性回归模型 (109) 参数向量 θ 的辅助模型递推最小二乘法 (Auxiliary Model based Recursive Least Squares algorithm, AM-RLS) 为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t) \hat{\varphi}(t) \cdot$$

$$[y^T(t) - \hat{\varphi}^T(t) \hat{\theta}(t-1)], \quad (110)$$

$$P^{-1}(t) = P^{-1}(t-1) + \hat{\varphi}(t) \hat{\varphi}^T(t) \quad P(0) = p_0 I, \quad (111)$$

$$\hat{\varphi}(t) = [-\hat{x}(t-1) \ , -\hat{x}(t-2) \ ; \dots \ , -\hat{x}(t-n_a) \ , \ u(t-1) \ \mu(t-2) \ ; \dots \ \mu(t-n_b)]^T, \quad (112)$$

$$\hat{x}(t) = \hat{\varphi}^T(t) \hat{\theta}(t). \quad (113)$$

定义

$$P_0^{-1}(t) := P_0^{-1}(t-1) + \varphi_0(t) \varphi_0^T(t) \quad P_0(0) = p_0 I,$$

$$r(t) := \text{tr}[P^{-1}(t)] =$$

$$\frac{n}{p_0} + \sum_{i=1}^t \|\hat{\varphi}(i)\|^2 = r(t-1) + \|\hat{\varphi}(t)\|^2,$$

$$r_0(t) := \text{tr}[\mathbf{P}_0^{-1}(t)] =$$

$$\frac{n}{p_0} + \sum_{i=1}^t \|\varphi_0(i)\|^2 = r_0(t-1) + \|\varphi_0(t)\|^2.$$

AM-RLS 算法(110) — (113) 的收敛结论与定理 16 和定理 17 类似.

定理 19^[15] 对于系统(109) 和 AM-RLS 算法(110) — (113) 假设(C1) 和(C2) 成立 $A^{-1}(z) - \frac{1}{2}$ 是严格正实函数, $[\ln r_0(t)]^c = o(\lambda_{\min}[\mathbf{P}_0^{-1}(t)])$ 那么对于 $c > 1$, AM-RLS 算法给出的参数估计误差满足

$$1) \quad \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = O\left(\frac{[\ln r_0(t)]^c}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

$$2) \quad \|\hat{\boldsymbol{\alpha}}(t) - \boldsymbol{\theta}\|^2 = O\left(\frac{\ln r_0(t) [\ln \ln r_0(t)]^c}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

$$3) \quad \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 =$$

$$O\left(\frac{\ln r_0(t) \ln \ln r_0(t) [\ln \ln \ln r_0(t)]^c}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

$$4) \quad \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 =$$

$$O\left(\frac{\ln r_0(t) \ln \ln r_0(t) \ln \ln \ln r_0(t) [\ln \ln \ln \ln r_0(t)]^c}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

定理 20^[15] 对于系统(109) 和 AM-RLS 算法(110) — (113) 假设(C1) 和(C3) 成立 $D^{-1}(z) - \frac{1}{2}$ 是严格正实函数, $[\ln r_0(t)]^c = o(\lambda_{\min}[\mathbf{P}_0^{-1}(t)])$ 那么 AM-RLS 算法给出的参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{\lambda_{\max}[\mathbf{P}_0^{-1}(t)]}{\lambda_{\min}[\mathbf{P}_0^{-1}(t)]}\right) \text{ a. s.}$$

将矩阵求逆公式(39) 应用于(111), 可以得到避免求协方差矩阵逆的 AM-RLS 算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}(t) [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (114)$$

$$\mathbf{L}(t) = \mathbf{P}(t-1) \hat{\boldsymbol{\varphi}}(t) [1 + \hat{\boldsymbol{\varphi}}^T(t) \mathbf{P}(t-1) \hat{\boldsymbol{\varphi}}(t)]^{-1}, \quad (115)$$

$$\mathbf{P}(t) = [\mathbf{I} - \mathbf{L}(t) \hat{\boldsymbol{\varphi}}^T(t)] \mathbf{P}(t-1) \quad \mathbf{P}(0) = p_0 \mathbf{I} \quad (116)$$

$$\hat{\boldsymbol{\varphi}}(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \dots, -\hat{x}(t-n_a)],$$

$$u(t-1) \quad \mu(t-2) \quad \dots \quad \mu(t-n_b)]^T, \quad (117)$$

$$\hat{x}(t) = \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (118)$$

5.3.2 辅助模型随机梯度辨识算法

估计伪线性回归模型(109) 参数向量 $\boldsymbol{\theta}$ 的辅助模型随机梯度算法(Auxiliary Model based Stochastic Gradient algorithm, AM-SG) 为

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (119)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad r(0) = 1, \quad (120)$$

$$\hat{\boldsymbol{\varphi}}(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \dots, -\hat{x}(t-n_a)],$$

$$u(t-1) \quad \mu(t-2) \quad \dots \quad \mu(t-n_b)]^T, \quad (121)$$

$$\hat{x}(t) = \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (122)$$

定理 21 对于系统(109) 和 AM-SG 算法(119) — (122) 定义

$$\mathbf{R}(t) := \sum_{j=1}^t \hat{\boldsymbol{\varphi}}(j) \hat{\boldsymbol{\varphi}}^T(j),$$

$$\mathbf{R}_0(t) := \sum_{j=1}^t \varphi_0(j) \varphi_0^T(j), \quad r_0(t) := \text{tr}[\mathbf{R}_0(t)].$$

假设(C1) 和(C2) 成立 $A(z)$ 严格正实, 且 $r(t) \rightarrow \infty$ $r(t) = O(\lambda_{\min}[\mathbf{R}(t)])$ a. s. 那么 AM-SG 算法给出的参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{r(t)}{\lambda_{\min}[\mathbf{R}(t)]}\right) \rightarrow 0 \text{ a. s.}$$

如果 $[\ln r_0(t)]^c = o(\lambda_{\min}[\mathbf{P}_0^{-1}(t)])$ 且 $r_0(t) \rightarrow \infty$, $r_0(t) = O(\lambda_{\min}[\mathbf{R}_0(t)])$ a. s. 则有

$$\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 = o\left(\frac{r(t)}{\lambda_{\min}[\mathbf{R}(t)]}\right) \rightarrow 0 \text{ a. s.}$$

我们可以得到辅助模型遗忘因子随机梯度算法(Auxiliary Model based Forgetting Factor Stochastic Gradient algorithm, AM-FFSG) 简称辅助模型遗忘梯度辨识算法(Auxiliary Model based Forgetting Gradient algorithm, AM-FG):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (123)$$

$$r(t) = \lambda r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad \rho < \lambda < 1 \quad r(0) = 1, \quad (124)$$

$$\hat{\boldsymbol{\varphi}}(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \dots, -\hat{x}(t-n_a)],$$

$$u(t-1) \quad \mu(t-2) \quad \dots \quad \mu(t-n_b)]^T, \quad (125)$$

$$\hat{x}(t) = \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (126)$$

辅助模型修正随机梯度算法(Auxiliary Model based Modified Stochastic Gradient algorithm, AM-M-SG) 或称为辅助模型 Epsilon 随机梯度算法(Auxiliary Model based Epsilon Extended Stochastic Gradient algorithm, AM- ε -ESG):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r^\varepsilon(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t-1)],$$

$$\frac{1}{2} < \varepsilon \leq 1, \quad (127)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 \quad r(0) = 1, \quad (128)$$

$$\hat{\boldsymbol{\varphi}}(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \dots, -\hat{x}(t-n_a)],$$

$$u(t-1) \quad \mu(t-2) \quad \dots \quad \mu(t-n_b)]^T, \quad (129)$$

$$\hat{x}(t) = \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\theta}}(t). \quad (130)$$

AM-FFSG 算法和 AM-M-SG 算法的收敛性分析仍是有待解决的辨识难题.

5.4 伪线性回归模型 III

伪线性回归模型 I 的信息向量中只涉及未知白噪声项 $v(t-i)$, 伪线性回归模型 II 的信息向量中只涉及未知真实输出项 $x(t-i)$, 我们把信息向量涉及真实输出项 $x(t-i)$ 和(或)未知白噪声项 $v(t-i)$ 和(或)其他相关噪声项 $w(t-i)$ 的模型都归结为伪线性回归模型 III.

方程误差类的 CARAR 模型:

$$A(z)y(t) = B(z)u(t) + \frac{1}{C(z)}v(t).$$

CARARMA 模型:

$$A(z)y(t) = B(z)u(t) + \frac{D(z)}{C(z)}v(t).$$

输出误差类的输出误差滑动平均模型(Output Error Moving Average model, OEMA) [17-21]:

$$y(t) = \frac{B(z)}{A(z)}u(t) + D(z)v(t).$$

输出误差自回归模型(Output Error Autoregressive model, OEAR):

$$y(t) = \frac{B(z)}{A(z)}u(t) + \frac{1}{C(z)}v(t).$$

Box-Jenkins 模型(B-J model):

$$y(t) = \frac{B(z)}{A(z)}u(t) + \frac{D(z)}{C(z)}v(t).$$

这些线性系统都可以写成伪线性回归模型 III, 其中

$$\begin{aligned} \text{真实输出是 } x(t) &= \frac{B(z)}{A(z)}u(t) \text{ 相关噪声项为 } w(t): \\ &= \frac{1}{C(z)}v(t) \text{ 或 } w(t) = \frac{D(z)}{C(z)}v(t). \end{aligned}$$

6 结语

控制学科是一门基于数学模型的学科, 而系统辨识是建立系统数学模型的理论与方法. 因此, 辨识方法的收敛性性能、参数估计精度和系统可辨识性都是系统辨识的基本问题, 也给辨识的研究者提出了新的课题. 本文围绕这些问题, 简单讨论了辨识算法收敛结论, 总结和给出了一些基本的激励条件, 如强持续激励条件, 广义强持续激励条件, 弱持续激励条件, 广义弱持续激励条件, 衰减激励条件等, 希望能给辨识爱好者的深入研究提供一个轮廓.

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] 丁锋. 系统辨识(1): 辨识导引 [J]. 南京信息工程大学学报: 自然科学版, 2011, 3(1): 1-22

DING Feng. System identification. Part A: Introduction to the identification [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition 2011, 3(1): 1-22

[6] 丁锋. 系统辨识(2): 系统描述的基本模型 [J]. 南京信息工程大学学报: 自然科学版, 2011, 3(2): 97-117

DING Feng. System identification. Part B: Basic models for system description [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition 2011, 3(2): 97-117.

[7] 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

[8] 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

[9] 丁锋, 丁韬, 萧德云, 等. 时变系统有限数据窗最小二乘辨识算法的有界收敛性 [J]. 自动化学报, 2002, 28(5): 754-761

DING Feng, DING Tao, XIAO Deyun, et al. Bounded convergence of finite data window least squares identification for time-varying systems [J]. Acta Automatica Sinica 2002, 28(5): 754-761.

[10] Ding F, Liu X P, 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
- [11] 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
- [12] Wang D Q, Yang G W, Ding F. Gradient-based iterative parameter estimation for Box-Jenkins systems with finite measurement data [J]. *Computers & Mathematics with Applications* 2010 60(5): 1200-1208
- [13] 丁锋 谢新民. 传递函数阵子模型参数递推估计: 辅助模型方法[J]. *控制与决策*, 1991 6(6): 447-452
DING Feng, XIE Xinmin. Recursive parameter estimation of transfer matrix sub-submodels: An auxiliary model method [J]. *Control and Decision*, 1991 6(6): 447-452
- [14] 丁锋. 多变量系统的辅助模型辨识方法的收敛性分析[J]. *控制理论与应用*, 1997 14(2): 192-200
DING Feng. Convergence analysis of auxiliary model identification algorithms for multivariable systems [J]. *Control Theory and Applications*, 1997 14(2): 192-200
- [15] Ding F, Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model [J]. *Automatica* 2004 40(10): 1739-1748
- [16] 王冬青, 丁锋. Box-Jenkins 模型的基于辅助模型的多新息广义增广随机梯度算法[J]. *控制与决策*, 2008, 23(9): 999-1003, 1010
WANG Dongqing, DING Feng. Auxiliary model based multi-innovation generalized extended stochastic gradient (AM-MI-GESG) algorithms for Box-Jenkins models [J]. *Control and Decision* 2008 23(9): 999-1003, 1010
- [17] 王冬青. 基于辅助模型的递推增广最小二乘辨识方法[J]. *控制理论与应用*, 2009 26(1): 51-56
WANG Dongqing. Recursive extended least squares identification method based on auxiliary models [J]. *Control Theory and Applications* 2009 26(1): 51-56
- [18] 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
- [19] 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
- [20] Ding F, Chen T. Performance analysis of multi-innovation gradient type identification methods [J]. *Automatica*, 2007 43(1): 1-14
- [21] Ding F, Liu X P, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises [J]. *Signal Processing* 2009 89(10): 1883-1890
- [22] Wang D Q, Ding F. Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems [J]. *Digital Signal Processing* 2010 20(3): 750-762
- [23] Ding F, Liu X P, Liu G. Multi-innovation least squares identification for system modeling [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010 40(3): 767-778
- [24] Liu Y J, Yu L, Ding F. Multi-innovation extended stochastic gradient algorithm and its performance analysis [J]. *Circuits, Systems and Signal Processing*, 2010, 29(4): 649-667
- [25] Ding F. Several multi-innovation identification methods [J]. *Digital Signal Processing* 2010 20(4): 1027-1039
- [26] 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
- [27] Ding F, Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems [J]. *Automatica* 2005 41(2): 315-325
- [28] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems [J]. *IEEE Transactions on Automatic Control* 2005 50(3): 397-402
- [29] Wang L Y, Ding F, Liu X P. Consistency of HLS estimation algorithms for MIMO ARX-like systems [J]. *Applied Mathematics and Computation*, 2007, 190(2): 1081-1093
- [30] Han H Q, Xie L, Ding F et al. Hierarchical least squares based iterative identification for multivariable systems with moving average noises [J]. *Mathematical and Computer Modelling* 2010 51(9/10): 1213-1220
- [31] Xiang L L, Xie L B, Ding R F. Hierarchical least squares algorithms for single-input multiple-output systems based on the auxiliary model [J]. *Mathematical and Computer Modelling* 2010 52(5/6): 918-924
- [32] Zhang Z N, Ding F, Liu X G. Hierarchical gradient based iterative parameter estimation algorithm for multivariable output error moving average systems [J]. *Computers & Mathematics with Applications* 2011 61(3): 672-682
- [33] 丁锋 陈通文 萧德云. 一般双率系统状态空间模型及其递阶辨识[J]. *自动化学报*, 2004 30(5): 652-663
DING Feng, CHEN Tongwen, XIAO Deyun. State-space modeling and hierarchical identification for general dual-rate stochastic systems [J]. *Acta Automatica Sinica*, 2004 30(5): 652-663
- [34] 丁锋 陈通文 萧德云. 非均匀周期采样多率系统的一种递阶辨识方法[J]. *电子学报*, 2004, 32(9): 1414-1420
DING Feng, CHEN Tongwen, XIAO Deyun. Hierarchical identification of non-uniformly periodically sampled multirate systems [J]. *Acta Electronica Sinica*, 2004, 32(9): 1414-1420.
- [35] 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
- [36] Ding F, Liu P X, Liu G. Identification methods for Hammerstein nonlinear systems [J]. *Digital Signal Processing* 2011 21(2): 215-238
- [37] Moustafa K A F. Identification of stochastic time-varying systems [J]. *IEE Proceeding, Part D, Control Theory and*

- Applications ,1983 ,130(4) : 137-142
- [38] Han L L ,Ding F. Multi-innovation stochastic gradient algorithms for multi-input multi-output systems [J]. Digital Signal Processing 2009 ,19(4) : 545-554
- [39] 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
- [40] 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
- [41] 丁锋, 萧德云, 丁韬. 衰减激励条件下最小均方算法的收敛性 [J]. 控制理论与应用 2003 20(1) : 109-112
DING Feng ,Xiao Deyun ,DING Tao. Convergence of least mean square algorithm under attenuating excitation [J]. Control Theory and Applications 2003 20(1) : 109-112
- [42] Liu Y J ,Sheng J ,Ding R F. Convergence of stochastic gradient algorithm for multivariable ARX-like systems [J]. Computers & Mathematics with Applications 2010 , 59(8) : 2615-2627
- [43] Ding F ,Liu P X ,Shi Y. Convergence analysis of estimation algorithms of dual-rate stochastic systems [J]. Applied Mathematics and Computation ,2006 ,176(1) : 245-261
- [44] 丁锋, 杨家本. 衰减激励条件下确定性系统多新息算法的收敛性分析 [J]. 清华大学学报: 自然科学版 , 1998 38(9) : 111-115
DING Feng ,YANG Jiaben. Convergence analysis of multi-innovation identification under attenuating excitation conditions for deterministic systems [J]. Journal of Tsinghua University: Science and Technology ,1998 ,38(9) : 111-115
- [45] 丁锋, 丁韬, 杨家本, 等. 随机系统多新息辨识在衰减激励条件下的性能分析 [J]. 清华大学学报: 自然科学版 2001 41(9) : 115-117
DING Feng ,DING Tao ,YANG Jiaben et al. Performance analysis of multi-innovation identification under attenuating excitation for stochastic systems [J]. Journal of Tsinghua University: Science and Technology ,2001 ,41(9) : 115-117
- [46] 丁锋, 丁韬, 杨家本, 等. 衰减激励条件下递阶最小二乘辨识的均方收敛性 [J]. 控制与决策 2002 ,17(1) : 6-10
DING Feng ,DING Tao ,YANG Jiaben et al. Mean square convergence of hierarchical least squares identification under the attenuating excitation [J]. Control and Decision 2002 ,17(1) : 6-10
- [47] 丁锋, 丁韬. 衰减激励条件下随机系统最小二乘的收敛性 [J]. 湖北工学院学报 2001 ,16(1) : 5-7
DING Feng ,Ding Tao. Convergence of least squares algorithms for stochastic systems under the attenuating excitation [J]. Journal of Hubei University of Technology , 2001 ,16(1) : 5-7
- [48] 郭雷. 时变随机系统: 稳定性、估计与控制 [M]. 长春: 吉林科学技术出版社 ,1993
GUO Lei. Time-varying stochastic systems: Stability ,estimation and control [M]. Changchun: Jilin Science and Technology Press ,1993
- [49] Goodwin G C ,Payne. Dynamic system identification-experiment design and data analysis [M]. Academic Press , New York ,1977
- [50] Ding F ,Chen T. Performance bounds of the forgetting factor least squares algorithm for time-varying systems with finite measurement data [J]. IEEE Transactions on Circuits and Systems I: Regular Papers ,2005 ,52(3) : 555-566
- [51] 丁锋. 鞅超收敛定理与遗忘因子最小二乘算法的收敛性分析 [J]. 控制理论与应用 ,1997 ,14(1) : 90-95
DING Feng. Martingale hyperconvergence theorem and convergence of the forgetting factor least squares algorithm [J]. Control Theory and Applications ,1997 ,14(1) : 90-95
- [52] 丁锋. 系统辨识理论与方法 [M]. 北京: 电力出版社 2012
DING Feng. System identification theory and methods [M]. Beijing: China Electric Power Press 2012
- [53] Ljung L. Consistency of the least-squares identification method [J]. IEEE Transactions on Automatic Control , 1976 21(5) : 779-781
- [54] Solo V. The Convergence of AML [J]. IEEE Transactions on Automatic Control ,1979 24(6) : 958-962
- [55] Lai T L ,Wei C Z. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems [J]. The Annals of Statistics , 1982 ,10(1) : 154-166
- [56] Lai T L ,Wei C Z. Extended least squares and their applications to adaptive control and prediction in linear systems [J]. IEEE Transactions on Automatic Control , 1986 31(10) : 898-906
- [57] Wei C Z. Adaptive prediction by least squares prediction in stochastic regression models [J]. The Annals of Statistics ,1987 ,15(4) : 1667-1682
- [58] Lai T L ,Ying Z L. Recursive identification and adaptive prediction in linear stochastic systems [J]. SIAM Journal on Control and Optimization ,1991 29(5) : 1061-1090
- [59] Toussi K ,Ren W. On the convergence least squares estimates in white noise [J]. IEEE Transactions on Automatic Control ,1994 39(2) : 364-368
- [60] Ren W ,Kumar P R. Stochastic adaptive prediction and model reference control [J]. IEEE Transactions on Automatic Control ,1994 39(10) : 2047-2060
- [61] 丁锋, 丁韬, 杨家本, 等. 时变参数遗忘梯度估计算法的收敛性 [J]. 自动化学报 2002 28(6) : 962-968
DING Feng ,DING Tao ,YANG Jiaben et al. Convergence of forgetting gradient estimation algorithm for time-varying parameters [J]. Acta Automatica Sinica ,2002 ,28(6) : 962-968
- [62] Goodwin G C ,Sin K S. Adaptive filtering ,prediction and control [M]. Prentice-Hall: Englewood Cliffs ,NJ ,1984
- [63] 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]. Department of Automation ,Tsinghua University ,Beijing ,1994
- [64] Ding F ,Chen T. Parameter estimation of dual-rate sto-

- chastic systems by using an output error method [J]. IEEE Transactions on Automatic Control ,2005 ,50(9) : 1436-1441
- [65] Ding F ,Chen T. Identification of Hammerstein nonlinear ARMAX systems [J]. Automatica , 2005 , 41 (9) : 1479-1489
- [66] 丁锋 杨家本. 关于鞅超收敛定理与遗忘因子最小二乘算法的收敛性分析 [J]. 控制理论与应用 ,1999 ,16(4) : 569-572
DING Feng ,YANG Jiaben. Remarks on the martingale hyperconvergence theorem and convergence analysis of the forgetting factor least squares algorithms [J]. Control Theory and Applications ,1999 ,16(4) : 569-572
- [67] Golub G H ,Van Loan C F. Matrix computations [M]. 3rd ed. Baltimore ,MD: Johns Hopkins Univ Press ,1996
- [68] Ding F ,Chen T. Modeling and identification for multirate systems [J]. Acta Automatica Sinica ,2005 ,31 (1) : 105-122
- [69] 丁锋 杨慧中 刘飞. 弱条件下随机梯度算法性能分析 [J]. 中国科学 E 辑: 信息科学 ,2008 ,38 (12) : 2173-2184
- [70] Ding F ,Yang H Z ,Liu F. Performance analysis of stochastic gradient algorithms under weak conditions [J]. Science in China Series F: Information Sciences ,2008 ,51 (9) : 1269-1280
- [71] Ding J ,Shi Y ,Wang H G ,Ding F. A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems [J]. Digital Signal Processing ,2010 ,20(4) : 1238-1247
- [72] Chen J ,Ding F. Modified stochastic gradient algorithms with fast convergence rates [J]. Journal of Vibration and Control ,2011 ,DOI: 10. 1177/1077546310376989
- [73] Ding F ,Chen T. Author's reply to "comments on 'identification of hammerstein nonlinear ARMAX systems'" [J]. Automatica ,2007 ,43(8) : 1497

System identification. Part C: Identification accuracy and basic problems

DING Feng^{1,2}

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

2 Control Science and Engineering Research Center ,Jiangnan University ,Wuxi 214122

Abstract System identification is the theory and methods of establishing the mathematical models of systems. This paper discusses some basic issues involved in system identification ,including the identification accuracy ,the identification methods ,the input signal design ,the parameter identifiability and system identifiability ,the open-loop identifiability and closed-loop identifiability ,the identifiability and the controllability and observability ,the identifiability and the input signal ,and the excitation signal and the excitation conditions relate to the convergence of identification methods ,and the convergene theorems of some typical identification algorithms.

Key words mathemaitcal model; system identification; parameter estimation; CAR model; CARMA model; CARAR model; CARARMA model; output error model; OEMA model; OEAR model