文章编号: 1674-7070(2012)01-0001-28

系统辨识(6):多新息辨识理论与方法

丁锋[1,2,3]

摘要

多新息辨识是系统辨识的一个重要分支.新息是能够改善参数估计精度或状态估计精度的有用信息.首先,详细讨论了线性回归模型的各种多新息辨识方法,包括多新息投影算法、多新息随机梯度算法、多新息遗忘梯度算法、变递推间隔多新息随机梯度算法、多新息最小二乘辨识方法、变递推间隔多新息最小二乘算法等;然后,给出了方程误差类系统、输出误差类系统、输入非线性系统的随机梯度辨识算法、多新息随机梯度算法和多新息最小二乘辨识算法;最后,简单说明了多新息辨识理论可以发展到多新息观测器和多新息卡尔曼滤波理论.

关键词

迭代辨识;递推辨识;参数估计;FIR模型;方程误差模型;CAR模型;CARMA模型;CARAR模型;CARARMA模型;输出误差模型;OEMA模型;OEAR模型;辅助模型辨识;多新息辨识;递阶辨识;耦合辨识

中图分类号 TP273

文献标志码 A

收稿日期 2011-12-14

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

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

3 江南大学 教育部轻工过程先进控制重点实验室,无锡,214122

0 引言

这是一个信息时代,是一个知识爆炸时代,归根结底是信息科学和自动化科学高度发展的时代.控制论和控制科学给我们认识世界提供了系统的方法论,给我们改造世界提供了最高效和最有力的手段.控制科学跨越时空的伟大成就——自动化电子产品的问世、电子设备计算能力和信息处理能力的提升、自动化设备和装备的出现,彻底改变了我们的生活方式.新思想、新理论、新原理、新概念的诞生都是科学史上的重要里程碑.就研究建立系统数学模型的理论与方法的系统辨识而言,辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念的诞生,有助于推动系统辨识学科的研究进程[1-6].

动态系统的数学模型是控制科学的基础.系统辨识是使用观测信息(即系统的输入输出数据)建立描述事物运动规律的数学模型.典型的随机梯度辨识方法、最小二乘辨识方法就是利用系统输入输出信息计算模型的参数.为了实时(real-time)获得系统模型参数,提出了递推辨识方法,它可以在线(on-line)递推计算模型参数.它的基本思想是当前时刻模型参数估计等于前一时刻参数估计加上增益向量与新息(innovation)的乘积进行校正,这样的递推计算方式可以提高计算效率.所谓新息就是指能够改善参数估计精度或状态估计精度的有用信息.

回顾辨识的发展史,自1967年国际自动控制联合会(IFAC,International Federation of Automatic Control)每3年组织一次"辨识与系统参数估计"专题讨论会以来,系统辨识参数估计方法和各种辨识应用软件工具得到长足发展,辨识方法的(有界)收敛性、收敛速率、估计误差上界的研究也取得了丰富的成果.但是,诞生新辨识方法族的确不多,可见提出新的辨识思想、辨识理论、辨识原理、辨识概念是极其重要的.

近年来,本文作者等提出和创立了辅助模型辨识思想(auxiliary model identification idea)[7-11]、多新息辨识理论(multi-innovation identification theory)[12-14]、递阶辨识原理(hierarchical identification principle)[15-18]、耦合辨识概念(coupled identification concept)[19]和参数估计误差界理论(parameter estimation error bound theory),发展了时不变系统的鞅收敛定理(martingale convergence theorem),建立了研究时变系统参数估计误差界的鞅超收敛定理(martingale hypercovergence theorem)等,进而提出和发展了辅助模型辨识方法[20-23]、多新息辨识方法[24-30]、递阶辨识方法[31-36]、耦合辨识方法[19]以及现存方法和新提出方法在不同条件下的性能分析等一系列研究成果,形成了一套理论体系. 这些方法与最小二乘法、卡尔曼滤波算法、最小均方算法一样,可用于解决多种模型的参数估计、自适应滤波和预测(adaptive filtering and prediction)、自适应信号处理(adaptive signal processing)和构成自适应控制等问题.

本文作者拓展了新息辨识的概念,将标量新息扩展到新息向量,将向量新息扩展到新息矩阵,从而提出和建立了一种基于新息的辨识理论与方法,简称为多新息辨识理论与方法. 最近,作者的多新息随机梯度型辨识方法和多新息最小二乘辨识方法 Regular Paper 分别发表在控制领域国际期刊《Automatica》2007 年第 1 期[12]和《IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics》2010 年第 3 期[13].

本文首先介绍用于线性回归模型的多新息辨识理论,详细推导了多新息随机梯度辨识方法,介绍了一些多新息梯度型辨识算法,推导了(变递推间隔)多新息最小二乘算法,给出了一些派生的多新息最小二乘类辨识算法;其次将多新息辨识理论用于有色噪声干扰的方程误差类系统和输出误差类系统,以及输入非线性系统的辨识,讨论了多新息随机梯度算法和多新息最小二乘算法;最后将多新息辨识理论应用于观测器设计和卡尔曼滤波,提出了多新息观测器和多新息卡尔曼滤波器. 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 多新息辨识理论
  1.1 什么是多新息辨识方法?
  1.2 变递推间隔多新息辨识方法
  1.3 多新息辨识的相关成果
2 多新息随机梯度辨识方法(MISG)
3 多新息梯度型辨识算法(MIGT)
4 多新息最小二乘辨识方法(MILS)
5 变递推间隔多新息最小二乘辨识方法(V-MILS)
6 多新息最小二乘类辨识算法(MILST)
7 方程误差类系统(EET)
  7.1 受控自回归系统(CAR)
  7.2 受控自回归滑动平均系统(CARMA)
  7.3 受控自回归自回归系统(CARAR)
  7.4 受控自回归自回归滑动平均系统(CARARMA)
8 输出误差类系统(OET)
  8.1 输出误差系统(OE)
  8.2 输出误差滑动平均系统(OEMA)
  8.3 输出误差自回归系统(OEAR)
  8.4 Box-Jenkins 系统(BJ)
9 输入非线性受控自回归自回归滑动平均系统(IN-CARARMA)
10 多新息观测器和多新息卡尔曼滤波器(MI-Observer and MI-KF)
  10.1 多新息观测器(MI-Observer)
  10.2 多新息卡尔曼滤波器(Multi-innovation Kalman filter)
11 结语

1 多新息辨识理论

多新息辨识理论是本文作者 1994 年在其博士论文《时变参数系统辨识及其应用》中提出的[37],其第 1 篇多新息辨识论文"时变系统辨识的多新息方法"发表在《自动化学报》1996 年第 1 期上[24],而另一篇关于多新息辨识方法的 Regular paper "Performance analysis of multi-innovation gradient type identification methods(多新息梯度型辨识方法:多新息随机梯度、多新息遗忘梯度算法的性能分析)"发表在控制界国际期刊《Automatica》2007 年第 1 期上[12]. 多新息辨识方法是受文献[38]算法间断迭代思想的启发,最初用类比方法,直接给出了变递推间隔多新息广义投影辨识算法的数学表达式[24]. 当时尚无法给出详细的理论推导,后经过深入研究,从理论上详细推导了多新息投影辨识算法、多新息随机梯度算法、多新息最小二乘辨识算法、变递推间隔多新息最小二乘算法等. 这使得多新息辨识算法有了严密的数学基础[12,14,23,27,39]. 本文研究多新息辨识理论与一些多新息辨识方法.

1.1 什么是多新息辨识方法?

在此文之前的连载论文讨论的一些辨识算法,如最小二乘类和随机梯度类算法的一个共同特点:都是利用单新息修正技术的单新息辨识方法(single innovation identification method),即对于标量系统

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t),$$

其中 $y(t) \in \mathbf{R}$ 为输出,$\boldsymbol{\varphi}(t) \in \mathbf{R}^n$ 为输入输出数据

构成的信息向量,$\boldsymbol{\theta} \in \mathbf{R}^n$ 为待辨识的参数向量,$v(t) \in \mathbf{R}$ 为零均值随机噪声. 估计上式参数向量 $\boldsymbol{\theta}$ 的最小二乘辨识算法或随机梯度等辨识算法有下列形式:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{L}(t)e(t),$$

其中 $\boldsymbol{L}(t) \in \mathbf{R}^n$ 为算法增益向量(gain vector),$e(t) := y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \in \mathbf{R}$ 为标量新息(scalar innovation),即单新息(single innovation).

这个算法可以这样描述: t 时刻的参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 是用增益向量 $\boldsymbol{L}(t)$ 与标量新息 $e(t)$ 的乘积,对 $t-1$ 时刻参数估计向量 $\hat{\boldsymbol{\theta}}(t-1)$ 进行修正,即 $\hat{\boldsymbol{\theta}}(t)$ 是在 $\hat{\boldsymbol{\theta}}(t-1)$ 的基础上加上增益向量 $\boldsymbol{L}(t)$ 与新息 $e(t)$ 的乘积. 这种方法也称为新息修正辨识方法或新息辨识方法.

上述算法中新息 $e(t)$ 是标量,我们把这个标量新息加以推广,就导出了多新息辨识方法(multi-innovation identification method)[24]. 多新息辨识理论(multi-innovation identification theory)就是将单新息修正技术加以推广,从新息修正角度提出多新息修正技术辨识的概念,建立多新息修正辨识方法,简称多新息辨识方法.

顾名思义,多新息算法就是将新息加以推广. 对标量系统而言,将算法中的标量新息 $e(t) \in \mathbf{R}$ 推广为新息向量 $\boldsymbol{E}(p,t) \in \mathbf{R}^p$,即多新息(multi-innovation),为使矩阵乘法维数兼容,增益向量 $\boldsymbol{L}(t) \in \mathbf{R}^n$ 须推广为增益矩阵(gain matrix) $\boldsymbol{\Gamma}(p,t) \in \mathbf{R}^{n\times p}$,那么多新息辨识算法可以写作

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

其中 $\boldsymbol{\Gamma}(p,t) \in \mathbf{R}^{n\times p}$ 为增益矩阵(gain matrix),$\boldsymbol{E}(p,t) \in \mathbf{R}^p$ 为新息向量(innovation vector),$p \geqslant 1$ 为新息长度(innovation length). 多新息辨识算法就是从这里命名的.

多新息辨识算法 t 时刻参数估计 $\hat{\boldsymbol{\theta}}(t)$ 是用增益矩阵 $\boldsymbol{\Gamma}(p,t)$ 与新息向量 $\boldsymbol{E}(p,t)$ 的乘积对 $t-1$ 时刻参数估计 $\hat{\boldsymbol{\theta}}(t-1)$ 进行修正的. 我们把基于多新息的辨识理论称为多新息辨识理论(multi-innovation identification theory),把基于多新息的辨识方法称为多新息辨识方法.

1.2 变递推间隔多新息辨识方法

传统递推辨识方法(如最小二乘算法)的一个典型特征是,参数估计每更新计算一次,即当 $t=1,2,3,\cdots$ 都计算一次参数估计,亦即递推间隔(recursive interval)为 1. 这类辨识算法有一个最大缺点: 当系统输入输出包含了不可信数据,即"坏数据"时,或某些数据采集不到,即存在损失数据(missing-data)时,算法无法跳过这些数据点,以避开坏数据对参数估计的影响. 为此,我们提出了变递推间隔的递推辨识算法,它与多新息辨识算法相结合,便导出本文要讨论的变递推间隔多新息辨识算法. 多新息辨识方法又可分为多新息最小二乘法、多新息随机梯度方法、多新息投影算法等.

定义可得到的数据点序列(正整数序列):

$$1 = t_0 < t_1 < t_2 < \cdots < t_{s-1} < t_s < \cdots, \qquad (1)$$

即所有 $y(t_s)$ 和 $\boldsymbol{\varphi}(t_s)$,$s=1,2,\cdots$ 都可量测得到. 变递推间隔多新息辨识算法(interval-varying multi-innovation identification algorithm)是通过新息向量,即多新息来进行修正的一种参数估计方法,后一时刻的参数估计 $\hat{\boldsymbol{\theta}}(t_s)$ 是在 $\hat{\boldsymbol{\theta}}(t_{s-1})$ 的基础上依靠增益矩阵 $\boldsymbol{\Gamma}(p,t_s) \in \mathbf{R}^{n\times p}$ 与新息向量(多新息) $\boldsymbol{E}(p,t_s) \in \mathbf{R}^p$ 的乘积来修正的,亦即

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \boldsymbol{\Gamma}(p,t_s)\boldsymbol{E}(p,t_s),$$

其中 t_s 为计算参数估计的时间点,$t_s^* := t_s - t_{s-1} \geqslant 1$,$s=1,2,3,\cdots$ 为递推间隔.

传统辨识方法是变递推间隔多新息辨识方法的一个特例. 例如,当 $t_s^* \equiv 1$ 和 $p=1$ 时,我们就得到传统辨识方法; 当 $p=1$ 时,就得到变递推间隔辨识方法; 当 $t_s^* \equiv 1$ 时,就得到多新息辨识方法.

由于变递推间隔多新息辨识方法引入了新息长度参量,采用了间断迭代、变递推间隔方式,使得其具有克服坏数据对参数估计的影响,可以提高参数估计精度,在处理损失数据系统(missing-data system)以及不规则采样系统(irregularly sampled-data systems)[10]的辨识问题上具有独到的特点.

1.3 多新息辨识的相关成果

多新息辨识已经成为一个崭新的辨识领域,多新息辨识理论也可以用于研究各种模型的辨识问题. 例如:

1) 多新息随机梯度辨识算法、多新息遗忘梯度辨识算法的性能分析[12];

2) 输出误差滑动平均模型的辅助模型多新息增广随机梯度辨识方法[29];

3) 基于辅助模型的多输入单输出系统多新息最小二乘辨识方法[23];

4) 多输入多输出系统多新息随机梯度辨识算法的一致收敛性[39];

5) 多率多输入系统多新息随机梯度辨识算法

及其收敛性[40];

6) 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[41];

7) 基于前向神经网络的多新息随机梯度辨识算法[42];

8) 随机系统多新息辨识在衰减激励条件下的性能分析[43];

9) 时变多变量系统多新息投影算法的均方收敛性[26];

10) 多新息随机梯度辨识方法[27];

11) 衰减激励条件下确定性系统多新息算法的收敛性分析[25];

12) 时变系统辨识的多新息方法[24];

13) 基于辅助模型的输出误差系统多新息随机梯度算法及其收敛性[44];

14) 基于辅助模型的多输入单输出系统多新息随机梯度辨识方法[45];

15) 非均匀采样数据系统的辅助模型多新息广义增广随机梯度算法[41,46];

16) 多变量输入非线性系统的辅助模型多新息随机梯度辨识算法[47];

17) 基于辅助模型的多率多输入系统多新息随机梯度参数估计[48];

18) 多新息随机梯度型辨识方法与多新息最小二乘类辨识方法[49].

2 多新息随机梯度辨识方法(*MISG*)

最小二乘辨识算法有快的收敛速度,但计算量大,因为需要计算协方差阵. 随机梯度算法的计算量小,但收敛速度慢. 为了改进随机梯度辨识方法的收敛速度,可引入新息长度,从而导出多新息随机梯度算法. 多新息随机梯度算法(MISG)就是在随机梯度算法(SG)与最小二乘算法(LS)的收敛速度和计算量之间找到折中. 本节利用多新息理论,通过引入新息长度,研究线性回归模型的多新息随机梯度算法,以及派生的多新息遗忘梯度算法、多新息投影算法、多新息广义投影算法等.

本节主要内容引自《Automatica》2007 年第 1 期上论文“Performance analysis of multi-innovation gradient type identification methods”[12].

考虑下列线性回归模型(linear regression model)的辨识问题:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{2}$$

其中 $y(t)\in\mathbf{R}$ 为输出,$\boldsymbol{\varphi}(t)\in\mathbf{R}^n$ 为输入输出数据构成的回归信息向量,$\boldsymbol{\theta}\in\mathbf{R}^n$ 为待辨识的参数向量,$v(t)\in\mathbf{R}$ 为零均值随机噪声.

辨识系统(2)参数向量 $\boldsymbol{\theta}$ 的随机梯度算法(SG)如下.

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}e(t), \tag{3}$$

$$e(t)=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{4}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1. \tag{5}$$

众所周知,随机梯度算法的收敛性慢,为了提高参数估计收敛速度,将标量新息 $e(t)\in\mathbf{R}$ 扩展为新息向量[12]:

$$\boldsymbol{E}(p,t)=\begin{bmatrix}e(t)\\ e(t-1)\\ \vdots\\ e(t-p+1)\end{bmatrix}\in\mathbf{R}^p,$$

其中正整数 p 表征新息长度,且

$$e(t-i)=y(t-i)-\boldsymbol{\varphi}^{\mathrm{T}}(t-i)\hat{\boldsymbol{\theta}}(t-i-1).$$

一般情况下,人们总是认为 $(t-1)$ 时刻的参数估计值 $\hat{\boldsymbol{\theta}}(t-1)$ 比之前 $t-i$ 时刻($i=2,3,4,\cdots,p-1$)的估计值 $\hat{\boldsymbol{\theta}}(t-i)$ 更接近真参数向量 $\boldsymbol{\theta}$. 因此,为简化,新息向量可以合理取为

$$\boldsymbol{E}(p,t)=\begin{bmatrix}y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\\ y(t-1)-\boldsymbol{\varphi}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}(t-1)\\ \vdots\\ y(t-p+1)-\boldsymbol{\varphi}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}(t-1)\end{bmatrix}\in\mathbf{R}^p. \tag{6}$$

定义信息矩阵 $\boldsymbol{\Phi}(p,t)$ 和堆积输出向量 $\boldsymbol{Y}(p,t)$ 为

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]\in\mathbf{R}^{n\times p},$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^p.$$

那么新息向量可以表述成

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1).$$

因为 $\boldsymbol{E}(1,t)=e(t)$,$\boldsymbol{\Phi}(1,t)=\boldsymbol{\varphi}(t)$,$\boldsymbol{Y}(1,t)=y(t)$,所以随机梯度算法(3)可以等价为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(1,t)}{r(t)}[\boldsymbol{Y}(1,t)-\boldsymbol{\Phi}^{\mathrm{T}}(1,t)\hat{\boldsymbol{\theta}}(t-1)].$$

这就是新息长度为 1 的“多”新息随机梯度算法. 把上式 $\boldsymbol{\Phi}(1,t)$ 和 $\boldsymbol{Y}(1,t)$ 中的“1”换为 p,就得到新息长度为 p 的多新息随机梯度算法:

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

线性回归模型(2)的多新息随机梯度算法(MISG,Multi-Innovation Stochastic Gradient algorithm)的基本方程为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{7}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{8}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Phi}(p,t)\|^2,\quad r(0)=1, \tag{9}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{10}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]. \tag{11}$$

由于这个算法里 $\boldsymbol{E}(p,t)\in\mathbf{R}^p$ 是一个新息向量,即多新息,又是从随机梯度算法派生出来的,这就是多新息随机梯度算法名称的来历. 当 $p=1$ 时,多新息随机梯度算法就退化为随机梯度算法.

为方便起见,设 t 为当前时刻,我们把 $y(t)$ 和 $\boldsymbol{\varphi}(t)$ 称为当前数据,$y(t-i)$ 和 $\boldsymbol{\varphi}(t-i)$ $(t=1,2,\cdots)$ 称为过去数据. 式(6)新息向量 $\boldsymbol{E}(p,t)$ 第1元为当前新息,其余为过去新息.

与随机梯度算法相比,多新息随机梯度算法有下列优点[12].

1) 在每步递推计算参数估计时,随机梯度算法(3)—(5)只使用了当前数据 $y(t)$ 和 $\boldsymbol{\varphi}(t)$,以及当前新息;而多新息随机梯度算法(7)—(11)不仅使用了当前数据和新息,而且使用了过去数据 $\{y(t-i),\boldsymbol{\varphi}(t-i):i=1,2,\cdots,p-1\}$ 和新息. 这是潜在改善算法收敛性的原因.

2) 多新息随机梯度辨识算法重复使用了系统数据:在时刻 t,MISG 算法使用的数据为 $\{y(t-i),\boldsymbol{\varphi}(t-i):i=0,1,\cdots,p-1\}$;而在时刻 $t+1$,MISG 算法使用的数据为 $\{y(t+1-i),\boldsymbol{\varphi}(t+1-i):i=0,1,\cdots,p-1\}$. 因此,在两次相邻时刻递推计算参数估计时,重复利用的数据为 $\{y(t-i),\boldsymbol{\varphi}(t-i):i=0,1,\cdots,p-2\}$. 这是多新息随机梯度算法改善参数估计精度原因.

3) 在相同数据长度下,增加 p 能减小参数估计误差. 换句话说,大 p 导致高精度的参数估计,因此,新息长度 p 的引入能改善参数估计精度. 然而,MISG 算法比 SG 算法计算量有所增加,这种增加的计算量是可容忍的,是计算机完全可以胜任的.

下面给出派生的几个典型多新息梯度型辨识算法.

1) 多新息随机梯度算法

为了进一步加快随机梯度算法的收敛速度,可将式(9)修改为 $r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2$. 修改后的算法也称为多新息随机梯度算法(MISG 算法)[12]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{12}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{13}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1, \tag{14}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{15}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]. \tag{16}$$

2) 多新息遗忘梯度算法

在式(9)或(14)中引入遗忘因子 λ,可得到多新息遗忘因子随机梯度算法,简称多新息遗忘梯度算法(MIFG,Multi-Innovation Forgetting Gradient algorithm)(MIFG 算法)[12]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{17}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{18}$$

$$r(t)=\lambda r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad 0<\lambda<1,\quad r(0)=1, \tag{19}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{20}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]. \tag{21}$$

3) 多新息投影算法

比较随机梯度算法与投影算法的差别,直接用 $\|\boldsymbol{\Phi}(p,t)\|^2$ 代替式(7)中 $r(t)$,或将式(9)修改为 $r(t)=\|\boldsymbol{\Phi}(p,t)\|^2$,就得到多新息投影算法(MI-Proj,Multi-Innovation Projection algorithm)[12]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{\|\boldsymbol{\Phi}(p,t)\|^2}\boldsymbol{E}(p,t), \tag{22}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{23}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{24}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]. \tag{25}$$

从下节可知,这是一个简化多新息投影算法.

4) 多新息广义投影算法

如果取 $r(t)=\sum_{i=0}^{q-1}\|\boldsymbol{\Phi}(p,t-i)\|^2$,或直接取 $r(t)=\|\boldsymbol{\Phi}(q,t)\|^2$($q$ 为记忆长度),就得到多新息广义投影算法(MIGP,Multi-Innovation Generalized Projection algorithm):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{26}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{27}$$

$$r(t)=\|\boldsymbol{\Phi}(q,t)\|^2,\quad q\geqslant p, \tag{28}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{29}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]. \tag{30}$$

当然,我们还可以得到带遗忘因子的广义投影算法等. 多新息辨识算法的初值选择类似其他递推算法,可取为 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$,$p_0=10^6$.

MISG 算法(12)—(16)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的步骤如下.

① 令 $t=1$: $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$, $r(0)=1$, $p_0=10^6$, $\mathbf{1}_n$ 是一个元均为 1 的 n 维列向量.

② 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(15)构造堆积输出向量 $\boldsymbol{Y}(p,t)$,由式(16)构造信息矩阵 $\boldsymbol{\Phi}(p,t)$.

③ 由式(13)计算新息向量 $\boldsymbol{E}(p,t)$,由式(14)计算 $r(t)$.

④ 根据式(12)刷新参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.

⑤ t 增 1,转到第 2 步.

MISG 算法计算参数估计 $\hat{\boldsymbol{\theta}}(t)$ 的流程如图 1 所示.

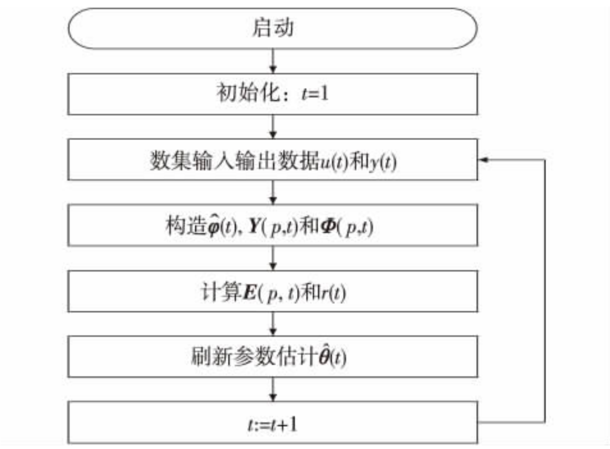

图 1 MISG 算法计算参数估计 $\hat{\boldsymbol{\theta}}(t)$ 的流程

Fig. 1 The flowchart of computing the parameter estimate $\hat{\boldsymbol{\theta}}(t)$ in the MISG algorithm

仿真试验如下.

例 1 考虑下列仿真对象[12]:

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

$$A(z)=1+a_1z^{-1}+a_2z^{-2}=1-1.35z^{-1}+0.75z^{-2},$$

$$B(z)=b_1z^{-1}+b_2z^{-2}=0.214z^{-1}+0.428z^{-2},$$

其中 $u(t)$ 和 $y(t)$ 分别是系统输入和输出,z^{-1} 为单位后移算子 $[z^{-1}y(t)=y(t-1)]$. 定义

$$\boldsymbol{\theta}:=[a_1,a_2,b_1,b_2]^{\mathrm{T}}=[-1.35,0.75,0.214,0.428]^{\mathrm{T}},$$

$$\boldsymbol{\varphi}(t):=[-y(t-1),-y(t-2),u(t-1),u(t-2)]^{\mathrm{T}},$$

那么这个系统可以写成式(2)的形式.

仿真时,输入 $\{u(t)\}$ 采用零均值、单位方差的不相关持续激励信号系列,$\{v(t)\}$ 采用零均值、方差为 $\sigma^2=0.50^2$ 的白噪声序列,系统的噪信比(noise-to-signal ratio)为 $\delta_{\mathrm{ns}}=82.17\%$. 应用随机梯度算法(3)—(5)(即在 MISG 算法中取 $p=1$)和多新息随机梯度算法(12)—(16)(新息长度 $p=2,3,5,8$)估计这个系统的参数,SG 参数估计和不同新息长度下的 MISG 参数估计及其误差如表 1—2,参数估计误差 $\delta=\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ 随 t 变化曲线如图 2 所示. 为与递推最小二乘算法(RLS)比较,图 2 画出了递推最小二乘算法(RLS)参数估计误差曲线(最下面一条曲线),RLS 参数估计及其误差如表 3 所示.

从表 1—3 和图 2,我们能看出: MISG 估计($p\geqslant 2$)比 SG 估计有更高的精度; 随着新息长度 p 增加,MISG 算法给出的参数估计误差越来越小,并随着数据长度 t 增加而趋于零. 随着新息长度增加,MISG 估计越来越接近 RLS 估计.

Matlab 程序如下.

把下列程序写到 MISG_SG_RLS_ex1.m 文件中,依次运行 $p=1,2,3,5$ 和 $p=8$,可得到上述例子的仿真结果(参数估计表和误差曲线图).

```
%————————————————————————————————————————————————*
% Filename: MISG_SG_RLS_ex1.m for the MISG,SG,RLS algorithms *
% The ARX models: A(z)y(t) = B(z)u(t) + v(t) *
% The noise variance sigma^2 = 0.50^2 *
% The MISG algorithm with the innovation length p = 1,2,3, 5 and 8 *
%————————————————————————————————————————————————*
clear; clf; format short g
M = 'The SG or MISG algorithm for ARX models'
sigma = 0.5; % The noise variance sigma^2 = 0.50^2
 FF = 1; % The forgetting factor \lambda
 p = 1; % The innovation length p = 1,2,3,5 and 8

 PlotLength = 3000; length1 = PlotLength + 100;
 na = 2; nb = 2; n = na + nb;
 a = [1, -1.35, 0.75]; b = [0, 0.214, 0.428]; d = 1;
 par0 = [a(2:na+1), b(2:nb+1)]';
 p0 = 1e6; P = eye(n)* p0; r = 1;
 par1 = ones(n,1)/p0; parLS = par1;
 %——Compute the noise-to-signal ratio
 sy = f_integral(a,b); sv = f_integral(a,d);
 delta_ns = sqrt(sv/sy)* 100* sigma;
 [sy, sv, delta_ns]
 %——Generate the input-output data
 rand('state',1); randn('state',0);
 %u = (rand(length1,1) - 0.5)* sqrt(12); %v = randn(length1,1);
 u = idinput(length1,'rgs', [0,1], [-1,1]);
 v = idinput(length1,'rgs', [0,1], [-1,1])* sigma;
 y = ones(n,1)/p0;
 for t = n+1: length1
```

```
    y(t) = par0'* [-y(t-1:-1:t-na); u(t-1:-1:t-nb)] + v(t);
end
% —Identification
jj = 0; j1 = 0;
for t = 10* n: length1,
    jj = jj + 1;
    varphi = [-y(t-1:-1:t-na); u(t-1:-1:t-nb)];
    Phi = varphi; Y = y(t:-1:t-p+1);
    for i = 1:p-1
       Phi = [Phi, [-y(t-1-i:-1:t-na-i); u(t-1-i:-1:t-nb-i)]];
    end
% ——Compute the SG or MISG parameter estimates
    r = FF* r + varphi'* varphi;
    par1 = par1 + Phi* (Y - Phi'* par1) /r;
    delta = norm(par1 - par0) /norm(par0); % The SG or MISG errors
% ——Compute the RLS parameter estimates
    L1 = P* varphi/(1 + varphi'* P* varphi);
    P = P - L1* (varphi'* P);
    parLS = parLS + L1* (y(t) - varphi'* parLS);
    deltaLS = norm(parLS - par0) /norm(par0); % The RLS error

    ls(jj,:) = [jj,par1',delta];
    lsLS(jj,:) = [jj,parLS',deltaLS];
    if (jj = = 100) | (jj = = 200) | (jj = = 500) | (mod(jj,1000) = = 0)
       j1 = j1 + 1;
       ls_100(j1,:) = [jj,par1',delta* 100];
       lsLS_100(j1,:) = [jj,parLS',deltaLS* 100];
    end
    if jj = = PlotLength
       break;
    end
end
ls_100(j1+1,:) = [0,par0',0];
lsLS_100(j1+1,:) = [0,par0',0];

fprintf('\n $ \\lambda = %5.2f $ ,( $ \\sigma_^2 = %5.2f^2 $ ,$ \\delta_{\\ns} = %6.2f% s',...
    FF,sigma,delta_ns,'\%$ )');
fprintf('\n $ t $ $ a_1 $ $ a_2 $ $ b_1 $ $ b_2 $ %s',...
    '$ \delta\ (\%) \ \ $ \\\hline');
fprintf('\nThe innovation length $ p = %d $ ,$ \\lambda = %6.2f $ ',p,FF);
fprintf('\n%5d %10.5f %10.5f %10.5f %10.5f %10.5f \\\\',ls_100');

fprintf('\n The RLS estimates');
fprintf('\n%5d %10.5f %10.5f %10.5f %10.5f %10.5f \\\\',lsLS_100');

figure(1); k0 = 12;
jk = (k0:1:PlotLength-1)';
plot(ls(jk,1),ls(jk,n+2),'m',lsLS(jk,1),lsLS(jk,n+2),'b')
xlabel('\it t'); ylabel('{\it\delta}')

if p = = 1 % Collect for plotting the error curves
    data1 = [ls(:,1),ls(:,n+2)];
    save data1 data1
elseif p = = 2
    load data1
    data2 = [data1,ls(:,n+2)];
    save data2 data2
elseif p = = 3
    load data2
    data3 = [data2,ls(:,n+2)];
    save data3 data3
elseif p = = 5
    load data3
    data4 = [data3,ls(:,n+2)];
    save data4 data4
elseif p = = 8
    load data4
    z0 = [data4,ls(:,n+2)];
    jk = (k0:10:PlotLength-1)';
    jkx = z0(jk,1); figure(2);
    plot(jkx,z0(jk,2),'b',jkx,z0(jk,3),'k',jkx,z0(jk,4),'b',...
    jkx,z0(jk,5),'k',jkx,z0(jk,6),'b',jkx,lsLS(jk,n+2),'k')
    axis([0,PlotLength,0,0.63])
    text(1000,z0(1000,2) + 0.02,'{SG (MISG,\itp} = 1)')
    text(1000,z0(1000,3) + 0.02,'{MISG,\itp} = 2')
    text(1000,z0(1000,4) + 0.02,'{MISG,\itp} = 3')
    text(1000,z0(1000,5) + 0.02,'{MISG,\itp} = 5')
    text(1000,z0(1000,6) + 0.02,'{MISG,\itp} = 8')
    text(10,0.015,'{RLS}')
 end
 xlabel('\it t'); ylabel('{\it \delta}')
```

表 1　例 1 参数的 SG 估计及其误差($\sigma^2=0.50^2$)

Table 1　The SG estimates and errors of Example 1 ($\sigma^2=0.50^2$)

t	a_1	a_2	b_1	b_2	$\delta/\%$
100	−0.764 72	0.161 51	0.241 23	0.405 23	51.382 41
200	−0.807 79	0.215 29	0.236 77	0.418 75	47.124 95
500	−0.840 15	0.263 87	0.230 90	0.428 06	43.584 17
1 000	−0.871 17	0.287 36	0.232 24	0.436 10	41.200 02
2 000	−0.899 37	0.316 52	0.233 88	0.446 22	38.710 31
3 000	−0.920 76	0.332 38	0.233 51	0.450 96	37.088 14
真值	−1.350 00	0.750 00	0.214 00	0.428 00	

表 2　例 1 参数的 MISG 估计及其误差($\sigma^2=0.50^2$)

Table 2　The MISG estimates and errors of Example 1($\sigma^2=0.50^2$)

p	t	a_1	a_2	b_1	b_2	$\delta/\%$
2	100	−0.935 49	0.315 84	0.245 71	0.524 37	37.653 67
	200	−0.987 00	0.394 83	0.240 08	0.525 30	32.023 70
	500	−1.023 05	0.457 32	0.230 33	0.515 89	27.698 62
	1 000	−1.062 38	0.479 97	0.231 08	0.512 14	24.972 33
	2 000	−1.090 87	0.511 27	0.231 48	0.510 85	22.412 88
	3 000	−1.116 73	0.527 70	0.229 83	0.510 78	20.600 53
3	100	−1.020 81	0.401 74	0.454 30	0.575 02	34.382 30
	200	−1.083 86	0.492 41	0.405 48	0.556 21	26.980 62
	500	−1.118 02	0.559 79	0.349 42	0.528 66	21.288 62
	1 000	−1.159 47	0.575 21	0.323 77	0.516 04	18.207 22
	2 000	−1.182 60	0.602 57	0.301 58	0.505 97	15.586 72
	3 000	−1.208 90	0.617 71	0.289 90	0.503 33	13.669 72
5	100	−1.103 87	0.496 17	0.452 60	0.536 75	27.226 14
	200	−1.177 00	0.603 29	0.384 09	0.508 68	18.232 38
	500	−1.212 96	0.660 78	0.311 10	0.478 44	12.169 75
	1 000	−1.257 24	0.662 72	0.281 58	0.467 72	9.250 15
	2 000	−1.269 23	0.683 09	0.261 80	0.461 50	7.424 32
	3 000	−1.293 84	0.697 09	0.250 68	0.462 43	5.697 01
8	100	−1.171 87	0.564 54	0.368 26	0.461 07	18.659 96
	200	−1.243 07	0.688 07	0.316 09	0.445 57	9.973 42
	500	−1.277 80	0.720 21	0.256 64	0.433 97	5.516 40
	1 000	−1.316 36	0.711 71	0.238 14	0.432 80	3.500 99
	2 000	−1.315 18	0.721 87	0.232 12	0.435 92	3.026 75
	3 000	−1.336 56	0.737 53	0.224 01	0.442 22	1.562 84
真值		−1.350 00	0.750 00	0.214 00	0.428 00	

表 3　例 1 参数的 RLS 估计及其误差($\sigma^2=0.50^2$)

Table 3　The RLS estimates and errors of Example 1($\sigma^2=0.50^2$)

t	a_1	a_2	b_1	b_2	$\delta/\%$
100	−1.338 64	0.699 60	0.235 81	0.373 16	4.851 74
200	−1.349 79	0.746 26	0.255 88	0.390 62	3.479 77
500	−1.329 66	0.752 81	0.226 35	0.403 22	2.131 94
1 000	−1.336 02	0.748 65	0.221 12	0.413 58	1.320 70
2 000	−1.326 07	0.741 30	0.223 13	0.425 84	1.678 29
3 000	−1.349 12	0.757 48	0.216 20	0.431 59	0.533 41
真值	−1.350 00	0.750 00	0.214 00	0.428 00	

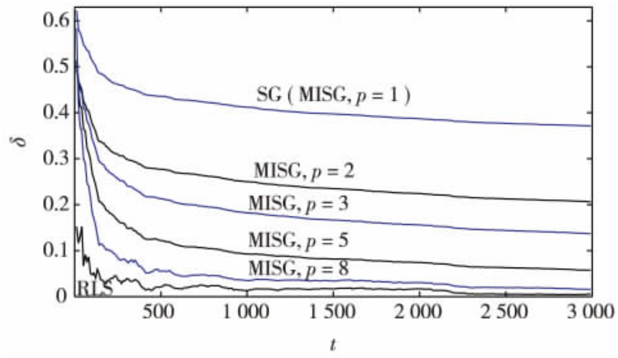

图2 例1参数估计误差 δ 随 t 变化曲线($\sigma^2=0.50^2$)

ig. 2 The parameter estimation errors δ versus t ($\sigma^2=0.50^2$)

下面利用最速下降法或梯度搜索原理推导多新息辨识方法.

1) 多新息投影算法

在多新息随机梯度算法中,我们考虑了从 $t-p+1$ 到 t 的长度为 p 的数据窗里共 p 组数据,定义了堆积输出向量(stacked output vector) $\boldsymbol{Y}(p,t)$ 和堆积信息矩阵(stacked information matrix) $\boldsymbol{\Phi}(p,t)$ 如下:

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

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

如果再定义堆积噪声向量(stacked noise vector):

$$\boldsymbol{V}(p,t):=\begin{bmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(t-p+1) \end{bmatrix}\in\mathbf{R}^p,$$

则由式(2)可得到矩阵方程:

$$\boldsymbol{Y}(p,t)=\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\theta}+\boldsymbol{V}(p,t). \tag{31}$$

这就是多新息方法的辨识模型(identification model).

设准则函数为

$$J(\boldsymbol{\theta}):=\|\boldsymbol{V}(p,t)\|^2=\|\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\theta}\|^2. \tag{32}$$

这是一个有限数据窗准则函数. 假设步长为 μ_t,求 $J(\boldsymbol{\theta})$ 最小值的梯度迭代算法可表示为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)-\frac{\mu_t}{2}\mathrm{grad}[J(\hat{\boldsymbol{\theta}}(t-1))]=$$

$$\hat{\boldsymbol{\theta}}(t-1)+\mu_t\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)]. \tag{33}$$

定义新息向量:

$$\boldsymbol{E}(p,t):=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}^p,$$

则有

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\mu_t\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t). \tag{34}$$

下面求最佳步长 μ_t. 将 $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)$ 代入式(32)可得

$$\begin{aligned} g(\mu_t)&:=J(\hat{\boldsymbol{\theta}}(t))= \\ &\|\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)[\hat{\boldsymbol{\theta}}(t-1)+\mu_t\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)]\|^2= \\ &\|[\boldsymbol{I}-\mu_t\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)]\boldsymbol{E}(p,t)\|^2= \\ &\boldsymbol{E}^{\mathrm{T}}(p,t)[\boldsymbol{I}-\mu_t\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)]^2\boldsymbol{E}(p,t)= \\ &\boldsymbol{E}^{\mathrm{T}}(p,t)[\boldsymbol{I}-2\mu_t\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)+ \\ &\mu_t^2\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)]\boldsymbol{E}(p,t)= \\ &\|\boldsymbol{E}(p,t)\|^2-2\mu_t\|\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2+ \\ &\mu_t^2\|\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2. \end{aligned}$$

极小化 $g(\mu_t)$,令 $g'(\mu_t)=0$,即

$$-2\|\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2+2\mu_t\|\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2=0.$$

由此可求得最佳步长为

$$\mu_t=\frac{\|\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2}{\|\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2}=$$

$$\frac{\boldsymbol{E}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)}{\boldsymbol{E}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)}. \tag{35}$$

将式(35)代入(34)即得多新息投影辨识算法(MI-Proj, Multi-Innovation Projection identification algorithm)[14]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\mu_t\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t), \tag{36}$$

$$\mu_t=\frac{\|\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2}{\|\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\Phi}(p,t)\boldsymbol{E}(p,t)\|^2}, \tag{37}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{38}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{39}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{40}$$

如果式(37)的分母为零,就令 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)$. 因为收敛因子 μ_t 的计算比较复杂,故对其进行简化. 由于对于任意实向量 x 和非负定对称矩阵 Q,下式成立.

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}\leqslant\lambda_{\max}[\boldsymbol{Q}]\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\leqslant\|\boldsymbol{Q}\|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}.$$

于是,收敛因子可以保守取为

$$\mu_t=\frac{1}{\lambda_{\max}[\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)]}.$$

因为计算矩阵的迹(trace)比计算特征值简单,所以收敛因子可以更保守取为

$$\mu_t=\frac{1}{\|\boldsymbol{\Phi}(p,t)\|^2}.$$

如果取上式μ_t作为收敛因子,就得到简化多新息投影算法(22)—(25).

由式(36)与(38)可得

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t) &= \\ \hat{\boldsymbol{\theta}}(t-1) &+ \mu_t \boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)] = \\ [\boldsymbol{I} &- \mu_t \boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)]\hat{\boldsymbol{\theta}}(t-1) + \mu_t \boldsymbol{\Phi}(p,t)\boldsymbol{Y}(p,t).\end{aligned}$$

如果$\mu_t \boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)$的特征值大于2,$\hat{\boldsymbol{\theta}}(t)$就不可能收敛.

2) 变递推间隔多新息投影算法

在实际问题中,可能发生数据丢失的情况. 也就是说,对每一个t,$y(t)$和$\varphi(t)$不可能总是可得到. 为了处理数据丢失情况,定义一个整数序列$\{t_s, s=0,1,2,\cdots\}$满足

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

且$t_s^* := t_s - t_{s-1} \geqslant 1$. 假设当$t = t_s(s=1,2,\cdots)$时,$y(t)$和$\varphi(t)$都可得到,即对任意$s=1,2,3,\cdots$,$y(t_s)$和$\varphi(t_s)$都可得到. 用$t_s$代替式(31)中$t$,可得一类损失数据系统(missing-data system)的辨识模型(identification model):

$$\boldsymbol{Y}(p,t_s) = \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{\theta} + \boldsymbol{V}(p,t_s). \quad (41)$$

定义准则函数(criterion function):

$$J(\boldsymbol{\theta}) = \| \boldsymbol{Y}(p,t_s) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{\theta} \|^2. \quad (42)$$

设收敛因子或迭代步长为μ_{t_s},采用与上类似的推导方法,可得变递推间隔多新息投影辨识算法(V-MI-Proj, interval-Varying Multi-Innovation Projection identification algorithm)[13-14]:

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \mu_{t_s}\boldsymbol{\Phi}(p,t_s)\boldsymbol{E}(p,t_s), \quad s=0,1,2,\cdots, \quad (43)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_s), \quad t \in T_s := \{t_s, t_s+1, \cdots, t_{s+1}-1\}, \quad (44)$$

$$\mu_{t_s} = \frac{\| \boldsymbol{\Phi}(p,t_s)\boldsymbol{E}(p,t_s) \|^2}{\| \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{\Phi}(p,t_s)\boldsymbol{E}(p,t_s) \|^2}, \quad (45)$$

$$\boldsymbol{E}(p,t_s) = \boldsymbol{Y}(p,t_s) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1}), \quad (46)$$

$$\boldsymbol{\Phi}(p,t_s) = [\boldsymbol{\varphi}(t_s), \boldsymbol{\varphi}(t_s-1), \cdots, \boldsymbol{\varphi}(t_s-p+1)], \quad (47)$$

$$\boldsymbol{Y}(p,t_s) = [y(t_s), y(t_s-1), \cdots, y(t_s-p+1)]^{\mathrm{T}}, \quad (48)$$

$$0 = t_0 < t_1 < t_2 < \cdots, \quad 1 \leqslant t_s^* := t_s - t_{s-1}. \quad (49)$$

变递推间隔多新息辨识算法也可以这样理解:实际中一些系统的控制周期与参数估计刷新周期不相同. 在这种情形下,t_s^*可认为是控制周期,但是在这一个周期t_s^*内,每t_s^*/N(N为一正整数)都可获得一组可用的数据信息,可利用其中p组数据所包含的信息进行参数辨识,即用这些信息产生的新息向量$\boldsymbol{E}(p,t_s)$来对$\hat{\boldsymbol{\theta}}(t_{s-1})$进行修正,多新息(修正)算法就是因此而得名. 这里$\boldsymbol{E}(p,t_s)$的每一元都是新息,与新息的定义相吻合,易于理解. 当$t_s^* = p = 1$时,多新息算法就是常规的单新息修正算法(投影算法). 式(44)表示在数据丢失的区间,保持参数估计不变.

同样,我们有简化的变递推间隔多新息投影辨识算法(V-MI-Proj):

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\boldsymbol{\Phi}(p,t_s)}{\| \boldsymbol{\Phi}(p,t_s) \|^2}\boldsymbol{E}(p,t_s), \quad s=0,1,2,\cdots, \quad (50)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_s), \quad t \in T_s = \{t_s, t_s+1, \cdots, t_{s+1}-1\}, \quad (51)$$

$$\boldsymbol{E}(p,t_s) = \boldsymbol{Y}(p,t_s) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1}), \quad (52)$$

$$\boldsymbol{\Phi}(p,t_s) = [\boldsymbol{\varphi}(t_s), \boldsymbol{\varphi}(t_s-1), \cdots, \boldsymbol{\varphi}(t_s-p+1)], \quad (53)$$

$$\boldsymbol{Y}(p,t_s) = [y(t_s), y(t_s-1), \cdots, y(t_s-p+1)]^{\mathrm{T}}, \quad (54)$$

$$0 = t_0 < t_1 < t_2 < \cdots, \quad 1 \leqslant t_s^* = t_s - t_{s-1}. \quad (55)$$

为了防止式(50)右边第2项分母为零,解决的方法之一是:当$\| \boldsymbol{\Phi}(p,t_s) \|^2 = 0$时,令$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1})$,或者在式(50)右边第2项分母上加上一个正常数,将式(50)修改为

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\boldsymbol{\Phi}(p,t_s)}{1 + \| \boldsymbol{\Phi}(p,t_s) \|^2}[\boldsymbol{Y}(p,t_s) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})].$$

3) 变递推间隔多新息广义投影算法

变递推间隔多新息投影辨识算法(50)—(55)可以推广为变递推间隔多新息广义投影辨识算法(V-MIGP, interval-Varying Multi-Innovation Generalized Projection identification algorithm)[11,24]:

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\boldsymbol{\Phi}(p,t_s)}{r(q,t_s)}[\boldsymbol{Y}(p,t_s) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})], \quad s=0,1,2,\cdots \quad (56)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_s), \quad t \in T_s = \{t_s, t_s+1, \cdots, t_{s+1}-1\}, \quad (57)$$

$$r(q,t_s) = \mathrm{tr}[\boldsymbol{\Phi}(q,t_s)\boldsymbol{\Phi}^{\mathrm{T}}(q,t_s)], \quad q \geqslant p, \quad (58)$$

$$\boldsymbol{\Phi}(p,t_s) = [\boldsymbol{\varphi}(t_s), \boldsymbol{\varphi}(t_s-1), \cdots, \boldsymbol{\varphi}(t_s-p+1)], \quad (59)$$

$$\boldsymbol{Y}(p,t_s) = [y(t_s), y(t_s-1), \cdots, y(t_s-p+1)]^{\mathrm{T}}, \quad (60)$$

$$0 = t_0 < t_1 < t_2 < \cdots, \quad 1 \leqslant t_s^* = t_s - t_{s-1}. \quad (61)$$

4) 变递推间隔多新息随机梯度算法

进一步可推广为变递推间隔多新息随机梯度算法(V-MISG, interval-Varying Multi-Innovation Stochastic Gradient identification algorithm)(V-MISG 算

法)[11,24]:

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\Phi}(p,t_s)}{r(t_s)}[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],\quad s=0,1,2,\cdots \tag{62}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\}, \tag{63}$$

$$r(t_s)=\|\boldsymbol{\Phi}(s,t_s)\|^2=\sum_{i=0}^{s-1}\|\boldsymbol{\varphi}(t_s-i)\|^2, \tag{64}$$

$$\boldsymbol{\Phi}(p,t_s)=[\boldsymbol{\varphi}(t_s),\boldsymbol{\varphi}(t_s-1),\cdots,\boldsymbol{\varphi}(t_s-p+1)], \tag{65}$$

$$\boldsymbol{Y}(p,t_s)=[y(t_s),y(t_s-1),\cdots,y(t_s-p+1)]^{\mathrm{T}}, \tag{66}$$

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*=t_s-t_{s-1}. \tag{67}$$

3 多新息梯度型辨识算法(*MIGT*)

多新息辨识算法派生出如下一些算法.

1) 多新息广义投影算法

当递推间隔 $t_s^*\equiv 1$ 时,从 V-MIGP 算法(56)—(61)得到多新息广义投影辨识算法(MIGP,Multi-Innovation Generalized Projection identification algorithm):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(q,t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{68}$$

$$r(q,t)=\mathrm{tr}[\boldsymbol{\Phi}(q,t)\boldsymbol{\Phi}^{\mathrm{T}}(q,t)],\quad q\geqslant p, \tag{69}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{70}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{71}$$

当记忆长度 $q=p$ 时,MIGP 算法退化为多新息投影辨识算法(MI-Proj,Multi-Innovation Projection identification algorithm).

2) 多新息随机梯度算法

当递推间隔 $t_s^*\equiv 1$ 时,从 V-MISG 算法(62)—(67)得到多新息随机梯度辨识算法(MISG,Multi-Innovation Stochastic Gradient identification algorithm)(MISG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{72}$$

$$r(t)=\sum_{i=0}^{t-1}\|\boldsymbol{\varphi}(t-i)\|^2=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1, \tag{73}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{74}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{75}$$

3) 变递推间隔广义投影算法

当新息长度 $p=1$ 时,从 V-MIGP 算法(56)—(61)得到变递推间隔广义投影辨识算法(V-GP,interval-Varying Generalized Projection identification algorithm):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\varphi}(t_s)}{r(q,t_s)}[y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})], \tag{76}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\}, \tag{77}$$

$$r(q,t_s)=\sum_{i=0}^{q-1}\|\boldsymbol{\varphi}(t_s-i)\|^2,\quad q\geqslant 1, \tag{78}$$

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*=t_s-t_{s-1}. \tag{79}$$

当记忆长度 $q=p=1$ 时,V-GP 算法退化为变递推间隔投影辨识算法(V-Proj: interval-Varying Projection identification algorithm).

4) 变递推间隔随机梯度算法

当新息长度 $p=1$ 时,从 V-MISG 算法(62)—(67)得到变递推间隔随机梯度辨识算法(V-SG,interval-Varying Stochastic Gradient identification algorithm)(V-SG 算法):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\varphi}(t_s)}{r(t_s)}[y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})], \tag{80}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\}, \tag{81}$$

$$r(t_s)=\sum_{i=0}^{s-1}\|\boldsymbol{\varphi}(t_s-i)\|^2,\quad q\geqslant 1, \tag{82}$$

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*=t_s-t_{s-1}. \tag{83}$$

5) 等递推间隔广义投影算法

当递推间隔 $t_s^*\equiv d$,$p=1$ 时,从 V-MIGP 算法(56)—(61)得到等递推间隔广义投影辨识算法(E-GP,interval-Equating Generalized Projection identification algorithm):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\frac{\boldsymbol{\varphi}(t)}{r(q,t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-d)], \tag{84}$$

$$r(q,t)=\sum_{i=0}^{q-1}\|\boldsymbol{\varphi}(t-i)\|^2,\quad q\geqslant 1. \tag{85}$$

或

$$\hat{\boldsymbol{\theta}}(id)=\hat{\boldsymbol{\theta}}((i-1)d)+\frac{\boldsymbol{\varphi}(id)}{r(q,id)}[y(id)-\boldsymbol{\varphi}^{\mathrm{T}}(id)\hat{\boldsymbol{\theta}}((i-1)d)], \tag{86}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}((i-1)d),\quad t\in T_i=\{(i-1)d,(i-1)d+1,\cdots,id-1\}, \tag{87}$$

$$r(q,id)=\sum_{j=0}^{q-1}\|\boldsymbol{\varphi}(id-j)\|^2,\quad q\geqslant 1. \tag{88}$$

当记忆长度 $q=p=1$ 时,E-GP 算法退化为等递推间隔投影辨识算法(E-Proj,interval-Equating Projection identification algorithm).

6) 等递推间隔随机梯度算法

当递推间隔 $t_s^* \equiv d, p=1$ 时,从 V-MISG 算法(62)—(67)得到等递推间隔随机梯度辨识算法(E-SG, interval-Equating Stochastic Gradient identification algorithm)(E-SG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-d)], \tag{89}$$

$$r(t)=\sum_{i=0}^{t-1}\|\boldsymbol{\varphi}(t-i)\|^2=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1. \tag{90}$$

7) 等递推间隔多新息广义投影算法

当递推间隔 $t_s^* \equiv d\ (s=1,2,\cdots)$ 时,从 V-MIGP 算法(56)—(61)得到等递推间隔多新息广义投影辨识算法(E-MIGP, interval-Equating Multi-Innovation Generalized Projection identification algorithm):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\frac{\boldsymbol{\Phi}(p,t)}{r(q,t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-d)], \tag{91}$$

$$r(q,t)=\mathrm{tr}[\boldsymbol{\Phi}(q,t)\boldsymbol{\Phi}^{\mathrm{T}}(q,t)]=\sum_{i=0}^{q-1}\|\boldsymbol{\varphi}(t-i)\|^2,\quad q\geqslant p, \tag{92}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{93}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{94}$$

或

$$\hat{\boldsymbol{\theta}}(id)=\hat{\boldsymbol{\theta}}((i-1)d)+\frac{\boldsymbol{\Phi}(p,id)}{r(q,id)}[\boldsymbol{Y}(p,id)-\boldsymbol{\Phi}^{\mathrm{T}}(p,id)\hat{\boldsymbol{\theta}}((i-1)d)], \tag{95}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}((i-1)d),\quad t\in T_i=\{(i-1)d,(i-1)d+1,\cdots,id-1\}, \tag{96}$$

$$\boldsymbol{\Phi}(p,id)=[\boldsymbol{\varphi}(id),\boldsymbol{\varphi}(id-1),\cdots,\boldsymbol{\varphi}(id-p+1)], \tag{97}$$

$$r(q,id)=\mathrm{tr}[\boldsymbol{\Phi}(q,id)\boldsymbol{\Phi}^{\mathrm{T}}(q,id)],\quad q\geqslant p, \tag{98}$$

$$\boldsymbol{Y}(p,id)=[y(id),y(id-1),\cdots,y(id-p+1)]^{\mathrm{T}}. \tag{99}$$

当记忆长度 $q=p$ 时,E-MIGP 算法退化为等递推间隔多新息投影辨识算法(E-MI-Proj, interval-Equating Multi-Innovation Projection identification algorithm).

8) 等递推间隔多新息随机梯度算法

当递推间隔 $t_s^* \equiv d$ 时,从 V-MISG 算法(62)—(67)得到等递推间隔多新息随机梯度辨识算法(E-MISG, interval-Equating Multi-Innovation Stochastic Gradient identification algorithm)(E-MISG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-d)], \tag{100}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1, \tag{101}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{102}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{103}$$

9) 多新息遗忘梯度算法

在 MISG 算法中引入遗忘因子 λ,得到多新息遗忘因子随机梯度辨识算法(MI-FFSG, Multi-Innovation Forgetting Factor Stochastic Gradient identification algorithm)(MI-FFSG 算法)简称为多新息遗忘梯度辨识算法(MIFG, Multi-Innovation Forgetting Gradient identification algorithm)(MIFG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{104}$$

$$r(t)=\lambda r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad 0<\lambda\leqslant 1, r(0)=1, \tag{105}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{106}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{107}$$

10) 等递推间隔多新息遗忘梯度算法

在 E-MISG 算法中引入遗忘因子 λ,得到等递推间隔多新息遗忘梯度辨识算法(E-MIFG, interval-Equating Multi-Innovation Forgetting factor stochastic Gradient identification algorithm)(E-MIFG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\frac{\boldsymbol{\Phi}(p,t)}{r(t)}[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-d)], \tag{108}$$

$$r(t)=\lambda r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r(0)=1, \tag{109}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)], \tag{110}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \tag{111}$$

此外,当递推间隔 $t_s^*=1$,新息长度 $p=1$ 时,V-MIGP 算法退化为广义投影辨识算法(GP),V-MISG 算法退化为随机梯度辨识算法(SG).

4 多新息最小二乘辨识方法(*MILS*)

考虑下列线性回归模型描述的标量系统

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{112}$$

其中 $y(t)\in\mathbf{R}$ 为系统输出,$\boldsymbol{\theta}\in\mathbf{R}^n$ 为待辨识的参数向量,$\boldsymbol{\varphi}(t)\in\mathbf{R}^n$ 是由系统输入 $u(t)\in\mathbf{R}$ 和输出 $y(t)$ 构成的回归信息向量,$v(t)\in\mathbf{R}$ 为零均值随机噪声.

辨识的目标是:利用系统的输入输出数据

$\{u(i),y(i),0\leqslant i\leqslant t\}$或$\{y(i),\varphi(i),0\leqslant i\leqslant t\}$,提出多新息最小二乘辨识算法,对系统的未知参数向量$\boldsymbol{\theta}$进行实时估计.

考虑$t-p+1$到t时共p组数据,令

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

$$\boldsymbol{V}(p,t):=\begin{bmatrix} v(t)\\ v(t-1)\\ \vdots \\ v(t-p+1)\end{bmatrix}\in\mathbf{R}^p,$$

$$\boldsymbol{\Phi}(p,t):=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]\in\mathbf{R}^{n\times p}.$$

则由式(112)可得到矩阵方程:

$$\boldsymbol{Y}(p,t)=\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\theta}+\boldsymbol{V}(p,t).$$

上式称为多新息辨识方法的辨识模型或辨识表达式. 取准则函数为

$$J(\boldsymbol{\theta}):=\sum_{i=1}^{t}[\boldsymbol{Y}(p,i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,i)\boldsymbol{\theta}]^{\mathrm{T}}[\boldsymbol{Y}(p,i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,i)\boldsymbol{\theta}]=\sum_{i=1}^{t}\|\boldsymbol{Y}(p,i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,i)\boldsymbol{\theta}\|^2.$$

如不作特别申明,矩阵X的范数均定义为$\|X\|^2:=\mathrm{tr}[XX^{\mathrm{T}}]$. 定义矩阵

$$\boldsymbol{Z}_t:=\begin{bmatrix}\boldsymbol{Y}(p,1)\\ \boldsymbol{Y}(p,2)\\ \vdots\\ \boldsymbol{Y}(p,t)\end{bmatrix},\quad \boldsymbol{H}_t:=\begin{bmatrix}\boldsymbol{\Phi}^{\mathrm{T}}(p,1)\\ \boldsymbol{\Phi}^{\mathrm{T}}(p,2)\\ \vdots\\ \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\end{bmatrix}.$$

则准则函数可以写为

$$J(\boldsymbol{\theta})=(\boldsymbol{Z}_t-\boldsymbol{H}_t\boldsymbol{\theta})^{\mathrm{T}}(\boldsymbol{Z}_t-\boldsymbol{H}_t\boldsymbol{\theta}).$$

设$\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}$使$J(\boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}}=\min$. 令

$$\left.\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}}=\left.\frac{\partial[(\boldsymbol{Z}_t-\boldsymbol{H}_t\boldsymbol{\theta})^{\mathrm{T}}(\boldsymbol{Z}_t-\boldsymbol{H}_t\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}}=\mathbf{0}.$$

展开之,并运用如下2个向量微分公式:

$$\frac{\partial}{\partial\boldsymbol{x}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})=\boldsymbol{a},\quad \frac{\partial}{\partial\boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x},$$

可得

$$(\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t)\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}=\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Z}_t,$$

或

$$\hat{\boldsymbol{\theta}}_{\mathrm{MILS}}=(\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t)^{-1}\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Z}_t=\left[\sum_{i=1}^{t}\boldsymbol{\Phi}(p,i)\boldsymbol{\Phi}^{\mathrm{T}}(p,i)\right]^{-1}\left[\sum_{i=1}^{t}\boldsymbol{\Phi}(p,i)\boldsymbol{Y}^{\mathrm{T}}(p,i)\right].$$

这就是多新息辨识一次完成最小二乘估计. 令

$$\boldsymbol{P}^{-1}(t)=\sum_{i=1}^{t}\boldsymbol{\Phi}(p,i)\boldsymbol{\Phi}^{\mathrm{T}}(p,i).$$

则有

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t).$$

仿照递推最小二乘辨识方法的推导过程,能够得到多新息最小二乘辨识算法(MILS,Multi-Innovation Least Squares identification algorithm)(MILS算法)如下[13]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{113}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{114}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{115}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}.\tag{116}$$

或

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{117}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)=\boldsymbol{P}(t-1)\boldsymbol{\Phi}(p,t)[\boldsymbol{I}_p+\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\boldsymbol{\Phi}(p,t)]^{-1},\tag{118}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{P}(t-1)\boldsymbol{\Phi}(p,t)[\boldsymbol{I}_p+\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\boldsymbol{\Phi}(p,t)]^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\tag{119}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{120}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}.\tag{121}$$

$\boldsymbol{L}(t)\in\mathbf{R}^{n\times p}$为系统增益矩阵,$\boldsymbol{P}(t)\in\mathbf{R}^{n\times n}$为协方差矩阵,$p\geqslant 1$为新息长度,$\hat{\boldsymbol{\theta}}(t)$为$\boldsymbol{\theta}$在$t$时刻的估计. 算法的初值选择同常规最小二乘算法,如取$p_0\gg 1$,$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$. 当$p=1$时,上述算法退化为标准递推最小二乘算法.

值得指出的是:对于量测数据$\{y(t),\boldsymbol{\varphi}(t):t=1,2,\cdots,L\}$($L$为数据长度),随机梯度算法从$t=1$到$t=L$递推计算出参数估计,比递推最小二乘算法递推计算出的参数估计精度低,因为随机梯度算法收敛慢. 也就是说,两个算法同样使用了一批数据,但随机梯度算法从量测数据中提取的信息要少,故随机梯度算法使用数据的"效率"低,而多新息随机梯度算法提高了数据使用的效率,能提高参数估计精度. 最小二乘算法的收敛速度快,使用数据的效率本身就很高,所以多新息最小二乘算法对参数估计精度的改进是很有限的. 只有在数据缺失情况的间断递推或变递推间隔时,递推计算步数相同的情况下,才显示出计算效率的特点. 这就引出了下面的变递推间隔多新息最小二乘辨识方法.

5 变递推间隔多新息最小二乘辨识方法(*V-MILS*)

上述多新息最小二乘法与常规递推最小二乘法都是采用逐步递推计算,不能克服坏数据对参数估计的影响. 下面介绍具有克服坏数据能力或处理损失数据情形的变递推间隔最小二乘法.

考虑下列线性回归模型描述的标量系统:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{122}$$

其中 $y(t)\in\mathbf{R}$ 为系统输出,$\boldsymbol{\theta}\in\mathbf{R}^n$ 为待辨识的参数向量,$\boldsymbol{\varphi}(t)\in\mathbf{R}^n$ 是由系统输入 $u(t)\in\mathbf{R}$ 和输出 $y(t)$ 构成的回归信息向量,$v(t)\in\mathbf{R}$ 为零均值随机噪声.

辨识的目标是利用系统的输入输出数据$\{y(i),\boldsymbol{\varphi}(i),0\leqslant i\leqslant t\}$,提出变递推间隔多新息最小二乘辨识算法,对系统的未知参数向量 $\boldsymbol{\theta}$ 进行实时估计.

考虑 $t-p+1$ 到 t 时共 p 组数据,令

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

$$\boldsymbol{V}(p,t):=\begin{bmatrix}v(t)\\v(t-1)\\\vdots\\v(t-p+1)\end{bmatrix}\in\mathbf{R}^p,$$

$$\boldsymbol{\Phi}(p,t):=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]\in\mathbf{R}^{n\times p}.$$

由式(122)可得辨识模型:

$$\boldsymbol{Y}(p,t)=\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\theta}+\boldsymbol{V}(p,t). \tag{123}$$

定义整数序列$\{t_s,s=0,1,2,\cdots\}$满足

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*:=t_s-t_{s-1}.$$

用 t_s 代替式(123)中 t,可得变递推间隔多新息辨识方法的辨识模型:

$$\boldsymbol{Y}(p,t_s)=\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{\theta}+\boldsymbol{V}(p,t_s).$$

取准则函数为

$$J(\boldsymbol{\theta}):=\sum_{i=1}^{s}[\boldsymbol{Y}(p,t_i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_i)\boldsymbol{\theta}]^{\mathrm{T}}[\boldsymbol{Y}(p,t_i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_i)\boldsymbol{\theta}]=\sum_{i=1}^{s}\|\boldsymbol{Y}(p,t_i)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_i)\boldsymbol{\theta}\|^2.$$

参照递推最小二乘辨识方法的推导过程,能够得到变递推间隔多新息最小二乘辨识算法(V-MILS,interval-Varying Multi-Innovation Least Squares identification algorithm)(V-MILS 算法)如下[13]:

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{P}(t_s)\boldsymbol{\Phi}(p,t_s)[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\},$$

$$\boldsymbol{P}^{-1}(t_s)=\boldsymbol{P}^{-1}(t_{s-1})+\boldsymbol{\Phi}(p,t_s)\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s),$$

$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n,$$

$$\boldsymbol{\Phi}(p,t_s)=[\boldsymbol{\varphi}(t_s),\boldsymbol{\varphi}(t_s-1),\cdots,\boldsymbol{\varphi}(t_s-p+1)],$$

$$\boldsymbol{Y}(p,t_s)=[y(t_s),y(t_s-1),\cdots,y(t_s-p+1)]^{\mathrm{T}},$$

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*:=t_s-t_{s-1}.$$

或

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{L}(t_s)[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})], \tag{124}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\}, \tag{125}$$

$$\boldsymbol{L}(t_s)=\boldsymbol{P}(t_s)\boldsymbol{\Phi}(p,t_s)=\boldsymbol{P}(t_{s-1})\boldsymbol{\Phi}(p,t_s)[\boldsymbol{I}_p+\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{P}(t_{s-1})\boldsymbol{\Phi}(p,t_s)]^{-1}, \tag{126}$$

$$\boldsymbol{P}(t_s)=\boldsymbol{P}(t_{s-1})-\boldsymbol{L}(t_s)\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\boldsymbol{P}(t_{s-1}), \tag{127}$$

$$\boldsymbol{\Phi}(p,t_s)=[\boldsymbol{\varphi}(t_s),\boldsymbol{\varphi}(t_s-1),\cdots,\boldsymbol{\varphi}(t_s-p+1)], \tag{128}$$

$$\boldsymbol{Y}(p,t_s)=[y(t_s),y(t_s-1),\cdots,y(t_s-p+1)]^{\mathrm{T}}, \tag{129}$$

$$0=t_0<t_1<t_2<\cdots,\quad 1\leqslant t_s^*:=t_s-t_{s-1}. \tag{130}$$

$\boldsymbol{L}(t_s)\in\mathbf{R}^{n\times p}$为变递推间隔多新息最小二乘辨识算法的增益矩阵,$\boldsymbol{P}(t_s)\in\mathbf{R}^{n\times n}$为协方差矩阵.

从这个算法中,我们可以看出:修正的新息向量可以表达为

$$[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})]=\begin{bmatrix}y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})\\y(t_s-1)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s-1)\hat{\boldsymbol{\theta}}(t_{s-1})\\\vdots\\y(t_s-p+1)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s-p+1)\hat{\boldsymbol{\theta}}(t_{s-1})\end{bmatrix},$$

其中$\{y(t_s-i+1)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s-i+1)\hat{\boldsymbol{\theta}}(t_{s-1}),i=1,2,\cdots,p\}$分别为 $t=t_s,t_s-1,\cdots,t_s-p+1$ 时的新息. 因此从这个角度出发,我们把 $\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})$ 称为 $t=t_s$ 时刻的多新息,多新息辨识算法也是因此而得名的. 此外,t_s 不必取连续的自然数(即 t_s^* 不恒等于 1),该算法的递推间隔是变化的,这是与最小二乘算法的最大不同之处. 当遇到坏数据或不可信数据时,变递推间隔多新息辨识方法可跳过这部分数据,因而具有鲁棒性.

6 多新息最小二乘类辨识算法(*MILST*)

变递推间隔多新息最小二乘辨识可以派生出多新息最小二乘类辨识算法(MILST,Multi-Innovation Least Squares Type identification algorithm),有下列几种特别形式.

1) **递推最小二乘算法**

当递推间隔 $t_s^*=1$**,新息长度** $p=1$ **时,得到递推最小二乘辨识算法**(RLS,Recursive Least Squares identification algorithm):

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

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),\boldsymbol{P}(0)=p_0\boldsymbol{I}.$$

2) **变递推间隔最小二乘算法**

当新息长度 $p=1$ **时,得到变递推间隔最小二乘辨识算法**(V-LS,interval-Varying Least Squares identification algorithm)(V-LS **算法**):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{P}(t_s)\boldsymbol{\varphi}(t_s)[y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],\tag{131}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\tag{132}$$

$$\boldsymbol{P}^{-1}(t_s)=\boldsymbol{P}^{-1}(t_{s-1})+\boldsymbol{\varphi}(t_s)\boldsymbol{\varphi}^{\mathrm{T}}(t_s),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}.\tag{133}$$

3) **多新息最小二乘算法**

当递推间隔 $t_s^*=1(s=1,2,\cdots)$ **时,得到多新息最小二乘辨识算法**(MILS,Multi-Innovation Least Squares identification algorithm)(MILS **算法**):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{134}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{135}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{136}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}.\tag{137}$$

4) **等递推间隔多新息最小二乘算法**

当递推间隔 $t_s^*\equiv d$ **常数时,得到等递推间隔多新息最小二乘辨识算法**(E-MILS,interval-Equating Multi-Innovation Least Squares identification algorithm)(E-MILS **算法**):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-d)],\tag{138}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-d)+\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{139}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{140}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}.\tag{141}$$

5) **等递推间隔最小二乘算法**

当递推间隔间隔 $t_s^*\equiv d$**,新息长度** $p=1$ **时,得到等递推间隔最小二乘辨识算法**(E-LS,interval-Equating Least Squares identification algorithm)(E-LS **算法**):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-d)],\tag{142}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-d)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}.\tag{143}$$

6) **变递推间隔有限数据窗多新息最小二乘方法**

当数据窗长度为 q **时,可得到变递推间隔有限数据窗多新息最小二乘辨识方法**(V-FDW-MILS,interval-Varying Multi-Innovation Least Squares identification algorithm over the Finite Data Window)(V-FDW-MILS **算法**):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{P}(t_s)\boldsymbol{\Phi}(p,t_s)[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],\tag{144}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\tag{145}$$

$$\boldsymbol{P}^{-1}(t_s)=\sum_{i=0}^{q-1}\boldsymbol{\Phi}(p,t_{s-i})\boldsymbol{\Phi}^{\mathrm{T}}(p,t_{s-i})=\boldsymbol{P}^{-1}(t_{s-1})+\boldsymbol{\Phi}(p,t_s)\boldsymbol{\Phi}^{\mathrm{T}}(p,t_s)-\boldsymbol{\Phi}(p,t_{s-q+1})\boldsymbol{\Phi}^{\mathrm{T}}(p,t_{s-q+1}),\tag{146}$$

$$\boldsymbol{\Phi}(p,t_s)=[\boldsymbol{\varphi}(t_s),\boldsymbol{\varphi}(t_s-1),\cdots,\boldsymbol{\varphi}(t_s-p+1)],\tag{147}$$

$$\boldsymbol{Y}(p,t_s)=[y(t_s),y(t_s-1),\cdots,y(t_s-p+1)]^{\mathrm{T}}.\tag{148}$$

7) **有限数据窗多新息最小二乘算法**

当递推间隔 $t_s^*=1$**,数据窗长度为** q **时,得到有限数据窗多新息最小二乘辨识算法**(FDW-MILS,Multi-Innovation Least Squares identification algorithm over the Finite Data Window):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{149}$$

$$P^{-1}(t)=\sum_{i=0}^{q}\boldsymbol{\Phi}(p,t-i)\boldsymbol{\Phi}^{\mathrm{T}}(p,t-i)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}(p,t)\boldsymbol{\Phi}^{\mathrm{T}}(p,t)-\boldsymbol{\Phi}(p,t-q+1)\boldsymbol{\Phi}^{\mathrm{T}}(p,t-q+1),\tag{150}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{151}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}.\tag{152}$$

8) **等递推间隔有限数据窗多新息最小二乘算法**

当递推间隔 $t_s^*\equiv d$**,数据窗长度为** q **时,得到等递推间隔有限数据窗多新息最小二乘辨识算法**(E-FDW-MILS, interval-Equating Multi-Innovation Least Squares identification algorithm over the Finite Data Window)(E-FDW-MILS **算法**):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-d)+\boldsymbol{P}(t)\boldsymbol{\Phi}(p,t)[\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-d)],\tag{153}$$

$$\boldsymbol{P}^{-1}(t)=\sum_{i=0}^{q}\boldsymbol{\Phi}(p,t-i)\boldsymbol{\Phi}^{\mathrm{T}}(p,t-i),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{154}$$

$$\boldsymbol{\Phi}(p,t)=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)],\tag{155}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}. \quad (156)$$

9) 等递推间隔投影算法

当新息长度$p=1$,记忆长度(或数据窗长度)$q=1$时,得到等递推间隔投影辨识算法(E-Proj,interval-Equating Projection identification algorithm).

此外,在上述一些算法中引入遗忘因子λ,可以得到变递推间隔遗忘因子最小二乘辨识算法(V-FF-RLS)、遗忘因子多新息最小二乘辨识算法(FF-MILS)、等递推间隔遗忘因子递推最小二乘辨识算法(E-FF-RLS)等,这里不一一介绍了.

7 方程误差类系统(*EET*)

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,$\{v(t)\}$是零均值方差为σ^2的白噪声序列,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:

$$A(z):=1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a}\in\mathbf{R},$$
$$B(z):=b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b}\in\mathbf{R},$$
$$C(z):=1+c_1z^{-1}+c_2z^{-2}+\cdots+c_{n_c}z^{-n_c}\in\mathbf{R},$$
$$D(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d}\in\mathbf{R}.$$

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

$$A(z)y(t)=(1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a})y(t)=y(t)+a_1y(t-1)+a_2y(t-2)+\cdots+a_{n_a}y(t-n_a),$$
$$B(z)u(t)=(b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b})u(t)=b_1u(t-1)+b_2u(t-2)+\cdots+b_{n_b}u(t-n_b),$$
$$D(z)v(t)=(1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d})v(t)=v(t)+d_1v(t-1)+d_2v(t-2)+\cdots+d_{n_d}v(t-n_d).$$

设阶次n_a,n_b,n_c和n_d已知,且$t\leqslant0$时,$y(t)=0$,$u(t)=0$,$v(t)=0$.

上述讨论的各种多新息梯度型辨识方法和多新息最小二乘类辨识方法可以推广用于各种模型(包括有色噪声模型、非线性模型)的参数估计.为简化起见,下面的讨论只给出多新息随机梯度型辨识方法和多新息最小二乘类辨识方法,其他的变递推间隔多新息辨识方法、变递推间隔多新息最小二乘辨识方法等,这里不加讨论.

7.1 受控自回归系统(*CAR*)

考虑下列受控自回归系统(CAR)[14]:

$$A(z)y(t)=B(z)u(t)+v(t). \quad (157)$$

定义参数向量$\boldsymbol{\theta}$和信息向量$\boldsymbol{\varphi}(t)$如下:

$$\boldsymbol{\theta}:=[a_1,\cdots,a_{n_a},b_1,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^n,n:=n_a+n_b,$$
$$\boldsymbol{\varphi}(t):=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n,$$

把多项式$A(z)$和$B(z)$表达式代入式(157),并利用移位算子的性质得到CAR系统的辨识模型:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \quad (158)$$

这个线性回归模型的信息向量$\boldsymbol{\varphi}(t)$是由可测的输入输出数据构成的,故前面讨论的各种多新息梯度型算法和多新息最小二乘类算法都可用于上述模型的辨识.

7.2 受控自回归滑动平均系统(*CARMA*)

考虑下列受控自回归滑动平均系统(CARMA)[28]:

$$A(z)y(t)=B(z)u(t)+D(z)v(t). \quad (159)$$

定义增广参数向量$\boldsymbol{\theta}$和包含噪声项的信息向量$\boldsymbol{\varphi}(t)$如下:

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^n,\quad n:=n_a+n_b+n_d,$$
$$\boldsymbol{\varphi}(t):=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b),v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^n,$$

把多项式$A(z)$,$B(z)$和$D(z)$表达式代入式(159),并利用移位算子的性质得到CARMA系统的辨识模型:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \quad (160)$$

估计CARMA系统参数向量$\boldsymbol{\theta}$的增广随机梯度算法(ESG,Extended Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \quad (161)$$
$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \quad (162)$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \quad (163)$$
$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \quad (164)$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \quad (165)$$

为了改善ESG算法的收敛速度,仿照上面的多新息随机梯度算法的推导步骤[12,29-30],扩展标量新息(scalar innovation)$e(t)\in\mathbf{R}$为一个新息向量(即多新息)(innovation vector,i.e.,also multi-innovation)

$$\boldsymbol{E}(p,t)=$$

$$\begin{bmatrix} y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \\ y(t-1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}(t-1) \\ \vdots \\ y(t-p+1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}(t-1) \end{bmatrix} \in \mathbf{R}^p.$$

通过定义堆积输出向量 $\boldsymbol{Y}(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$：

$$\boldsymbol{Y}(p,t) = [y(t), y(t-1), \cdots, y(t-p+1)]^{\mathrm{T}},$$

$$\hat{\boldsymbol{\Phi}}(p,t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \cdots, \hat{\boldsymbol{\varphi}}(t-p+1)],$$

可以得到估计 CARMA 系统参数向量 $\boldsymbol{\theta}$ 的多新息增广随机梯度算法(MI-ESG，Multi-Innovation Extended Stochastic Gradient algorithm)[28]：

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{166}$$

$$\boldsymbol{E}(p,t) = \boldsymbol{Y}(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{167}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(p,t)\|^2, \quad r(0)=1, \tag{168}$$

$$\boldsymbol{Y}(p,t) = [y(t), y(t-1), \cdots, y(t-p+1)]^{\mathrm{T}}, \tag{169}$$

$$\hat{\boldsymbol{\Phi}}(p,t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \cdots, \hat{\boldsymbol{\varphi}}(t-p+1)], \tag{170}$$

$$\hat{\boldsymbol{\varphi}}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), u(t-1), u(t-2), \cdots, u(t-n_b), \hat{v}(t-1), \hat{v}(t-2), \cdots, \hat{v}(t-n_d)]^{\mathrm{T}}, \tag{171}$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{172}$$

式(168)可以修改为

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

MI-ESG **算法(166)—(172)随 t 增加，计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的步骤如下.**

1) **令** $t=1$. $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0$，$\hat{v}(t-i) = 1/p_0$ $(i=1,2,\cdots,n_d)$，$r(0)=1$，$p_0=10^6$.

2) **采集输入输出数据** $u(t)$ **和** $y(t)$，**由式**(169)**构造堆积输出向量** $\boldsymbol{Y}(p,t)$，**由式**(171)**构造信息向量** $\hat{\boldsymbol{\varphi}}(t)$，**由式**(170)**构造堆积信息矩阵** $\hat{\boldsymbol{\Phi}}(p,t)$.

3) **由式**(167)**计算新息向量** $\boldsymbol{E}(p,t)$，**由式**(168)**计算** $r(t)$.

4) **根据式**(166)**刷新参数估计向量** $\hat{\boldsymbol{\theta}}(t)$，**由式**(172)**计算残差** $\hat{v}(t)$.

5) t **增1，转到第2步.**

MI-ESG **算法与** MISG **算法有相似的性能.** MI-ESG **算法计算参数估计** $\hat{\boldsymbol{\theta}}(t)$ **的流程如图3所示.**

仿照多新息最小二乘算法(171)—(121)的推导，我们可以得到估计 CARMA 系统参数向量 $\boldsymbol{\theta}$ 的多新息增广最小二乘算法(MI-ELS，Multi-Innovation Extended Least Squares algorithm)[13]：

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{L}(t)[\boldsymbol{Y}(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{173}$$

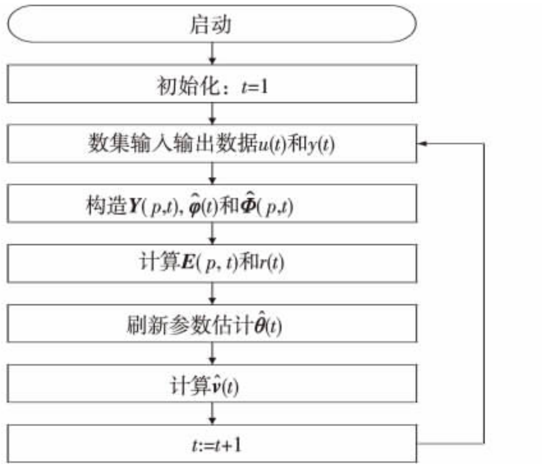

图3 MI-ESG **算法计算参数估计** $\hat{\boldsymbol{\theta}}(t)$ **的流程**

Fig. 3 The flowchart of computing the parameter estimate $\hat{\boldsymbol{\theta}}(t)$ in the MI-ESG algorithm

$$\boldsymbol{L}(t) = \boldsymbol{P}(t)\hat{\boldsymbol{\Phi}}(p,t) = \boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p + \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \tag{174}$$

$$\boldsymbol{P}(t) = \boldsymbol{P}(t-1) - \boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1), \quad \boldsymbol{P}(0) = p_0\boldsymbol{I}, \tag{175}$$

$$\boldsymbol{Y}(p,t) = [y(t), y(t-1), \cdots, y(t-p+1)]^{\mathrm{T}}, \tag{176}$$

$$\hat{\boldsymbol{\Phi}}(p,t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \cdots, \hat{\boldsymbol{\varphi}}(t-p+1)], \tag{177}$$

$$\hat{\boldsymbol{\varphi}}(t) = [-y(t-1), -y(t-2), \cdots, -y(t-n_a), u(t-1), u(t-2), \cdots, u(t-n_b), \hat{v}(t-1), \hat{v}(t-2), \cdots, \hat{v}(t-n_d)]^{\mathrm{T}}, \tag{178}$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{179}$$

这里 $\boldsymbol{E}(p,t) := \boldsymbol{Y}(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1) \in \mathbf{R}^p$ **为新息向量.**

7.3 受控自回归自回归系统(*CARAR*)

考虑下列受控自回归自回归系统(CARAR)，

$$A(z)y(t) = B(z)u(t) + \frac{1}{C(z)}v(t). \tag{180}$$

令

$$w(t) := \frac{1}{C(z)}v(t), \tag{181}$$

定义增广参数向量 $\boldsymbol{\theta}$ 和包含噪声项的信息向量 $\boldsymbol{\varphi}(t)$ 如下：

$$\boldsymbol{\theta} := [\boldsymbol{\theta}_{\mathrm{s}}^{\mathrm{T}}, c_1, c_2, \cdots, c_{n_c}]^{\mathrm{T}} \in \mathbf{R}^{n_a+n_b+n_c},$$

$$\boldsymbol{\theta}_{\mathrm{s}} := [a_1, a_2, \cdots, a_{n_a}, b_1, b_2, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}(t) = [\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t), -w(t-1), -w(t-2), \cdots, -w(t-n_c)]^{\mathrm{T}} \in \mathbf{R}^{n_a+n_b+n_c},$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t) := [-y(t-1), -y(t-2), \cdots, -y(t-n_a), u(t-1), u(t-2), \cdots, u(t-n_b)]^{\mathrm{T}} \in \mathbf{R}^{n_a+n_b}.$$

由式(180)—(181)可得

$$y(t)=[1-A(z)]y(t)+B(z)u(t)+w(t)=\boldsymbol{\varphi}_s^{\mathrm T}(t)\boldsymbol{\theta}_s+w(t)=\boldsymbol{\varphi}^{\mathrm T}(t)\boldsymbol{\theta}+v(t). \quad (182)$$

估计 CARAR 系统参数向量 $\boldsymbol{\theta}$ 的广义随机梯度算法(GSG,Generalized Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \quad (183)$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm T}(t)\hat{\boldsymbol{\theta}}(t-1), \quad (184)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \quad (185)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_s^{\mathrm T}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm T}, \quad (186)$$

$$\boldsymbol{\varphi}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm T}, \quad (187)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm T}(t)\hat{\boldsymbol{\theta}}_s(t), \quad (188)$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_s^{\mathrm T}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm T}, \quad (189)$$

$$\hat{\boldsymbol{\theta}}_s(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm T}. \quad (190)$$

估计 CARAR 系统参数向量 $\boldsymbol{\theta}$ 的多新息广义随机梯度算法(MI-GSG,Multi-Innovation Generalized Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \quad (191)$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm T}(p,t)\hat{\boldsymbol{\theta}}(t-1), \quad (192)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \quad (193)$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm T}, \quad (194)$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \quad (195)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_s^{\mathrm T}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm T}, \quad (196)$$

$$\boldsymbol{\varphi}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm T}, \quad (197)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm T}(t)\hat{\boldsymbol{\theta}}_s(t), \quad (198)$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_s^{\mathrm T}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm T}, \quad (199)$$

$$\hat{\boldsymbol{\theta}}_s(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm T}. \quad (200)$$

CARAR 系统参数向量 $\boldsymbol{\theta}$ 的多新息广义最小二乘算法(MI-GLS,Multi-Innovation Generalized Least Squares algorithm)[13]如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm T}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \quad (201)$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm T}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \quad (202)$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm T}(p,t)\boldsymbol{P}(t-1),$$

$$\boldsymbol{P}(0)=p_0\boldsymbol{I}, \quad (203)$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm T}, \quad (204)$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \quad (205)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_s^{\mathrm T}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm T}, \quad (206)$$

$$\boldsymbol{\varphi}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm T}, \quad (207)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm T}(t)\hat{\boldsymbol{\theta}}_s(t), \quad (208)$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_s^{\mathrm T}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm T}, \quad (209)$$

$$\hat{\boldsymbol{\theta}}_s(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm T}. \quad (210)$$

7.4 受控自回归自回归滑动平均系统(*CARARMA*)

考虑下列受控自回归自回归滑动平均系统(CARARMA):

$$A(z)y(t)=B(z)u(t)+\frac{D(z)}{C(z)}v(t). \quad (211)$$

定义中间相关噪声变量

$$w(t):=\frac{D(z)}{C(z)}v(t) \quad (212)$$

和参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_s\\ \boldsymbol{\theta}_n\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},$$

$$\boldsymbol{\theta}_s:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm T}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\theta}_n:=[c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm T}\in\mathbf{R}^{n_c+n_d},$$

$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_s(t)\\ \boldsymbol{\varphi}_n(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},$$

$$\boldsymbol{\varphi}_s(t):=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm T}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}_n(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c),v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm T}\in\mathbf{R}^{n_c+n_d}.$$

这里的下标 s 和 n 分别表示系统模型和噪声模型之意. 由式(211)—(212)可得辨识模型,

$$w(t)=[1-C(z)]w(t)+[D(z)-1]v(t)+v(t)=\boldsymbol{\varphi}_n^{\mathrm T}(t)\boldsymbol{\theta}_n+v(t), \quad (213)$$

$$y(t)=[1-A(z)]y(t)+B(z)u(t)+w(t)=\boldsymbol{\varphi}_s^{\mathrm T}(t)\boldsymbol{\theta}_s+w(t)=\boldsymbol{\varphi}_s^{\mathrm T}(t)\boldsymbol{\theta}_s+\boldsymbol{\varphi}_n^{\mathrm T}(t)\boldsymbol{\theta}_n+v(t)=\boldsymbol{\varphi}^{\mathrm T}(t)\boldsymbol{\theta}+v(t). \quad (214)$$

信息向量 $\boldsymbol{\varphi}(t)$ 中未知相关噪声项 $w(t-i)$ 和白噪声项 $v(t-i)$ 分别用其估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 代替,可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}$ 的广义增广随机梯度算法(GESG,Generalized Extended Stochastic Gradient algorithm):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \tag{215}$$
$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{216}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{217}$$
$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{218}$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{219}$$
$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{220}$$
$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{221}$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{222}$$
$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \tag{223}$$

估计 CARARMA 系统参数向量 $\boldsymbol{\theta}$ 的多新息广义随机梯度算法(MI-GESG,Multi-Innovation Generalized Extended Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{224}$$
$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{225}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{226}$$
$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{227}$$
$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{228}$$
$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{229}$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{230}$$
$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{231}$$
$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{232}$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{233}$$
$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \tag{234}$$

估计 CARARMA 系统参数向量 $\boldsymbol{\theta}$ 的多新息广义最小二乘算法(MI-GELS,Multi-Innovation Generalized Extended Least Squares algorithm)[13] 如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{235}$$
$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \tag{236}$$
$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}, \tag{237}$$
$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{238}$$
$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{239}$$
$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{240}$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{241}$$
$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{242}$$
$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{243}$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{244}$$
$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \tag{245}$$

8 输出误差类系统(OET)

8.1 输出误差系统(OE)

考虑输出误差系统[10-11,14]:

$$y(t)=\frac{B(z)}{A(z)}u(t)+v(t). \tag{246}$$

令

$$x(t):=\frac{B(z)}{A(z)}u(t). \tag{247}$$

定义参数向量 $\boldsymbol{\theta}$ 和包含噪声项的信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$
$$\boldsymbol{\varphi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b}.$$

由式(246)—(247)可得输出误差系统的辨识模型

$$x(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta},$$
$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{248}$$

估计输出系统参数向量 $\boldsymbol{\theta}$ 的辅助模型随机梯度算法(AM-SG,Auxiliary Model based Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \tag{249}$$
$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{250}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{251}$$
$$\hat{\boldsymbol{\varphi}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{252}$$
$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{253}$$

为了改善 AM-SG 算法的收敛速度,扩展标量新息 $e(t)\in\mathbf{R}$ 为一个新息向量(即多新息)$\boldsymbol{E}(p,t)\in\mathbf{R}^p$,可以得到估计输出系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息随机梯度算法(AM-MISG,Auxiliary Model based Multi-Innovation Stochastic Gradient algorithm)[44]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{254}$$
$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{255}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{256}$$
$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{257}$$
$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{258}$$
$$\hat{\boldsymbol{\varphi}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),$$

$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{259}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{260}$$

在这个算法的收敛性证明中,需要将式(260)修改为

$$\hat{x}(t-i)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-i)\hat{\boldsymbol{\theta}}(t),i=p-1,p-2,\cdots,1,0.$$

当新息长度 $p=1$ 时,AM-MISG 算法退化为 AM-SG 算法.

估计输出系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息最小二乘算法(AM-MILS,Auxiliary Model based Multi-Innovation Least Squares algorithm)[13]如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{261}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \tag{262}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}, \tag{263}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{264}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{265}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{266}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{267}$$

当新息长度 $p=1$ 时,AM-MILS 算法退化为 AM-RLS 算法.

8.2 输出误差滑动平均系统(*OEMA*)

考虑输出误差滑动平均模型(OEMA,Output Error Moving Average model)描述的系统

$$y(t)=\frac{B(z)}{A(z)}u(t)+D(z)v(t). \tag{268}$$

定义未知中间变量

$$x(t):=\frac{B(z)}{A(z)}u(t). \tag{269}$$

置系统参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}:=[\boldsymbol{\theta}_{\mathrm{s}}^{\mathrm{T}},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b+n_d},$$

$$\boldsymbol{\theta}_{\mathrm{s}}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}(t):=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b+n_d},$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b}.$$

借助于上述定义,式(269)和(268)可以写为下列辨识模型:

$$x(t)=[1-A(z)]x(t)+B(z)u(t) \tag{270}$$

$$=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}, \tag{271}$$

$$y(t)=x(t)+D(z)v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{272}$$

因为上式辨识模型信息向量 $\boldsymbol{\varphi}(t)$ 中不仅包含了不可测真实输出 $x(t-i)$,而且包含了噪声项 $v(t-i)$,我们用其估计值代替,可以得到估计 OEMA 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型增广随机梯度算法(AM-ESG,Auxiliary Model based Extended Stochastic gradient algorithm)[29]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \tag{273}$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{274}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{275}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{276}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{277}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{278}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{279}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}, \tag{280}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}. \tag{281}$$

估计 OEMA 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息增广随机梯度算法(AM-MI-ESG,Auxiliary Model based Multi-Innovation Extended Stochastic Gradient algorithm)[29]如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{282}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{283}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1, \tag{284}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{285}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{286}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{287}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{288}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{289}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{290}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}, \tag{291}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}. \tag{292}$$

估计 OEMA 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息增广最小二乘算法(AM-MI-ELS,Auxiliary Model based Multi-Innovation Extended Least Squares algorithm)[13]如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-$$

$$\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{293}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1},\tag{294}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{295}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{296}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{297}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{298}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\tag{299}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\theta}_{\mathrm{s}}(t),\tag{300}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{301}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}},\tag{302}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}.\tag{303}$$

8.3 输出误差自回归系统（*OEAR*）

考虑下列输出误差自回归模型（OEAR，Output Error AutoRegressive model）描述的系统：

$$y(t)=\frac{B(z)}{A(z)}u(t)+\frac{1}{C(z)}v(t).\tag{304}$$

定义系统真实输出 $x(t)$ 和噪声模型输出 $w(t)$ 分别为

$$x(t):=\frac{B(z)}{A(z)}u(t),\tag{305}$$

$$w(t):=\frac{1}{C(z)}v(t).\tag{306}$$

置系统参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下：

$$\boldsymbol{\theta}:=[\boldsymbol{\theta}_{\mathrm{s}}^{\mathrm{T}},c_1,c_2,\cdots,c_{n_c}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b+n_c},$$

$$\boldsymbol{\theta}_{\mathrm{s}}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}(t):=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),-w(t-1),-w(t-2),\cdots,-w(t-n_c)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b+n_c},$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b}.$$

借助于上述定义，式(305)—(306)可以写为

$$x(t)=[1-A(z)]x(t)+B(z)u(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}},$$

$$w(t)=[1-C(z)]w(t)+v(t).\tag{307}$$

则 OEAR 系统(304)可以写为下列辨识模型：

$$y(t)=x(t)+w(t)=x(t)+[1-C(z)]w(t)+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t).\tag{308}$$

式(308)辨识模型信息向量 $\boldsymbol{\varphi}(t)$ 中包含的不可测真实输出 $x(t-i)$ 和相关噪声项 $w(t-i)$ 用其估计值代替，可以得到辨识 OEAR 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型广义随机梯度算法（AM-GSG，Auxiliary Model based Generalized Stochastic Gradient algorithm）：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t),\tag{309}$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{310}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\tag{311}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}},\tag{312}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\tag{313}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\tag{314}$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\tag{315}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm{T}},\tag{316}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}.\tag{317}$$

辨识 OEAR 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义随机梯度算法（AM-MI-GSG，Auxiliary Model based Multi-Innovation Generalized Stochastic Gradient algorithm）如下：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t),\tag{318}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1),\tag{319}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\tag{320}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{321}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{322}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}},\tag{323}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\tag{324}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\tag{325}$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\tag{326}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm{T}},\tag{327}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}.\tag{328}$$

辨识 OEAR 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义最小二乘算法（AM-MI-GLS，Auxiliary Model based Multi-Innovation Generalized Least Squares algorithm）如下：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\tag{329}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1},\tag{330}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I},\tag{331}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\quad(332)$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\quad(333)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}},\quad(334)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\quad(335)$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\quad(336)$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\quad(337)$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm{T}},\quad(338)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}.\quad(339)$$

8.4 *Box-Jenkins* 系统(*BJ*)

考虑下列 Box-Jenkins 系统

$$y(t)=\frac{B(z)}{A(z)}u(t)+\frac{D(z)}{C(z)}v(t).\quad(340)$$

定义系统真实输出 $x(t)$ 和噪声模型输出 $w(t)$ 分别为

$$x(t):=\frac{B(z)}{A(z)}u(t),\quad(341)$$

$$w(t):=\frac{D(z)}{C(z)}v(t).\quad(342)$$

定义参数向量 $\boldsymbol{\theta}$ 和和信息向量 $\boldsymbol{\varphi}(t)$ 分别为

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},$$

$$\boldsymbol{\theta}_{\mathrm{s}}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\theta}_{\mathrm{n}}:=[c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$

$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \boldsymbol{\varphi}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}_{\mathrm{n}}(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c),v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}.$$

由式(341)—(342)可得

$$x(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}},\quad(343)$$

$$w(t)=[1-C(z)]w(t)+[D(z)-1]v(t)+v(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t),\quad(344)$$

把式(341)和(342)代入式(340)可得

$$y(t)=x(t)+w(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t).\quad(345)$$

辨识 Box-Jenkins 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型广义增广随机梯度算法(AM-GESG,Auxiliary Model based Generalized Extended Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t),\quad(346)$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\quad(347)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\quad(348)$$

$$\hat{\boldsymbol{\varphi}}(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix},\quad\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix},\quad(349)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\quad(350)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\quad(351)$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\quad(352)$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\quad(353)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).\quad(354)$$

辨识 Box-Jenkins 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义增广随机梯度算法(AM-MI-GESG,Auxiliary Model based Multi-Innovation Generalized Extended Stochastic Gradient algorithm)[41] 如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad(355)$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1),\quad(356)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\quad(357)$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\quad(358)$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\quad(359)$$

$$\hat{\boldsymbol{\varphi}}(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix},\quad\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix},\quad(360)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\quad(361)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\quad(362)$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\quad(363)$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\quad(364)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).\quad(365)$$

当新息长度 $p=1$ 时,AM-MI-GESG 算法退化为 AM-GESG 算法.

辨识 Box-Jenkins 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义增广最小二乘算法(AM-MI-GELS,Auxiliary Model based Multi-Innovation Generalized Extended Least Squares algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\quad(366)$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1},\quad(367)$$

$$P(t)=P(t-1)-L(t)\hat{\Phi}^{T}(p,t)P(t-1),\quad P(0)=p_0 I,\tag{368}$$

$$Y(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{T},\tag{369}$$

$$\hat{\Phi}(p,t)=[\hat{\varphi}(t),\hat{\varphi}(t-1),\cdots,\hat{\varphi}(t-p+1)],\tag{370}$$

$$\hat{\varphi}(t)=\begin{bmatrix}\hat{\varphi}_s(t)\\ \hat{\varphi}_n(t)\end{bmatrix},\quad \hat{\theta}(t)=\begin{bmatrix}\hat{\theta}_s(t)\\ \hat{\theta}_n(t)\end{bmatrix},\tag{371}$$

$$\hat{\varphi}_s(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{T},\tag{372}$$

$$\hat{\varphi}_n(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{T},\tag{373}$$

$$\hat{x}(t)=\hat{\varphi}_s^{T}(t)\hat{\theta}_s(t),\tag{374}$$

$$\hat{w}(t)=y(t)-\hat{x}(t),\tag{375}$$

$$\hat{v}(t)=y(t)-\hat{\varphi}^{T}(t)\hat{\theta}(t).\tag{376}$$

9 输入非线性受控自回归自回归滑动平均系统(IN-CARARMA)

考虑下列输入非线性受控自回归自回归滑动平均系统(IN-CARARMA，Input Nonlinear CARARMA system)[5]：

$$A(z)y(t)=B(z)\bar{u}(t)+\frac{D(z)}{C(z)}v(t),\tag{377}$$

其结构如图4所示，其非线性部分方程为

$$\bar{u}(t)=f(u(t))=\gamma_1 f_1(u(t))+\gamma_2 f_2(u(t))+\cdots+\gamma_m f_m(u(t))=f(u(t))\gamma,\tag{378}$$

其中 $f(u(t)):=[f_1(u(t)),f_2(u(t)),\cdots,f_m(u(t))]\in\mathbf{R}^{1\times m}$ 是基函数构成的行向量，$\gamma:=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{T}\in\mathbf{R}^{m}$ 是非线性部分的参数向量. 同样假定了多项式

$$B(z):=b_0+b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b}\tag{379}$$

的首项系数 $b_0=1$. 当然也可假定

$$B(z):=z^{-1}+b_2z^{-2}+b_3z^{-3}+\cdots+b_{n_b}z^{-n_b}.$$

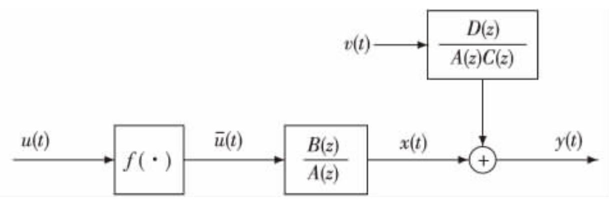

图4 输入非线性受控自回归自回归滑动平均系统(IN-CARARMA)
Fig. 4 The input nonlinear CARARMA system

定义中间相关噪声变量:

$$w(t):=\frac{D(z)}{C(z)}v(t).\tag{380}$$

定义参数向量 θ 和信息向量 $\varphi(t)$ 如下:

$$\theta:=\begin{bmatrix}\theta_s\\ \theta_n\end{bmatrix}\in\mathbf{R}^{n_a+n_b+m+n_c+n_d},\quad \theta_s:=\begin{bmatrix}a\\ b\\ \gamma\end{bmatrix}\in\mathbf{R}^{n_a+n_b+m},$$

$$\theta_n:=\begin{bmatrix}c\\ d\end{bmatrix}\in\mathbf{R}^{n_c+n_d},$$

$$a:=\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_{n_a}\end{bmatrix}\in\mathbf{R}^{n_a},\quad b:=\begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_{n_b}\end{bmatrix}\in\mathbf{R}^{n_b},$$

$$c:=\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_{n_c}\end{bmatrix}\in\mathbf{R}^{n_c},\quad d:=\begin{bmatrix}d_1\\ d_2\\ \vdots\\ d_{n_d}\end{bmatrix}\in\mathbf{R}^{n_d},$$

$$\varphi(t):=[\varphi_s^{T}(t),-w(t-1),-w(t-2),\cdots,-w(t-n_c),v(t-1),v(t-2),\cdots,v(t-n_d)]^{T}\in\mathbf{R}^{n_a+n_b+m+n_c+n_d},$$

$$\varphi_s(t):=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\bar{u}(t-1),\bar{u}(t-2),\cdots,\bar{u}(t-n_b),f(u(t))]\in\mathbf{R}^{n_a+n_b+m}.$$

将式(378)和(380)代入式(377)，可得辨识模型

$$y(t)=[1-A(z)]y(t)+[B(z)-1]\bar{u}(t)+\bar{u}(t)+w(t)=\varphi_s^{T}(t)\theta_s+w(t)=\tag{381}$$

$$\varphi_s^{T}(t)\theta_s+[1-C(z)]w(t)+[D(z)-1]v(t)+v(t)=\varphi^{T}(t)\theta+v(t).\tag{382}$$

信息向量 $\varphi(t)$ 中未知中间变量 $\bar{u}(t-i)$ 用辅助模型的输出 $\hat{\bar{u}}(t-i)$ 代替，未知相关噪声项 $w(t-i)$ 和白噪声项 $v(t-i)$ 分别用其估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 代替，可以得到估计 IN-CARARMA 系统参数向量 θ 的辅助模型广义增广随机梯度识算法(AM-GESG)：

$$\hat{\theta}(t)=\hat{\theta}(t-1)+\frac{\hat{\varphi}(t)}{r(t)}e(t),\tag{383}$$

$$e(t)=y(t)-\hat{\varphi}^{T}(t)\hat{\theta}(t-1),\tag{384}$$

$$r(t)=r(t-1)+\|\hat{\varphi}(t)\|^{2},\quad r(0)=1,\tag{385}$$

$$\hat{\varphi}(t)=[\hat{\varphi}_s^{T}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{T},\tag{386}$$

$$\hat{\varphi}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\hat{\bar{u}}(t-1),\hat{\bar{u}}(t-2),\cdots,\hat{\bar{u}}(t-n_b),f(u(t))]^{T},\tag{387}$$

$$\hat{\bar{u}}(t)=f(u(t))\hat{\gamma}(t),\tag{388}$$

$$f(u(t))=[f_1(u(t)),f_2(u(t)),\cdots,f_m(u(t))],\tag{389}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_s(t)\\ \hat{\boldsymbol{\theta}}_n(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_s(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\\ \hat{\boldsymbol{\gamma}}(t)\end{bmatrix}, \tag{390}$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{\varphi}}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t), \tag{391}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{392}$$

估计 IN-CARARMA 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义增广随机梯度算法(AM-MI-GESG, Auxiliary Model based Multi-Innovation Generalized Extended Stochastic Gradient algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{393}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{394}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\ r(0)=1, \tag{395}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{396}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{397}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_s^{\mathrm{T}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{398}$$

$$\hat{\boldsymbol{\varphi}}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\hat{\bar{u}}(t-1),\hat{\bar{u}}(t-2),\cdots,\hat{\bar{u}}(t-n_b),\boldsymbol{f}(u(t))]^{\mathrm{T}}, \tag{399}$$

$$\hat{\bar{u}}(t)=\boldsymbol{f}(u(t))\hat{\boldsymbol{\gamma}}(t), \tag{400}$$

$$\boldsymbol{f}(u(t))=[f_1(u(t)),f_2(u(t)),\cdots,f_m(u(t))], \tag{401}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_s(t)\\ \hat{\boldsymbol{\theta}}_n(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_s(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\\ \hat{\boldsymbol{\gamma}}(t)\end{bmatrix}, \tag{402}$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{\varphi}}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t), \tag{403}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{404}$$

估计 IN-CARARMA 系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息广义增广最小二乘算法(AM-MI-GELS, Auxiliary Model based Multi-Innovation Generalized Extended Least Squares algorithm)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \tag{405}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \tag{406}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1),\ \boldsymbol{P}(0)=p_0\boldsymbol{I}, \tag{407}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{408}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{409}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_s^{\mathrm{T}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{410}$$

$$\hat{\boldsymbol{\varphi}}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\hat{\bar{u}}(t-1),\hat{\bar{u}}(t-2),\cdots,\hat{\bar{u}}(t-n_b),\boldsymbol{f}(u(t))]^{\mathrm{T}}, \tag{411}$$

$$\hat{\bar{u}}(t)=\boldsymbol{f}(u(t))\hat{\boldsymbol{\gamma}}(t), \tag{412}$$

$$\boldsymbol{f}(u(t))=[f_1(u(t)),f_2(u(t)),\cdots,f_m(u(t))], \tag{413}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_s(t)\\ \hat{\boldsymbol{\theta}}_n(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_s(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\\ \hat{\boldsymbol{\gamma}}(t)\end{bmatrix}, \tag{414}$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{\varphi}}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t), \tag{415}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{416}$$

本节讨论了输入非线性 CARARMA 系统的辅助模型广义增广随机梯度算法、辅助模型多新息广义增广随机梯度算法、辅助模型广义增广最小二乘算法,它包括了输入非线性 FIR 系统(IN-FIR)、输入非线性 CARMA 系统(IN-CARMA)、输入非线性 CARAR 系统(IN-CARAR)等特殊情形. 这个方法可以推广到输入非线性输出误差系统(IN-OE)、输入非线性输出误差滑动平均系统(IN-OEMA)、输入非线性输出误差自回归系统(IN-OEAR)和输入非线性 Box-Jenkins 系统(IN-BJ). 这些辨识算法的收敛性分析(参数估计的一致收敛性,估计误差的有界收敛性)仍然是控制科学家有待解决的辨识难题.

10 多新息观测器和多新息卡尔曼滤波器(*MI-Observer and MI-KF*)

10.1 多新息观测器(*MI-Observer*)

考虑离散时间状态空间模型描述的系统

$$\begin{cases}\boldsymbol{x}(t+1)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{u}(t),\\ \boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{x}(t)+\boldsymbol{D}\boldsymbol{u}(t),\end{cases} \tag{417}$$

其中 $\boldsymbol{x}(t)\in\mathbf{R}^n$ 为状态向量, $\boldsymbol{u}(t)\in\mathbf{R}^r$ 为输入向量, $\boldsymbol{y}(t)\in\mathbf{R}^m$ 为输出向量, $\boldsymbol{A}\in\mathbf{R}^{n\times n}$, $\boldsymbol{B}\in\mathbf{R}^{n\times r}$, $\boldsymbol{C}\in\mathbf{R}^{m\times n}$ 和 $\boldsymbol{D}\in\mathbf{R}^{m\times r}$ 为系统参数矩阵.

系统(417)的闭环观测器为

$$\begin{cases}\hat{\boldsymbol{x}}(t+1)=\boldsymbol{A}\hat{\boldsymbol{x}}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\boldsymbol{L}[\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(t)]=:\\ \qquad \boldsymbol{A}\hat{\boldsymbol{x}}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\boldsymbol{L}\boldsymbol{e}(t),\\ \hat{\boldsymbol{y}}(t)=\boldsymbol{C}\hat{\boldsymbol{x}}(t)+\boldsymbol{D}\boldsymbol{u}(t).\end{cases}$$

其中 $\boldsymbol{L}\in\mathbf{R}^{n\times m}$ 为增益矩阵, $\boldsymbol{e}(t):=\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(t)=\boldsymbol{y}(t)-\boldsymbol{C}\hat{\boldsymbol{x}}(t)-\boldsymbol{D}\boldsymbol{u}(t)$ 为向量新息,用于反馈校正观测器状态偏差. 扩展 $\boldsymbol{e}(t)$ 则得到多新息观测器:

$$\begin{cases}\hat{\boldsymbol{x}}(t+1)=\boldsymbol{A}\hat{\boldsymbol{x}}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\sum_{i=1}^{p}\boldsymbol{L}_i\boldsymbol{e}(t-i+1),\\ \hat{\boldsymbol{y}}(t)=\boldsymbol{C}\hat{\boldsymbol{x}}(t)+\boldsymbol{D}\boldsymbol{u}(t),\\ \boldsymbol{e}(t)=\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(t)=\boldsymbol{y}(t)-\boldsymbol{C}\hat{\boldsymbol{x}}(t)-\boldsymbol{D}\boldsymbol{u}(t).\end{cases}$$

其中 $L_i \in \mathbf{R}^{n\times m}$ 为增益矩阵，p 为新息长度.

10.2 多新息卡尔曼滤波器（*Multi-innovation Kalman filter*）

考虑下列随机系统状态空间模型

$$\begin{cases} x(t+1)=Ax(t)+Bu(t)+w(t),\\ y(t)=Cx(t)+Du(t)+v(t). \end{cases} \quad (418)$$

其中 $x(t)\in\mathbf{R}^n$ 为状态向量，$u(t)\in\mathbf{R}^r$ 为输入向量，$y(t)\in\mathbf{R}^m$ 为输出向量，$w(t)\in\mathbf{R}^n$ 为零均值过程噪声向量（process noise vector），$v(t)\in\mathbf{R}^m$ 为零均值观测噪声向量（observation noise vector），$A\in\mathbf{R}^{n\times n}$，$B\in\mathbf{R}^{n\times r}$，$C\in\mathbf{R}^{m\times n}$ 和 $D\in\mathbf{R}^{m\times r}$ 为系统参数矩阵，$E[w(t)w^{\mathrm{T}}(t)]=R_w\in\mathbf{R}^{n\times n}$，$\mathrm{E}[v(t)v^{\mathrm{T}}(t)]=R_v\in\mathbf{R}^{m\times m}$.

假设白噪声过程 $\{w(t)\}$ 和 $\{v(t)\}$ 与系统输入 $\{u(t)\}$ 不相关，一步超前卡尔曼状态估计算法（one-step ahead Kalman state estimation algorithm）或卡尔曼状态滤波算法（卡尔曼新息滤波器）如下：

$$\begin{cases} \hat{x}(t+1)=A\hat{x}(t)+Bu(t)+L(t)e(t),\\ L(t)=AP(t)C^{\mathrm{T}}[R_v+CP(t)C^{\mathrm{T}}]^{-1},\\ P(t+1)=AP(t)A^{\mathrm{T}}+R_w-\\ \qquad AP(t)C^{\mathrm{T}}[R_v+CP(t)C^{\mathrm{T}}]^{-1}CP(t)A^{\mathrm{T}}. \end{cases}$$

其中 $L(t)\in\mathbf{R}^{n\times m}$ 为时变增益矩阵，$e(t):=y(t)-\hat{y}(t)=y(t)-C\hat{x}(t)-Du(t)$ 为向量新息. 扩展 $e(t)$ 为新息矩阵

$$E(p,t)=\begin{bmatrix} e(t)\\ e(t-1)\\ e(t-2)\\ \vdots\\ e(t-p+1)\end{bmatrix}=$$

$$\begin{bmatrix} y(t)-C\hat{x}(t)-Du(t)\\ y(t-1)-C\hat{x}(t-1)-Du(t-1)\\ y(t-2)-C\hat{x}(t-2)-Du(t-2)\\ \vdots\\ y(t-p+1)-C\hat{x}(t-p+1)-Du(t-p+1)\end{bmatrix}\in\mathbf{R}^{(mp)},$$

得到多新息卡尔曼状态估计算法（多新息卡尔曼状态滤波算法，多新息卡尔曼新息滤波器）

$$\hat{x}(t+1)=$$
$$A\hat{x}(t)+Bu(t)+[L_1(t),L_2(t),\cdots,L_p(t)]E(p,t)=$$
$$A\hat{x}(t)+Bu(t)+\sum_{i=1}^{p}L_i(t)e(t-i+1).$$

这里可取增益矩阵 $L_i(t)=L(t-i+1)$.

多新息卡尔曼滤波器优于（至少等于）卡尔曼滤波器，因为只需取 $L_1(t)=L(t)$，$L_2(t)=L_3(t)=\cdots=L_p(t)=\mathbf{0}$.

11 结语

系统辨识是利用系统采样数据信息辨识系统数学模型参数的过程. 多新息辨识方法是充分利用和扩展辨识新息的一种辨识方法，即通过对辨识新息的扩展，从标量新息到新息向量，从向量新息到新息矩阵，提出的一种基于新息的辨识理论与方法. 这种多新息辨识理论可以发展到观测器设计和卡尔曼滤波中，研究和提出多新息观测器设计和多新息卡尔曼滤波理论与方法.

多新息辨识方法是系统辨识的一个重要分支. 本文详细讨论了线性回归模型、方程误差类系统、输出误差类系统、输入非线性系统的多新息梯度型辨识方法和多新息最小二乘类辨识方法等，简单说明了多新息辨识理论可以应用于观测器设计和卡尔曼滤波器的设计. 这些方法也可推广到输出非线性方程误差类系统和输出非线性输出误差类系统，以及反馈非线性系统. 一些有色噪声干扰系统多新息辨识算法的收敛性分析仍然是辨识领域有待解决的困难课题.

参考文献
References

[1] 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

[2] 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

[3] 丁锋. 系统辨识（3）：辨识精度与辨识基本问题［J］. 南京信息工程大学学报：自然科学版，2011，3（3）：193-226
DING Feng. System identification. Part C：Identification accuracy and basic problems［J］. Journal of Nanjing University of Information Science & Technology：Natural Science Edition，2011，3（3）：193-226

[4] 丁锋. 系统辨识（4）：辅助模型辨识思想与方法［J］. 南京信息工程大学学报：自然科学版，2011，3（4）：289-318
DING Feng. System identification. Part D：Auxiliary model identification idea and methods［J］. Journal of Nanjing University of Information Science & Technology：Natural Science Edition，2011，3（4）：289-318

[5] 丁锋. 系统辨识(5):迭代搜索原理与辨识方法[J]. 南京信息工程大学学报:自然科学版,2011,3(6):481-510
DING Feng. System identification. Part E: Iterative search principle and identification methods[J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition,2011,3(6):481-510

[6] 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

[7] Ding F,Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model[J]. Automatica,2004,40(10):1739-1748

[8] Ding F,Chen T. Identification of dual-rate systems based on finite impulse response models[J]. International Journal of Adaptive Control and Signal Processing,2004,18(7):589-598

[9] Ding F,Chen T. Parameter estimation of dual-rate stochastic systems by using an output error method[J]. IEEE Transactions on Automatic Control,2005,50(9):1436-1441

[10] 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

[11] Ding F,Liu G,Liu X P. Parameter estimation with scarce measurements[J]. Automatica,2011,47(8):1646-1655

[12] Ding F,Chen T. Performance analysis of multi-innovation gradient type identification methods[J]. Automatica,2007,43(1):1-14

[13] 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

[14] Ding F. Several multi-innovation identification methods[J]. Digital Signal Processing,2010,20(4):1027-1039

[15] Ding F,Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems[J]. Automatica,2005,41(2):315-325

[16] Ding F,Chen T. Hierarchical least squares identification methods for multivariable systems[J]. IEEE Transactions on Automatic Control,2005,50(3):397-402

[17] Ding F,Chen T. Hierarchical identification of lifted state-space models for general dual-rate systems[J]. IEEE Transactions on Circuits and Systems Ⅰ: Regular Papers,2005,52(6):1179-1187

[18] Ding F,Qiu L,Chen T. Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems[J]. Automatica,2009,45(2):324-332.

[19] Ding F,Liu G,Liu X P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems[J]. IEEE Transactions on Automatic Control,2010,55(8):1976-1981

[20] 丁锋,谢新民. 传递函数阵子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

[21] 丁锋,谢新民. 多变量系统的辅助模型辨识算法[J]. 清华大学学报:自然科学版,1992,32(4):100-106
DING Feng,XIE Xinmin. Auxiliary model identification method for multivariable systems[J]. Journal of Tsinghua University: Science and Technology,1992,32(4):100-106

[22] 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

[23] Ding F,Chen H B,Li M. Multi-innovation least squares identification methods based on the auxiliary model for MISO systems[J]. Applied Mathematics and Computation,2007,187(2):658-668

[24] 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

[25] 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

[26] 丁锋,丁韬. 时变多变量系统多新息投影算法的均方收敛性[J]. 湖北工学院学报,2001,16(4):16-20
DING Feng,DING Tao. Mean square convergence of multi-innovation projection algorithms for time-varying multivariable systems[J]. Journal of Hubei University of Technology,2001,16(4):16-20

[27] 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

[28] 于丽,刘艳君,丁锋. CARMA 模型多新息增广随机梯度参数估计算法的收敛性分析[J]. 系统工程与电子技术,2009,31(6):1446-1449
YU Li,LIU Yanjun,DING Feng. Convergence of multi-

innovation extended stochastic gradient parameter estimation for CARMA models [J]. Systems Engineering and Electronics,2009,31(6):1446-1449

[29] 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

[30] Liu Y J,Yu L,Ding F. Multi-innovation extended stochastic gradient algorithm and its performance analysis [J]. Circuits, Systems and Signal Processing, 2010, 2010,29(4):649-667

[31] **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

[32] Ding F,Yang J B,Xu Y M. Convergence of hierarchical stochastic gradient identification for transfer function matrix models [J]. Control Theory and Applications,2001,18(6):949-953

[33] **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

[34] **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

[35] **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.

[36] **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

[37] **丁锋. 时变参数系统辨识及其应用 [D]. 北京:清华大学自动化系**,1994
DING Feng. Time-varying parameter system identification and its applications [D]. Beijing: Department of Automation,Tsinghua University,1994

[38] Moustafa K A F. Identification of stochastic time-varying systems [J]. IEE Proceeding,Part D,Control Theory and Applications,1983,130(4):137-142

[39] 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

[40] Han L L,Ding F. Identification for multirate multi-input systems using the multi-innovation identification theory [J]. Computers & Mathematics with Applications,2009,57(9):1438-1449

[41] **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

[42] **刘英玉,申东日,陈义俊,等. 基于前向神经网络的多新息随机梯度辨识算法 [J]. 哈尔滨商业大学学报:自然科学版**,2006,22(2):83-86
LIU Yingyu,SHEN Dongri,CHEN Yijun,et al. Multi-innovation stochastic gradient identification algorithm based on feedforward neural networks [J]. Journal of Harbin University of Commerce: Natural Sciences Edition,2006,22(2):83-86

[43] **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

[44] 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

[45] Liu Y J,Xiao Y S,Zhao X L. Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model [J]. Applied Mathematics and Computation,2009,215(4):1477-1483

[46] Xie L,Yang H Z,Ding F. Modeling and identification for non-uniformly periodically sampled-data systems [J]. IET Control Theory & Applications,2010,4(5):784-794

[47] Chen J,Zhang Y,Ding R F. Auxiliary model based multi-innovation algorithms for multivariable nonlinear systems [J]. Mathematical and Computer Modelling,52(9/10):1428-1434

[48] Han L L, Ding F. Parameter estimation for multirate multi-input systems using auxiliary model and multi-innovation [J]. Journal of Systems Engineering and Electronics,2010,21(6):1079-1083.

[49] **丁锋. 系统辨识理方法论 [M]. 北京:电力出版社**,2012
DING Feng. System identification theory and methods [M]. Beijing: China Electric Power Press,2012

System identification. Part F: Multi-innovation identification theory and methods

DING Feng[1,2,3]

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

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

3 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122

Abstract Multi-innovation identification is an important branch of system identification. The innovation is the useful information that can improve parameter estimation or state estimation accuracies. This paper discusses various multi-innovation identification methods for linear regression models, including the multi-innovation projection algorithm, the multi-innovation stochastic gradient algorithm, the multi-innovation forgetting gradient algorithm, the interval-varying multi-innovation stochastic gradient algorithm, the multi-innovation least squares algorithm, the interval-varying multi-innovation least squares algorithm, and so on. We give the stochastic gradient algorithm, the multi-innovation stochastic gradient algorithm and the multi-innovation least squares identification algorithm for equation error type systems, output error type systems and input nonlinear systems. Finally, we state that the multi-innovation identification theory can be developed to multi-innovation observer and multi-innovation Kalman filtering theory.

Key words iterative identification; recursive identification; parameter estimation; FIR model; equation error model; CAR model; CARMA model; CARAR model; CARARMA model; output error model; OEMA model; OEAR model; auxiliary model identification; multi-innovation identification; hierarchical identification; coupled identification