文章编号:1674-7070(2015)05-0385-23

丁锋[1,2,3]

方程误差系统的多新息辨识方法

摘要

多新息方法可以用于线性系统和非线性系统的自适应滤波、参数估计、自校正控制、自适应故障检测与诊断等.线性系统包括两种基本类型:方程误差类系统和输出误差类系统.本文将多新息辨识应用到方程误差滑动平均(EEMA)系统(即CARMA系统),研究多新息增广随机梯度算法和多新息增广最小二乘算法,应用到方程误差自回归滑动平均(EEARMA)系统(即CARARMA系统),提出基于分解的多新息广义增广随机梯度算法和基于分解的多新息广义增广最小二乘算法,以及基于滤波的多新息广义增广随机梯度算法和基于滤波的多新息广义增广最小二乘算法.

关键词

参数估计;递推辨识;梯度搜索;最小二乘;滤波;分解;辅助模型辨识思想;多新息辨识理论;递阶辨识原理;方程误差系统;线性系统

中图分类号 TP273

文献标志码 A

收稿日期 2015-09-20

资助项目 国家自然科学基金(61273194);江苏省自然科学基金(BK2012549);高等学校学科创新引智"111计划"(B12018)

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

新颖的辨识方法都是蕴藏在科学云端的璀璨明珠.新辨识思想、新辨识理论、新辨识原理、新辨识概念的诞生都是辨识史上的重要里程碑,都标志着大量新颖辨识方法的涌现.近年来,辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念的出现都标志着一个个辨识分支的诞生.相继提出的辅助模型辨识方法、多新息辨识方法、迭代辨识方法、递阶辨识方法、耦合辨识方法等构成了现代系统辨识的精髓,也是最近出版的《系统辨识新论》[1]的主要组成部分.一些典型辨识方法的收敛性分析汇聚在《系统辨识——辨识方法性能分析》[2]一书,而在《南京信息工程大学学报》上发表的关于系统辨识的连载论文[3-23],更是给系统辨识研究的主旋律添上了玫瑰般的色彩.限于篇幅,本文仅研究CARMA系统多新息辨识方法、CARARMA系统基于分解的多新息辨识方法和基于滤波的多新息辨识方法,这些方法的收敛性分析为辨识科学提供了广阔的研究空间,也是呈现在读者面前的耀眼明珠,期盼着辨识科学家来采撷.

最小二乘辨识方法和随机梯度辨识方法是两类重要的辨识方法,它们派生出一些相关联的辨识方法,如有限数据窗最小二乘算法和有限数据窗随机梯度算法、遗忘因子最小二乘算法和遗忘因子随机梯度算法、带遗忘因子的有限数据窗最小二乘算法和带遗忘因子的有限数据窗随机梯度算法等[2,24],以及这些方法在有色噪声干扰系统,如CARMA模型(CARAR模型、CARARMA模型)描述的系统中的应用,便产生遗忘因子和(或)有限数据窗、增广(广义、广义增广)最小二乘算法和增广(广义、广义增广)随机梯度算法等.最近诞生的辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念等与一些经典的辨识方法相结合,产生一系列新的辨识算法,如辅助模型增广随机梯度算法、辅助模型多新息增广随机梯度算法、辅助模型递推增广最小二乘算法、辅助模型多新息增广最小二乘算法、基于分解的多新息广义增广随机梯度算法、基于分解的多新息广义增广最小二乘算法、基于滤波的多新息广义增广随机梯度算法、基于滤波的多新息广义增广最小二乘算法等.本文主要研究CARMA系统的多新息随机梯度算法和多新息最小二乘算法,以及CARARMA系统的分解多新息辨识算法和滤波多新息辨识算法.

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:

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

多项式系数 a_i,b_i,c_i 和 d_i 为待辨识的模型参数,设阶次 n_a,n_b,n_c 和 n_d 已知,定义待辨识的参数向量:

$$\boldsymbol{a}:=[a_1,a_2,\cdots,a_{n_a}]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$
$$\boldsymbol{b}:=[b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_b},$$
$$\boldsymbol{c}:=[c_1,c_2,\cdots,c_{n_c}]^{\mathrm{T}}\in\mathbf{R}^{n_c},$$
$$\boldsymbol{d}:=[d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_d}.$$

定义信息向量:

$$\boldsymbol{\varphi}_y(t):=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$
$$\boldsymbol{\varphi}_u(t):=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_b},$$
$$\boldsymbol{\varphi}_w(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c)]^{\mathrm{T}}\in\mathbf{R}^{n_c},$$
$$\boldsymbol{\varphi}_v(t):=[v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_d}.$$

根据移位算子的性质,有

$$\begin{aligned}A(z)y(t)&=(1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a})y(t)=\\&y(t)+a_1y(t-1)+a_2y(t-2)+\cdots+a_{n_a}y(t-n_a)=\\&y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a},\\B(z)u(t)&=(b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b})u(t)=\\&b_1u(t-1)+b_2u(t-2)+\cdots+b_{n_b}u(t-n_b)=\\&\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b},\\D(z)v(t)&=(1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d})v(t)=\\&v(t)+d_1v(t-1)+d_2v(t-2)+\cdots+d_{n_d}v(t-n_d)=\\&v(t)+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}.\end{aligned}$$

1 有限数据窗递推辨识方法

考虑下列线性回归模型描述的标量系统:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t),\tag{1}$$

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.

1.1 有限数据窗递推最小二乘算法

递推最小二乘算法使用直到时刻 t 的所有数据$\{y(j),\boldsymbol{\varphi}(j):1\leqslant j\leqslant t\}$,即 $j=1$ 到 $j=t$ 所有数据(t 组数据)来构成准则函数

$$J_1(\boldsymbol{\theta}):=\sum_{j=1}^{t}[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2,$$

极小化 $J_1(\boldsymbol{\theta})$ 得到递推最小二乘算法(Recursive Least Squares algorithm,RLS 算法):

$$\begin{aligned}&\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\\&\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{\mathrm{n}}/p_0,\end{aligned}\tag{2}$$

$$\begin{aligned}&\boldsymbol{P}^{-1}(t)=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)=\\&\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t),\quad\boldsymbol{P}(0)=p_0\boldsymbol{I}_{\mathrm{n}}.\end{aligned}\tag{3}$$

而有限数据窗最小二乘算法是根据最近长度为 q 的动态窗口内的数据$\{y(j),\hat{\boldsymbol{\varphi}}(j):t-q+1\leqslant j\leqslant t\}$来构造准则函数,即考虑 $i=t-q+1$ 到 $i=t$ 有限数据窗(finite data window)内的 q 组数据(q 为数据窗长度)构成准则函数

$$J_2(\boldsymbol{\theta}):=\sum_{j=t-q+1}^{t}[y(j)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2.$$

观察 RLS 算法协方差阵的定义式(3),比较准则函数 $J_1(\boldsymbol{\theta})$和 $J_2(\boldsymbol{\theta})$,极小化 $J_2(\boldsymbol{\theta})$,通过简单类比可以得到一个简化的有限数据窗递推最小二乘算法(FDW-RLS 算法)[24]:

$$\begin{aligned}&\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\\&\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\end{aligned}\tag{4}$$

$$\begin{aligned}&\boldsymbol{P}^{-1}(t)=\sum_{j=t-q+1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)=\\&\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)-\boldsymbol{\varphi}(t-q)\boldsymbol{\varphi}^{\mathrm{T}}(t-q),\\&\boldsymbol{P}(0)=p_0\boldsymbol{I}_n.\end{aligned}\tag{5}$$

为了避免计算矩阵逆,定义中间矩阵

$$\boldsymbol{P}_1^{-1}(t):=\boldsymbol{P}^{-1}(t-1)-\boldsymbol{\varphi}(t-q)\boldsymbol{\varphi}^{\mathrm{T}}(t-q),$$

式(5)中的协方差阵 $\boldsymbol{P}(t)$可以表示为

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}_1^{-1}(t)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t).$$

应用矩阵求逆引理

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

于以上两式,可得

$$\begin{aligned}&\boldsymbol{P}_1(t)=\boldsymbol{P}(t-1)+\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)\boldsymbol{\varphi}^{\mathrm{T}}(t-q)\boldsymbol{P}(t-1)}{1-\boldsymbol{\varphi}^{\mathrm{T}}(t-q)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)},\\&\boldsymbol{P}(0)=p_0\boldsymbol{I}_n,\end{aligned}\tag{6}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}_1(t)-\frac{\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}_1(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)}.\tag{7}$$

定义增益向量 $\boldsymbol{L}(t):=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$,$\boldsymbol{L}_1(t):=\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t-q)\in\mathbf{R}^n$.使用式(7)和(6)可得

$$\begin{aligned}&\boldsymbol{L}(t)=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)=\\&\qquad\frac{\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)},\\&\boldsymbol{L}_1(t)=\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t-q)=\\&\qquad\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)}{1-\boldsymbol{\varphi}^{\mathrm{T}}(t-q)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)}.\end{aligned}$$

因此,简化的有限数据窗递推最小二乘算法(Recursive Least Squares algorithm over a Finite Data Window,FDW-RLS 算法)可归纳为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$
$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{8}$$
$$\boldsymbol{L}(t)=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)=\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}_1(t)\boldsymbol{\varphi}(t)]^{-1},$$
$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{P}_1(t), \tag{9}$$
$$\boldsymbol{L}_1(t)=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)[1-\boldsymbol{\varphi}^{\mathrm{T}}(t-q)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t-q)]^{-1},$$
$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{10}$$
$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_n+\boldsymbol{L}_1(t)\boldsymbol{\varphi}^{\mathrm{T}}(t-q)]\boldsymbol{P}(t-1). \tag{11}$$

这个算法及其收敛性证明可参见文献[2,25].

1.2 有限数据窗随机梯度辨识算法

RLS 算法(2)—(3)中的增益向量为 $\boldsymbol{L}(t):=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)$,若将增益向量修改为 $\boldsymbol{L}(t):=\dfrac{\boldsymbol{\varphi}(t)}{r(t)}$,$r(t)$ 取为式(3)中 $\boldsymbol{P}^{-1}(t)$ 的迹,即

$$r(t):=\mathrm{tr}[\boldsymbol{P}^{-1}(t)]=$$
$$\mathrm{tr}[\boldsymbol{P}^{-1}(t-1)]+\mathrm{tr}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]=$$
$$r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=n/p_0,$$

就得到随机梯度算法(Stochastic Gradient algorithm,SG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$
$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{12}$$
$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1. \tag{13}$$

如果 $r(t)$ 取为式(5)中 $\boldsymbol{P}^{-1}(t)$ 的迹,即

$$r(t):=\mathrm{tr}[\boldsymbol{P}^{-1}(t)]=$$
$$\mathrm{tr}[\boldsymbol{P}^{-1}(t-1)]+\mathrm{tr}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]-\mathrm{tr}[\boldsymbol{\varphi}(t-q)\boldsymbol{\varphi}^{\mathrm{T}}(t-q)]=$$
$$r(t-1)+\|\boldsymbol{\varphi}(t)\|^2-\|\boldsymbol{\varphi}(t-q)\|^2,\quad r(0)=n/p_0, \tag{14}$$

就得到有限数据窗随机梯度算法(FDW-SG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$
$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{15}$$
$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2-\|\boldsymbol{\varphi}(t-q)\|^2,$$
$$r(0)=1. \tag{16}$$

2 受控自回归滑动平均系统

本节针对受控自回归滑动平均(CARMA)系统,讨论增广随机梯度(ESG)算法、多新息增广随机梯度(MI-ESG)算法、递推增广最小二乘(RELS)算法、多新息增广最小二乘(MI-ELS)算法、有限数据窗递推增广最小二乘(FDW-RELS)算法、有限数据窗多新息增广最小二乘(FDW-MI-ELS)算法等.

2.1 系统描述与辨识模型

考虑下列受控自回归滑动平均模型(Controlled AutoRegressive Moving Average model,CARMA 模型)描述的动态随机系统:

$$A(z)y(t)=B(z)u(t)+D(z)v(t), \tag{17}$$

其中 $\{u(t)\}$ 和 $\{y(t)\}$ 分别是系统的输入和输出序列,$\{v(t)\}$ 是零均值方差为 σ^2 的随机白噪声序列,$A(z)$,$B(z)$ 和 $D(z)$ 是单位后移算子 z^{-1} 的多项式:

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

设阶次 n_a,n_b 和 n_d 已知,记 $n:=n_a+n_b+n_d$,且 $t\leqslant 0$ 时,$y(t)=0$,$u(t)=0$,$v(t)=0$.

CARMA 模型又称为方程误差滑动平均模型(Equation-Error Moving Average model,EEMA 模型),也称为具有外部输入的 ARMA 模型(ARMA model with eXogenous input),即 ARMAX 模型.

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:

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

式(17)可等价写为

$$y(t)=\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t)=$$
$$\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{18}$$

这就是 CARMA 系统的辨识模型.

注 1 CARMA 系统的干扰 $w(t):=D(z)v(t)\in\mathbf{R}$ 是一个滑动平均(MA)过程.CARMA 系统辨识模型(18)的信息向量 $\boldsymbol{\varphi}(t)$ 不仅包含系统的可测输入输出数据 $u(t-i)$ 和 $y(t-i)$,还包含了不可测噪声项 $v(t-i)$,故式(18)是一个伪线性回归模型(pseudo-linear regressive model).信息向量中未知量的存在是辨识的困难所在.

注 2 CARMA 模型不仅包含了 CAR 部分的参数 a_i 和 b_i,还包含了 MA 噪声模型 $D(z)v(t)$ 的参数 d_i,在增广辨识方法(如增广随机梯度算法、递推增广最小二乘算法)中,将噪声模型参数 d_i 增加到参数向量中,这就是“增广辨识方法”名称的来历.记住:“增广”对应于模型中“MA 噪声过程”.

2.2 增广随机梯度算法

下面推导增广随机梯度算法.

设 $\hat{\boldsymbol{\theta}}(t)$ 是 $\boldsymbol{\theta}$ 在时刻 t 的估计,$\mu(t)\geqslant 0$ 是收敛因子(convergence factor)或步长 (step-size).对于辨识模型(18),定义准则函数

$$J_3(\boldsymbol{\theta}):=\frac{1}{2}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}]^2.$$

使用负梯度搜索,可以得到下列递推关系:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)-\mu(t)\,\mathrm{grad}[J_3(\hat{\boldsymbol{\theta}}(t-1))]=$$
$$\hat{\boldsymbol{\theta}}(t-1)+\mu(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)].$$

取 $\mu(t):=1/r(t)$,$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2$,则有

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad(19)$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1.\quad(20)$$

然而,因为信息向量 $\boldsymbol{\varphi}(t)$ 包含了不可测噪声项 $v(t-i)$,使得这个递推计算无法实现.解决的方法是使用其估计代替.

2.2.1 基于残差的增广随机梯度辨识算法

设 $\hat{v}(t)$ 是 $v(t)$ 的估计,根据 $\boldsymbol{\varphi}(t)$ 的定义,用系统的可测输入输出数据 $u(t-i)$ 和 $y(t-i)$,以及不可测噪声项 $v(t-i)$ 的估计 $\hat{v}(t-i)$ 定义 $\boldsymbol{\varphi}(t)$ 的估计:

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

从式(18)可得

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

上式中未知的 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{\theta}$ 分别用其估计 $\hat{\boldsymbol{\varphi}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$ 代替,则 $v(t)$ 的估计(即残差)可由下式计算:

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

因此参数向量 $\boldsymbol{\theta}$ 的辨识可以使用 $\hat{\boldsymbol{\varphi}}(t)$ 而非 $\boldsymbol{\varphi}(t)$ 进行计算.用 $\hat{\boldsymbol{\varphi}}(t)$ 代替式(19)—(20)中未知 $\boldsymbol{\varphi}(t)$,或使用负梯度搜索,极小化修正的准则函数

$$J_4(\boldsymbol{\theta}):=\frac{1}{2}[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{\theta}]^2,$$

可以得到辨识 CARMA 系统参数向量 $\boldsymbol{\theta}$ 的基于残差的增广随机梯度算法(Residual based Extended Stochastic Gradient algorithm,R-ESG 算法)[2,26]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t),\quad(21)$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\quad(22)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad(23)$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b),$$
$$\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\quad(24)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).\quad(25)$$

R-ESG 算法(21)—(25)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$,$r(0)=1$,$\hat{v}(t-i)=1/p_0(i=1,2,\cdots,n_d)$,$p_0=10^6$,$\mathbf{1}_n$ 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(24)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$.

3) 由式(22)计算新息 $e(t)$,由式(23)计算 $r(t)$.

4) 根据式(21)刷新参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.由式(25)计算残差 $\hat{v}(t)$.

5) t 增 1,转到第 2)步.

2.2.2 基于新息的增广随机梯度辨识算法

注意到式(24)估计的信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 中不可测噪声项是用残差 $\hat{v}(t-i)$ 代替的,故称为基于残差的增广随机梯度算法.如果 $\hat{\boldsymbol{\varphi}}(t)$ 中不可测噪声项用新息 $e(t-i)$ 代替,就得到估计 CARMA 系统参数向量 $\boldsymbol{\theta}$ 的基于新息的增广随机梯度算法(Innovation based Extended Stochastic Gradient algorithm, I-ESG 算法)[2]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t),\quad(26)$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\quad(27)$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\quad(28)$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b),$$
$$e(t-1),e(t-2),\cdots,e(t-n_d)]^{\mathrm{T}}.\quad(29)$$

I-ESG 算法(26)—(29)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$,$r(0)=1$,$e(t-i)=1/p_0(i=1,2,\cdots,n_d)$,$p_0=10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(29)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$.

3) 由式(27)计算新息 $e(t)$,由式(28)计算 $r(t)$.

4) 根据式(26)刷新参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.

5) t 增 1,转到第 2)步.

注 3 基于残差的增广随机梯度算法和基于新息的增广随机梯度算法统称为增广随机梯度算法(Extended Stochastic Gradient algorithm,ESG 算法).二者性能类似,其仿真结果的参数估计难以看出明显的差别.在参数估计时序上,由于残差中使用当前时刻的估计,所以 R-ESG 算法研究的更多.文献[2,

26]研究了基于残差的增广随机梯度辨识算法的收敛性,读者可以研究基于新息的增广随机梯度辨识算法的收敛性.文献[27]基于过参数化模型研究了Hammerstein非线性ARMAX系统的梯度迭代算法和增广随机梯度算法及其收敛性.

注4 为提高ESG算法的参数估计精度,可引入遗忘因子(forgetting factor)λ,将式(23)和(28)修改为

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

就得到遗忘因子增广随机梯度算法(Forgetting Factor ESG algorithm,FF-ESG算法).也可引入收敛指数(convergence index)ε,将式(21)和(26)修改为

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

就得到修正增广随机梯度算法(Modified ESG algorithm,M-ESG算法).这里$1/r^{\varepsilon}(t)$是收敛因子或步长.收敛指数ε的引入,使得算法的增益$\frac{\hat{\boldsymbol{\varphi}}(t)}{r^{\varepsilon}(t)}$增大,从而提高收敛速度[2,28-29].

注5 在CARMA系统中,噪声模型$w(t):=D(z)v(t)$是一个滑动平均过程(moving average process).辨识中将噪声模型的参数d_i增加到参数向量中,这就是"增广"二字的来历.如使用梯度搜索,推导的方法称为增广随机梯度辨识算法;如使用最小二乘原理,推导的方法就是递推增广最小二乘辨识算法;如使用多新息辨识理论和梯度搜索,推导的方法称为多新息增广随机梯度辨识算法;如使用递阶辨识原理和梯度搜索,推导的方法称为递阶增广随机梯度辨识算法;如使用递阶辨识原理和最小二乘原理,推导的方法称为递阶增广最小二乘辨识算法.

2.3 多新息增广随机梯度算法

为了改善ESG算法的收敛速度,借鉴多新息辨识理论[1,30-32],这里给出基于残差的多新息增广随机梯度算法和基于新息的多新息增广随机梯度算法.

2.3.1 基于残差的多新息增广随机梯度算法

设整数$p\geqslant 1$为新息长度.参照文献[1,30,33]中多新息随机梯度算法的推导,基于R-ESG算法(21)—(25),将输出$y(t)$和信息向量$\hat{\boldsymbol{\varphi}}(t)$扩展为堆积输出向量$\boldsymbol{Y}(p,t)$和堆积信息矩阵$\hat{\boldsymbol{\Phi}}(p,t)$:

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

将标量新息(scalar innovation)$e(t)\in\mathbf{R}$扩展为一个新息向量(innovation vector),即多新息(multi-innovation)

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

就可以得到估计CARMA系统参数向量$\boldsymbol{\theta}$的基于残差的多新息增广随机梯度算法(Residual based Multi-Innovation ESG algorithm, R-MI-ESG算法)[33-34]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t),\tag{30}$$
$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1),\tag{31}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2,\tag{32}$$
$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{33}$$
$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{34}$$
$$\begin{aligned}\hat{\boldsymbol{\varphi}}(t)=[&-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned}\tag{35}$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).\tag{36}$$

当新息长度$p=1$时,R-MI-ESG算法退化为R-ESG算法(21)—(25).

R-MI-ESG算法(30)—(36)随t增加,计算参数估计向量$\hat{\boldsymbol{\theta}}(t)$的步骤如下:

1) 初始化:令$t=1$,给定新息长度p.置初值$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$,$r(0)=1$,$\hat{v}(t-i)=1/p_0$($i=1,2,\cdots,n_d$),$p_0=10^6$.

2) 采集输入输出数据$u(t)$和$y(t)$,由式(35)构造信息向量$\hat{\boldsymbol{\varphi}}(t)$,由式(33)构造堆积输出向量$\boldsymbol{Y}(p,t)$,由式(34)构造堆积信息矩阵$\hat{\boldsymbol{\Phi}}(p,t)$.

3) 由式(31)计算新息向量$\boldsymbol{E}(p,t)$,由式(32)计算$r(t)$.

4) 根据式(30)刷新参数估计向量$\hat{\boldsymbol{\theta}}(t)$,由式(36)计算残差$\hat{v}(t)$.

5) t增1,转到第2)步.

2.3.2 基于新息的多新息增广随机梯度算法

基于I-ESG算法(26)—(29),将输出$y(t)$和信息向量$\hat{\boldsymbol{\varphi}}(t)$扩展为堆积输出向量$\boldsymbol{Y}(p,t)$和堆积信息矩阵$\hat{\boldsymbol{\Phi}}(p,t)$,将标量新息$e(t)\in\mathbf{R}$扩展为一个新息向量

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

就可以得到估计CARMA系统参数向量$\boldsymbol{\theta}$的基于新息的多新息增广随机梯度算法(Innovation based Multi-Innovation ESG algorithm,I-MI-ESG算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{37}$$

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

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2,\quad r(0)=1, \tag{39}$$

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

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{41}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b),\\&e(t-1),e(t-2),\cdots,e(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{42}$$

$$\begin{aligned}&e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\quad e(-i)=1/p_0,\\&i=0,1,\cdots,n_d.\end{aligned} \tag{43}$$

当新息长度 $p=1$ 时,I-MI-ESG 算法退化为 I-ESG 算法(26)—(29).

注 6 对于 R-MI-ESG 辨识算法,信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 在 $\hat{\boldsymbol{\Phi}}(p,t)$ 中要反复使用,考虑到参数估计的收敛性,期望残差 $\hat{v}(t)$ 逼近 $v(t)$,因此在每次递推过程中,可以用当前时刻最新的参数估计更新所有残差,这样式(36)和(35)可以修改为

$$\begin{aligned}&\hat{v}(j)=y(j)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\hat{\boldsymbol{\theta}}(t),\quad j=1,2,\cdots,t,\\&\hat{\boldsymbol{\varphi}}(j)=[-y(j-1),-y(j-2),\cdots,-y(j-n_a),\\&\qquad u(j-1),u(j-2),\cdots,u(j-n_b),\\&\qquad \hat{v}(j-1),\hat{v}(j-2),\cdots,\hat{v}(j-n_d)]^{\mathrm{T}}.\end{aligned}$$

如更新在算法中使用的残差,则上式可以修改为以残差向量的形式进行更新:

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

类似地,对于 I-MI-ESG 辨识算法,读者可考虑新息应该如何更新?

注 7 基于残差的多新息增广随机梯度算法和基于新息的多新息增广随机梯度算法统称为多新息增广随机梯度算法(Multi-Innovation ESG algorithm, MI-ESG 算法).文献[33]研究了基于残差的多新息增广随机梯度辨识算法的收敛性,读者可以研究基于新息的多新息增广随机梯度辨识算法的收敛性.

注 8 多新息增广随机梯度算法与多新息随机梯度算法性能相似,如既使用当前新息,又使用过去新息,还重复使用系统输入输出数据,来改进参数估计精度.特别是,引入新息长度能够提高算法的收敛速度.

注 9 对于参数向量维数高的问题,多新息增广随机梯度算法改进参数估计精度的能力是有限的.为了加快算法的收敛速度,式(32)和(39)可以修改为

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

可进一步引入遗忘因子 λ,即

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

就得到遗忘因子多新息增广随机梯度算法(Forgetting Factor MI-ESG algorithm, FF-MI-ESG 算法).当新息长度 $p=1$ 时,遗忘因子多新息增广随机梯度算法退化为遗忘因子增广随机梯度算法.

遗忘因子增广随机梯度算法、遗忘因子多新息增广随机梯度算法参数估计的有界收敛性是控制领域的研究难题.

2.4 递推增广最小二乘算法

2.4.1 递推增广最小二乘算法

对于 CARMA 系统辨识模型(18),定义最小二乘准则函数(least squares criterion function)

$$J_5(\boldsymbol{\theta}):=\sum_{j=1}^{t}[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2.$$

仿照文献[1-2]中 RLS 算法的推导,可以得到下列最小二乘递推关系:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{44}$$

$$\boldsymbol{L}(t)=\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}, \tag{45}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1). \tag{46}$$

信息向量 $\boldsymbol{\varphi}(t)$ 包含了不可测噪声项 $v(t-i)$,故递推关系(44)—(46)不可实现.解决的办法是未知向量 $\boldsymbol{\varphi}(t)$ 用其估计

$$\begin{aligned}\hat{\boldsymbol{\varphi}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^n\end{aligned}$$

代替,未知噪声项 $v(t)$ 用其估计(即残差)

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

代替,就可以得到估计 CARMA 系统参数向量 $\boldsymbol{\theta}$ 的递推增广最小二乘算法(Recursive Extended Least Squares algorithm, RELS 算法)[1]:

$$\begin{aligned}&\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\\&\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\end{aligned} \tag{47}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \tag{48}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{49}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{50}$$

$$\begin{aligned}&\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{v}(-i)=1/p_0,\\&i=0,1,\cdots,n_d.\end{aligned} \tag{51}$$

注 10 RELS 算法(47)—(51)比 R-ESG 算法(21)—(25)和 I-ESG 算法(26)—(29)有更快的收敛速度,但计算量也更大.

2.4.2 有限数据窗递推增广最小二乘算法

观看有限数据窗递推最小二乘算法(4)—(5)的构成,根据 RELS 算法(47)—(51),可以得到有限数据窗递推增广最小二乘算法(Recursive Extended Least Squares algorithm over a Finite Data Window, FDW-RELS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{52}$$

$$\boldsymbol{P}^{-1}(t)=\sum_{j=0}^{q}\hat{\boldsymbol{\varphi}}(t-j)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-j)=\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)-\hat{\boldsymbol{\varphi}}(t-q+1)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-q+1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{53}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{54}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{v}(-i)=1/p_0,\quad i=0,1,\cdots,n_d. \tag{55}$$

式中 $q>n$ 取为数据窗长度.

参考简化的有限数据窗递推最小二乘算法(8)—(11),FDW-RELS 算法可等价表示为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{56}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}_1(t)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}_1(t)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \tag{57}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}_1(t), \tag{58}$$

$$\boldsymbol{L}_1(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t-q)[1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-q)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t-q)]^{-1},\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{59}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_n+\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-q)]\boldsymbol{P}(t-1), \tag{60}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{61}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{v}(-i)=1/p_0,\quad i=0,1,\cdots,n_d. \tag{62}$$

注 11 有限数据窗最小二乘类辨识方法(如 FDW-RELS 算法)具有跟踪时变参数的能力.q 越小算法的跟踪能力越强,但参数估计波动也越大,所以应该动态改变数据窗长度或适当选择数据窗大小.参数不变时,数据窗应该大一些,否则小一点.FDW-RELS 算法、有限数据窗递推广义最小二乘算法、有限数据窗递推广义增广最小二乘算法的收敛性分析是控制领域的研究难题.

2.5 多新息增广最小二乘算法

2.5.1 多新息增广最小二乘算法

借助于多新息辨识理论,基于 RELS 算法(47)—(51),将系统输出 $y(t)$ 和信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$,将式(47)中标量新息 $e(t):=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}$ 扩展为新息向量(innovation vector)

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

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(Multi-Innovation Extended Least Squares algorithm,MI-ELS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{62}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t)\hat{\boldsymbol{\Phi}}(p,t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}, \tag{63}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)]\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{64}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{65}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{66}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{67}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{v}(-i)=1/p_0,\quad i=0,1,\cdots,n_d. \tag{68}$$

当新息长度 $p=1$ 时,MI-ELS 算法退化为 RELS 算法(47)—(51).

2.5.2 有限数据窗多新息增广最小二乘算法

基于 MI-ELS 算法(62)—(68),可以得到有限数据窗多新息增广最小二乘算法(Multi-Innovation Extended Least Squares algorithm over a Finite Data Window,FDW-MI-ELS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{69}$$

$$\boldsymbol{P}^{-1}(t)=\sum_{i=0}^{q}\hat{\boldsymbol{\Phi}}(p,t-i)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t-i)=\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\Phi}}(p,t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)-\hat{\boldsymbol{\Phi}}(p,t-q+1)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t-q+1), \tag{70}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{71}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)], \tag{72}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$

$$u(t-1),u(t-2),\cdots,u(t-n_b),$$
$$\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{73}$$
$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{v}(-i)=1/p_0,$$
$$i=0,1,\cdots,n_d. \tag{74}$$

3 基于模型分解的多新息辨识方法

本节针对 CARARMA 系统,基于辨识模型分解技术,推导基于分解的广义增广随机梯度(D-GESG)算法、基于分解的多新息广义增广随机梯度(D-MI-GESG)算法、基于分解的递推广义增广最小二乘(D-RGELS)算法、基于分解的多新息广义增广最小二乘(D-MI-GELS)算法,它们分别辨识系统模型的参数和噪声模型的参数,其特点是降低算法的计算量.

针对 CARARMA 系统,笔者提出了一种精确的两阶段最小二乘迭代算法[35],提出了基于分解的梯度迭代算法(即两阶段梯度迭代算法)、基于分解的最小二乘迭代算法(即两阶段最小二乘迭代算法),也称为递阶梯度迭代算法和递阶最小二乘迭代算法[36].

3.1 CARARMA 系统描述与分解辨识模型

考虑下列受控自回归自回归滑动平均模型(Controlled AutoRegressive AutoRegressive Moving Average model, CARARMA 模型)描述的动态随机系统:

$$A(z)y(t)=B(z)u(t)+\frac{D(z)}{C(z)}v(t), \tag{75}$$

其中$\{u(t)\}$和$\{y(t)\}$分别是系统的输入和输出序列,$\{v(t)\}$是零均值方差为σ^2的随机白噪声序列,$A(z),B(z),C(z)$和$D(z)$是后移算子z^{-1}的多项式:

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

设阶次n_a,n_b,n_c和n_d已知,记$n:=n_a+n_b+n_c+n_d$,且$t\leqslant 0$时,$y(t)=0,u(t)=0,v(t)=0$.

定义相关噪声项

$$w(t):=\frac{D(z)}{C(z)}v(t)\in\mathbf{R}, \tag{76}$$

定义系统模型的参数向量$\boldsymbol{\theta}_{\mathrm{s}}$和噪声模型的参数向量$\boldsymbol{\theta}_{\mathrm{n}}$,系统模型信息向量$\boldsymbol{\varphi}_{\mathrm{s}}(t)$和噪声模型信息向量$\boldsymbol{\varphi}_{\mathrm{n}}(t)$如下:

$$\boldsymbol{\theta}_{\mathrm{s}}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$
$$\boldsymbol{\theta}_{\mathrm{n}}:=[c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$
$$\boldsymbol{\varphi}_{\mathrm{n}}(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c),$$
$$v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}.$$

式(76)和(75)可分别写作

$$w(t)=[1-C(z)]w(t)+D(z)v(t)=$$
$$\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t), \tag{77}$$
$$y(t)=[1-A(z)]y(t)+B(z)u(t)+w(t)$$
$$=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+w(t) \tag{78}$$
$$=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t). \tag{79}$$

这个辨识模型包含了两个参数向量$\boldsymbol{\theta}_{\mathrm{s}}$和$\boldsymbol{\theta}_{\mathrm{n}}$,前者包含系统模型$G(z):=\dfrac{B(z)}{A(z)}$的参数,后者包含噪声模型$H(z):=\dfrac{D(z)}{C(z)}$的参数.$\boldsymbol{\varphi}_{\mathrm{s}}(t)$是由输出输入数据$y(t-i)$和$u(t-i)$构成的,故是已知的,$\boldsymbol{\varphi}_{\mathrm{n}}(t)$是由不可测噪声$w(t-i)$和$v(t-i)$构成的,故是未知的.

引入中间变量$y_1(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}$,式(79)可以等价写为两个子辨识模型:

$$y_1(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+v(t), \tag{80}$$
$$w(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}} \tag{81}$$
$$=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t). \tag{82}$$

这两个辨识模型的分解结构如图 1 所示.

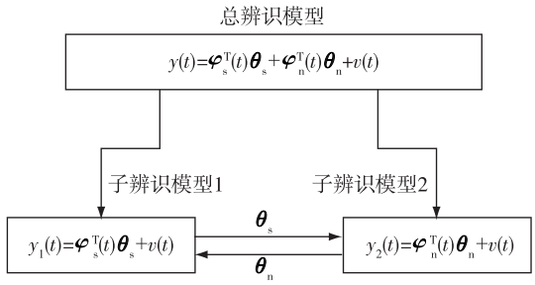


图 1 辨识模型分解为子辨识模型的递阶结构

Fig. 1 The hierarchical structure of the identification model

注 12 对辨识模型(79)进行分解,得到两个子辨识模型(80)和(82),一个包含系统模型的参数向量$\boldsymbol{\theta}_{\mathrm{s}}$和可测信息向量$\boldsymbol{\varphi}_{\mathrm{s}}(t)$,另一个包含噪声模型的参数向量$\boldsymbol{\theta}_{\mathrm{n}}$和不可测信息向量$\boldsymbol{\varphi}_{\mathrm{n}}(t)$,可以分别推导它们的辨识算法,并利用递阶辨识原理协调两个子辨识算法间的关联项.

3.2 基于分解的广义增广随机梯度算法

根据辨识模型(80)和(82),引入两个梯度准则

函数:

$$J_6(\boldsymbol{\theta}_s):=\frac{1}{2}[y_1(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\boldsymbol{\theta}_s]^2,$$

$$J_7(\boldsymbol{\theta}_n):=\frac{1}{2}[w(t)-\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n]^2.$$

读者可能会注意到这两个准则函数是相等的:$J_6(\boldsymbol{\theta}_s)=J_7(\boldsymbol{\theta}_n)$.这样做的目的是使读者易于理解:每个辨识模型(每个准则函数)辨识一个参数向量.

设 $\hat{\boldsymbol{\theta}}_s(t)$ 和 $\hat{\boldsymbol{\theta}}_n(t)$ 分别是 $\boldsymbol{\theta}_s$ 和 $\boldsymbol{\theta}_n$ 在时刻 t 的估计,$1/r_1(t)$ 和 $1/r_2(t)$ 为收敛因子(convergence factor)或步长(step-size).使用负梯度搜索,极小化 $J_6(\boldsymbol{\theta}_s)$ 和 $J_7(\boldsymbol{\theta}_n)$,可以得到下列递推关系:

$$\hat{\boldsymbol{\theta}}_s(t)=\hat{\boldsymbol{\theta}}_s(t-1)-\frac{1}{r_1(t)}\mathrm{grad}[J_6(\hat{\boldsymbol{\theta}}_s(t-1))]=$$

$$\hat{\boldsymbol{\theta}}_s(t-1)+\frac{\boldsymbol{\varphi}_s(t)}{r_1(t)}[y_1(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)]= \tag{83}$$

$$\hat{\boldsymbol{\theta}}_s(t-1)+\frac{\boldsymbol{\varphi}_s(t)}{r_1(t)}[y(t)-\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)]=$$

$$\hat{\boldsymbol{\theta}}_s(t-1)+\frac{\boldsymbol{\varphi}_s(t)}{r_1(t)}[y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)-\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n], \tag{84}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\varphi}_s(t)\|^2, \tag{85}$$

$$\hat{\boldsymbol{\theta}}_n(t)=\hat{\boldsymbol{\theta}}_n(t-1)-\frac{1}{r_2(t)}\mathrm{grad}[J_7(\hat{\boldsymbol{\theta}}_n(t-1))]=$$

$$\hat{\boldsymbol{\theta}}_n(t-1)+\frac{\boldsymbol{\varphi}_n(t)}{r_2(t)}[w(t)-\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t-1)]=$$

$$\hat{\boldsymbol{\theta}}_n(t-1)+\frac{\boldsymbol{\varphi}_n(t)}{r_2(t)}[y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\boldsymbol{\theta}_s-\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t-1)], \tag{86}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\varphi}_n(t)\|^2. \tag{87}$$

因为 $\boldsymbol{\varphi}_s(t)$ 是由系统可测输入输出数据 $y(t-i)$ 和 $u(t-i)$ 构成的,是已知的,而 $\boldsymbol{\varphi}_n(t)$ 是由不可测噪声项 $w(t-i)$ 和 $v(t-i)$ 构成的,是未知的.故关系式(84)—(87)右边包含了未知信息向量 $\boldsymbol{\varphi}_n(t)$,以及未知参数向量 $\boldsymbol{\theta}_n$ 和 $\boldsymbol{\theta}_s$,故无法实现.

本文的方法是利用辅助模型辨识思想,利用系统的可测信息(包括计算得到的信息)建立两个辅助模型,用辅助模型的输出 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 代替信息向量 $\boldsymbol{\varphi}_n(t)$ 中的未知项,从而实现参数估计的递推计算.具体细节如下.

用辅助模型的输出 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$(即 $w(t-i)$ 和 $v(t-i)$ 的估计)定义 $\boldsymbol{\varphi}_n(t)$ 的估计:

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

根据式(81),未知 $\boldsymbol{\theta}_s$ 用其估计 $\hat{\boldsymbol{\theta}}_s(t)$ 代替,便得到计算 $\hat{w}(t)$ 的辅助模型

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

由式(79)可得

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

上式中未知的 $\boldsymbol{\varphi}_n(t)$,$\boldsymbol{\theta}_s$ 和 $\boldsymbol{\theta}_n$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_n(t)$,$\hat{\boldsymbol{\theta}}_s(t)$ 和 $\hat{\boldsymbol{\theta}}_n(t)$ 代替,就得到计算 $\hat{v}(t)$ 的辅助模型

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t)-\hat{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t)=$$
$$\hat{w}(t)-\hat{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t).$$

上式也可以直接根据式(77),未知量 $w(t)$,$\boldsymbol{\varphi}_n(t)$ 和 $\boldsymbol{\theta}_n$ 分别用其估计 $\hat{w}(t)$,$\hat{\boldsymbol{\varphi}}_n(t)$ 和 $\hat{\boldsymbol{\theta}}_n(t)$ 代替得到.

因此,式(84)—(87)右边未知的 $\boldsymbol{\varphi}_n(t)$,$\boldsymbol{\theta}_n$ 和 $\boldsymbol{\theta}_s$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_n(t)$,$\hat{\boldsymbol{\theta}}_n(t-1)$ 和 $\hat{\boldsymbol{\theta}}_s(t-1)$ 代替,并在 $r_1(t)$ 和 $r_2(t)$ 中引入遗忘因子 λ(当然可以采用两个不同的遗忘因子),定义新息

$$e(t):=y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)-\hat{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t-1)\in\mathbf{R},$$

就可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_s$ 和 $\boldsymbol{\theta}_n$ 的基于分解的广义增广随机梯度算法(Decomposition based GESG algorithm,D-GESG 算法):

$$\hat{\boldsymbol{\theta}}_s(t)=\hat{\boldsymbol{\theta}}_s(t-1)+\frac{\boldsymbol{\varphi}_s(t)}{r_1(t)}e(t), \tag{88}$$

$$e(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)-\hat{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t-1), \tag{89}$$

$$r_1(t)=\lambda r_1(t-1)+\|\boldsymbol{\varphi}_s(t)\|^2,\quad 0\leqslant\lambda\leqslant 1, \tag{90}$$

$$\hat{\boldsymbol{\theta}}_n(t)=\hat{\boldsymbol{\theta}}_n(t-1)+\frac{\hat{\boldsymbol{\varphi}}_n(t)}{r_2(t)}e(t), \tag{91}$$

$$r_2(t)=\lambda r_2(t-1)+\|\hat{\boldsymbol{\varphi}}_n(t)\|^2, \tag{92}$$

$$\boldsymbol{\varphi}_s(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{93}$$

$$\hat{\boldsymbol{\varphi}}_n(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),$$
$$\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{94}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t), \tag{95}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_n(t), \tag{96}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_s(t)\\ \hat{\boldsymbol{\theta}}_n(t)\end{bmatrix}. \tag{97}$$

D-GESG 算法(88)—(97)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_s(t)$ 和 $\hat{\boldsymbol{\theta}}_n(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置初值 $\hat{\boldsymbol{\theta}}_s(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_n(0)=\mathbf{1}_{n_c+n_d}/p_0$,$r_1(0)=1$,$r_2(0)=1$,$\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$i=1,2,\cdots,\max[n_c,n_d]$,$p_0=10^6$.给定遗忘因子 λ.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式

(93)—(94)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$.

3) 由式(89)计算新息 $e(t)$,由式(90)计算 $r_1(t)$.

4) 根据式(88)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

5) 由式(92)计算 $r_2(t)$.

6) 根据式(91)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

7) 由式(95)—(96)计算辅助模型的输出 $\hat{w}(t)$ 和 $\hat{v}(t)$.

8) t 增 1,转到第 2)步.

注 13 递阶辨识原理在辨识中使用了分解和协调原理,能够处理结构复杂的线性或非线性系统、多变量系统的辨识问题.这里的基于分解的辨识方法(如两阶段辨识方法、三阶段辨识方法、多阶段辨识方法等)也可称为递阶辨识方法.因此,有下列等价名称:

1) 基于分解的广义增广随机梯度算法也可称为递阶广义增广随机梯度算法(H-GESG 算法);

2) 基于分解的多新息广义增广随机梯度算法也可称为递阶多新息广义增广随机梯度算法(HMI-GESG 算法);

3) 基于分解的递推广义增广最小二乘算法可称为递阶广义增广最小二乘算法(HGELS 算法)(因为这里递阶含有递推的意思,故省略"递推"二字);

4) 基于分解的多新息广义增广最小二乘算法可称为递阶多新息广义增广最小二乘算法(HMI-GELS 算法).

3.3 基于分解的多新息广义增广随机梯度算法

为改进基于分解的广义增广随机梯度辨识算法的收敛速度,下面借助多新息辨识理论,通过扩展新息维数,导出基于分解的多新息随机梯度算法.

设正整数 p 表示新息长度.基于 D-GESG 算法(88)—(97),将系统输出 $y(t)$,信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$:

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

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t):=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)]\in\mathbf{R}^{(n_a+n_b)\times p},$$

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

将式(88)和(91)中标量新息

$$e(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$$

扩展为新息向量

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

$$\begin{bmatrix}y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\\y(t-1)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-2)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-2)\\\vdots\\y(t-p+1)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-p)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-p)\end{bmatrix}\in\mathbf{R}^{p}.$$

通常可认为时刻 $(t-1)$ 的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$ 比时刻 $(t-i)$ $(i\geqslant 2)$ 的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-i)$ 更接近真值 $\boldsymbol{\theta}_{\mathrm{s}}$,估计 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$ 比时刻 $(t-i)$ $(i>2)$ 的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-i)$ 更接近真值 $\boldsymbol{\theta}_{\mathrm{n}}$.因此,将 $\boldsymbol{E}(p,t)$ 表达式中 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-i)$ $(i\geqslant 2)$ 换为 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$,将 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-i)$ $(i\geqslant 2)$ 换为 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$,多新息向量可合理修改为

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

$$\begin{bmatrix}y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\\y(t-1)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\\\vdots\\y(t-p+1)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\end{bmatrix}=$$

$$\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}^{p}.$$

注意到 $\boldsymbol{Y}(1,t)=y(t)$,$\boldsymbol{E}(1,t)=e(t)$,$\boldsymbol{\Phi}_{\mathrm{s}}(1,t)=\boldsymbol{\varphi}_{\mathrm{s}}(t)$,$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)=\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,式(88)和(91)可以等价表达为

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\boldsymbol{\Phi}_{\mathrm{s}}(1,t)}{r_1(t)}\boldsymbol{E}(1,t),$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)}{r_2(t)}\boldsymbol{E}(1,t).$$

这是新息长度 $p=1$ 的基于分解的"多新息"随机梯度算法.将两式中 $\boldsymbol{\Phi}_{\mathrm{s}}(1,t)$,$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)$ 和 $\boldsymbol{E}(1,t)$ 里的"1"换为 p,就得到新息长度为 p 的基于分解的多新息广义增广随机梯度算法:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\boldsymbol{\Phi}_{\mathrm{s}}(p,t)}{r_1(t)}\boldsymbol{E}(p,t),$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)}{r_2(t)}\boldsymbol{E}(p,t).$$

于是,可以总结出新息长度为 p 的,估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于分解的多新息广义增广随机梯度算法(Decomposition based MI-GESG algorithm,D-MI-GESG 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\boldsymbol{\Phi}_{\mathrm{s}}(p,t)}{r_1(t)}\boldsymbol{E}(p,t),\tag{98}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\tag{99}$$

$$r_1(t)=\lambda r_1(t-1)+\|\boldsymbol{\varphi}_{\mathrm{s}}(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\tag{100}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)}{r_2(t)}\boldsymbol{E}(p,t), \tag{101}$$

$$r_2(t)=\lambda r_2(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\|^2, \tag{102}$$

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

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t)=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)], \tag{104}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-p+1)], \tag{105}$$

$$\begin{aligned}\boldsymbol{\varphi}_{\mathrm{s}}(t)=[&-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\end{aligned} \tag{106}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[&-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{107}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{108}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{109}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{110}$$

这个算法中引入了遗忘因子,全称应该是基于分解的遗忘因子 MI-GESG 算法(D-FF-MI-GESG 算法).当新息长度 $p=1$ 时,D-MI-GESG 算法退化为 D-GESG 算法(88)—(87).

D-MI-GESG 算法(98)—(110)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p.置 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\boldsymbol{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\boldsymbol{1}_{n_c+n_d}/p_0$,$r_1(0)=1$,$r_2(0)=1$,$\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$i=1,2,\cdots,\max[n_c,n_d]$,$p_0=10^6$.给定遗忘因子 λ.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(106)—(107)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,由式(103)—(105)构造堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$.

3) 由式(99)计算 $\boldsymbol{E}(p,t)$,由式(100)计算 $r_1(t)$.

4) 根据式(98)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

5) 由式(102)计算 $r_2(t)$.

6) 根据式(101)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

7) 由式(108)—(109)计算辅助模型的输出 $\hat{w}(t)$ 和 $\hat{v}(t)$.

8) t 增 1,转到第 2)步.

3.4 基于分解的递推广义增广最小二乘算法

根据 CARARMA 系统辨识模型(80)和(82),引入两个最小二乘准则函数:

$$J_8(\boldsymbol{\theta}_{\mathrm{s}}):=\sum_{j=1}^{t}[y_1(j)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(j)\boldsymbol{\theta}_{\mathrm{s}}]^2,$$

$$J_9(\boldsymbol{\theta}_{\mathrm{n}}):=\sum_{j=1}^{t}[w(j)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(j)\boldsymbol{\theta}_{\mathrm{n}}]^2.$$

仿照 RLS 算法的推导,极小化 $J_8(\boldsymbol{\theta}_{\mathrm{s}})$ 和 $J_9(\boldsymbol{\theta}_{\mathrm{n}})$,可以得到下列最小二乘递推关系:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)[y_1(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)]= \tag{111}$$

$$\begin{aligned}&\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)]=\\&\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\\&\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}],\end{aligned} \tag{112}$$

$$\boldsymbol{L}_1(t)=\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_{\mathrm{s}}(t)[1+\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_{\mathrm{s}}(t)]^{-1}, \tag{113}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1), \tag{114}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_2(t)[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)]= \tag{115}$$

$$\begin{aligned}&\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_2(t)[y(t)-\\&\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)],\end{aligned}$$

$$\boldsymbol{L}_2(t)=\boldsymbol{P}_2(t-1)\boldsymbol{\varphi}_{\mathrm{n}}(t)[1+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_2(t-1)\boldsymbol{\varphi}_{\mathrm{n}}(t)]^{-1}, \tag{116}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_2(t)\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_2(t-1). \tag{117}$$

注意到信息向量 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 是未知的,故用辅助模型的输出 $\hat{w}(t)$ 和 $\hat{v}(t)$(即 $w(t)$ 和 $v(t)$ 的估计)定义 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 的估计:

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

根据式(81),未知 $\boldsymbol{\theta}_{\mathrm{s}}$ 用其估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 代替,便得到计算 $\hat{w}(t)$ 的辅助模型

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t).$$

由式(79)可得

$$v(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}},$$

上式中未知的 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$,$\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 代替,就得到计算 $\hat{v}(t)$ 的辅助模型

$$\begin{aligned}\hat{v}(t)=&y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\\&\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t).\end{aligned}$$

于是,式(112)—(117)中右边的未知向量 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$,$\boldsymbol{\theta}_{\mathrm{n}}$ 和 $\boldsymbol{\theta}_{\mathrm{s}}$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$ 代替,定义新息

$$e(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R},$$

可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于分解的递推广义增广最小二乘算法(Decomposition based RGELS algorithm,D-RGELS 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)e(t), \tag{118}$$

$$e(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1), \tag{119}$$

$$\boldsymbol{L}_1(t)=\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_{\mathrm{s}}(t)[1+\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_{\mathrm{s}}(t)]^{-1}, \tag{120}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1), \tag{121}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_2(t)e(t),\tag{122}$$

$$\boldsymbol{L}_2(t)=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)[1+\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)]^{-1},\tag{123}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_2(t)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_2(t-1),\tag{124}$$

$$\begin{aligned}\boldsymbol{\varphi}_{\mathrm{s}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\end{aligned}\tag{125}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=&[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned}\tag{126}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t),\tag{127}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\tag{128}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{129}$$

D-RGELS 算法(118)—(129)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置初值 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\boldsymbol{1}_{n_a+n_b}/p_0$, $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\boldsymbol{1}_{n_c+n_d}/p_0$, $\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$, $\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{n_c+n_d}$, $\hat{w}(t-i)=1/p_0$, $\hat{v}(t-i)=1/p_0$, $i=1,2,\cdots,\max[n_c,n_d]$, $p_0=10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(125)—(126)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$.

3) 由式(119)计算新息 $e(t)$,由式(120)计算增益向量 $\boldsymbol{L}_1(t)$,由式(121)计算协方差阵 $\boldsymbol{P}_1(t)$.

4) 根据式(118)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

5) 由式(123)计算增益向量 $\boldsymbol{L}_2(t)$,由式(124)计算协方差阵 $\boldsymbol{P}_2(t)$.

6) 根据式(122)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

7) 由式(127)—(128)计算辅助模型的输出 $\hat{w}(t)$ 和 $\hat{v}(t)$.

8) t 增 1,转到第 2)步.

表 1 给出了 D-RGELS 算法每递推计算一步的计算量($n=n_a+n_b+n_c+n_d$).

3.5 基于分解的多新息广义增广最小二乘算法

基于 D-RGELS 算法(118)—(129),将输出 $y(t)$,信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$:

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

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t):=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)]\in\mathbf{R}^{(n_a+n_b)\times p},$$

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

将式(118)和(122)中标量新息

$$e(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$$

扩展为新息向量(innovation vector)

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

可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于分解的多新息广义增广最小二乘算法(Decomposition based MI-GELS algorithm,D-MI-GELS 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)\boldsymbol{E}(p,t),\tag{130}$$

$$\begin{aligned}\boldsymbol{E}(p,t)=&\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\\&\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\end{aligned}\tag{131}$$

$$\begin{aligned}\boldsymbol{L}_1(t)=&\boldsymbol{P}_1(t-1)\boldsymbol{\Phi}_{\mathrm{s}}(p,t)[\boldsymbol{I}_p+\\&\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)\boldsymbol{P}_1(t-1)\boldsymbol{\Phi}_{\mathrm{s}}(p,t)]^{-1},\end{aligned}\tag{132}$$

表 1 D-RGELS 算法每步的计算量

Table 1 The computational efficiency of the D-RGELS algorithm

变量	表达式	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$	$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_1(t)e(t)\in\mathbf{R}^{n_a+n_b}$	n_a+n_b	n_a+n_b
	$e(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$	n	n
$\boldsymbol{L}_1(t)$	$\boldsymbol{L}_1(t)=\boldsymbol{\zeta}_1(t)/[1+\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\zeta}_1(t)]\in\mathbf{R}^{n_a+n_b}$	$2(n_a+n_b)$	n_a+n_b
	$\boldsymbol{\zeta}_1(t):=\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_{\mathrm{s}}(t)\in\mathbf{R}^{n_a+n_b}$	$(n_a+n_b)^2$	$(n_a+n_b)(n_a+n_b-1)$
$\boldsymbol{P}_1(t)$	$\boldsymbol{P}_1(t)=\boldsymbol{P}_1(t-1)-\boldsymbol{L}_1(t)\boldsymbol{\zeta}_1^{\mathrm{T}}(t)\in\mathbf{R}^{(n_a+n_b)\times(n_a+n_b)}$	$(n_a+n_b)^2$	$(n_a+n_b)^2$
$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$	$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_2(t)e(t)\in\mathbf{R}^{n_c+n_d}$	n_c+n_d	n_c+n_d
$\boldsymbol{L}_2(t)$	$\boldsymbol{L}_2(t)=\boldsymbol{\zeta}_2(t)/[1+\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\zeta}_2(t)]\in\mathbf{R}^{n_c+n_d}$	$2(n_c+n_d)$	n_c+n_d
	$\boldsymbol{\zeta}_2(t):=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\in\mathbf{R}^{n_c+n_d}$	$(n_c+n_d)^2$	$(n_c+n_d)(n_c+n_d-1)$
$\boldsymbol{P}_2(t)$	$\boldsymbol{P}_2(t)=\boldsymbol{P}_2(t-1)-\boldsymbol{L}_2(t)\boldsymbol{\zeta}_2^{\mathrm{T}}(t)\in\mathbf{R}^{(n_c+n_d)\times(n_c+n_d)}$	$(n_c+n_d)^2$	$(n_c+n_d)^2$
$\hat{w}(t)$	$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\in\mathbf{R}$	n_a+n_b	n_a+n_b
$\hat{v}(t)$	$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\in\mathbf{R}$	n_c+n_d	n_c+n_d
总数		$2(n_a+n_b)^2+2(n_c+n_d)^2+5n$	$2(n_a+n_b)^2+2(n_c+n_d)^2+3n$
总 flop 数		$N_2:=4(n_a+n_b)^2+4(n_c+n_d)^2+8n$	

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)]\boldsymbol{P}_1(t-1), \tag{133}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_2(t)\boldsymbol{E}(p,t), \tag{134}$$

$$\boldsymbol{L}_2(t)=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)]^{-1}, \tag{135}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_2(t)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)]\boldsymbol{P}_2(t-1), \tag{136}$$

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

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t)=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)], \tag{138}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-p+1)], \tag{139}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{140}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{141}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{142}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{143}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{144}$$

当新息长度 $p=1$ 时,D-MI-GELS 算法退化为 D-RGELS 算法(118)—(129).

D-MI-GELS 算法(130)—(144)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p. 置 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\boldsymbol{1}_{n_a+n_b}/p_0$, $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\boldsymbol{1}_{n_c+n_d}/p_0$, $\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$, $\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{n_c+n_d}$, $\hat{w}(t-i)=1/p_0$, $\hat{v}(t-i)=1/p_0$, $i=1,2,\cdots,\max[n_c,n_d]$, $p_0=10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(140)—(141)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,由式(137)—(139)构造堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$.

3) 由式(131)计算新息向量 $\boldsymbol{E}(p,t)$,由式(132)—(133)计算增益向量 $\boldsymbol{L}_1(t)$ 和协方差阵 $\boldsymbol{P}_1(t)$.

4) 根据式(130)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

5) 由式(135)计算增益向量 $\boldsymbol{L}_2(t)$,由式(136)计算协方差阵 $\boldsymbol{P}_2(t)$.

6) 根据式(134)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

7) 由式(142)—(143)计算辅助模型的输出 $\hat{w}(t)$ 和 $\hat{v}(t)$.

8) t 增 1,转到第 2)步.

4 基于数据滤波的多新息辨识方法

基于数据滤波的辨识方法一般是针对有色噪声干扰的系统而言的.其基本思想是,在系统方程两边同乘以一个滤波器(通常取为噪声模型传递函数的逆),对输入输出数据进行滤波,滤波后的系统是一个白噪声干扰的系统,而原系统的传递函数不变,然后使用滤波后的输入输出数据等进行辨识.因为滤波器(噪声模型的传递函数)是未知的,滤波后的输入输出数据也是未知的,实际中通常采用噪声模型传递函数的估计进行滤波,所以基于数据滤波的辨识方法只能以递推方式或迭代方式实现.

如果知道系统先验知识,滤波器也可以采用参数已知的固定模型,否则可采用动态变化模型,如即将讨论的用噪声传递函数的估计作为滤波器.

辨识中数据滤波与通信和信号处理中的滤波有本质不同.后者滤波是剔除被污染信号中的噪声,使得噪声在信号中的比重降低,如低通滤波器和高通滤波器,而滤波辨识方法只是改变系统模型的结构,不改变系统输入-输出间关系.总之,滤波辨识是将有色噪声干扰的系统模型变换为白噪声干扰的模型,即模型白色化(不是数据白色化),它能够提高参数辨识精度.

滤波辨识方法(filtering based identification method)是笔者 1988 年在清华大学攻读硕士学位时,受萧德云老师讲授的《过程辨识》[37]中“CARAR 系统的广义最小二乘方法”(即基于数据滤波的 CARAR 系统最小二乘方法)的启发,经过多年思考和发展,提炼形成的一个辨识分支——基于数据滤波的辨识方法.目前基于滤波的辨识方法在线性系统和非线性系统中都有广泛应用,如:

1) 基于数据滤波和辅助模型的 OEAR 系统递推最小二乘辨识方法[38];

2) 输出误差滑动平均(OEMA)系统的基于数据滤波的辅助模型递推最小二乘辨识方法和基于数据滤波的辅助模型最小二乘迭代辨识方法[39];

3) 基于数据滤波的线性参数自回归滑动平均系统的最小二乘迭代辨识方法[40];

4) 基于数据滤波的 Hammerstein CARAR 系统的多新息随机梯度辨识方法[41];

5) 基于数据滤波的输入非线性输出误差自回归系统的多新息随机梯度辨识方法[42];

6) 基于数据滤波的 Hammerstein CARMA 系统的多新息随机梯度辨识方法[43];

7) Hammerstein OEAR 系统的基于分解的多新息随机梯度辨识方法和基于数据滤波的多新息随机

梯度辨识方法[44].

本节讨论基于数据滤波的 CARARMA 系统辨识方法,包括基于数据滤波的广义增广随机梯度(F-GESG)辨识方法、基于数据滤波的多新息广义增广随机梯度(F-MI-GESG)辨识方法、基于数据滤波的递推广义增广最小二乘(F-RGELS)辨识方法、基于数据滤波的多新息广义增广最小二乘(F-MI-GELS)辨识方法.其中基于数据滤波的递推广义增广最小二乘辨识方法是笔者先写入尚未出版的《系统辨识方法论》一书中,后授权于访问学者王冬青教授补充成稿,发表在《Digital Signal Processing》2010 年第 4 期上[45].

4.1 CARARMA 系统描述与滤波辨识模型

考虑上节的受控自回归自回归滑动平均模型(Controlled AutoRegressive AutoRegressive Moving Average model, CARARMA 模型)描述的动态随机系统:

$$A(z)y(t)=B(z)u(t)+\frac{D(z)}{C(z)}v(t), \tag{145}$$

其中$\{u(t)\}$和$\{y(t)\}$分别是系统的输入和输出序列,$\{v(t)\}$是零均值方差为σ^2的随机白噪声序列,$A(z),B(z),C(z)$和$D(z)$是后移算子z^{-1}的多项式:

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

设阶次n_a,n_b,n_c和n_d已知,记$n:=n_a+n_b+n_c+n_d$,且$t\leqslant 0$时,$y(t)=0,u(t)=0,v(t)=0$.

定义相关噪声项

$$w(t):=\frac{D(z)}{C(z)}v(t)\in\mathbf{R}. \tag{146}$$

定义系统模型的参数向量$\boldsymbol{\theta},\boldsymbol{\theta}_{\mathrm{s}}$和$\boldsymbol{\theta}_{\mathrm{n}}$,信息向量$\boldsymbol{\varphi}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t)$和$\boldsymbol{\varphi}_{\mathrm{n}}(t)$如下:

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^n,$$
$$\boldsymbol{\theta}_{\mathrm{s}}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$
$$\boldsymbol{\theta}_{\mathrm{n}}:=[c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$
$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \boldsymbol{\varphi}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^n,$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},$$
$$\boldsymbol{\varphi}_{\mathrm{n}}(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c),$$
$$v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}.$$

式(146)和(145)可写作

$$w(t)=[1-C(z)]w(t)+D(z)v(t)=$$
$$\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t), \tag{147}$$
$$y(t)=[1-A(z)]y(t)+B(z)u(t)+w(t)$$
$$=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+w(t) \tag{148}$$
$$=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t) \tag{149}$$
$$=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{150}$$

取滤波器$L(z):=\dfrac{C(z)}{D(z)}$,它是噪声模型传递函数的倒数.定义滤波输入$u_{\mathrm{f}}(t)$、滤波输出$y_{\mathrm{f}}(t)$、滤波信息向量$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$分别为

$$u_{\mathrm{f}}(t):=L(z)u(t)=\frac{C(z)}{D(z)}u(t),$$
$$y_{\mathrm{f}}(t):=L(z)y(t)=\frac{C(z)}{D(z)}y(t),$$
$$\boldsymbol{\varphi}_{\mathrm{sf}}(t):=L(z)\boldsymbol{\varphi}_{\mathrm{s}}(t)=\frac{C(z)}{D(z)}\boldsymbol{\varphi}_{\mathrm{s}}(t)=$$
$$[-y_{\mathrm{f}}(t-1),-y_{\mathrm{f}}(t-2),\cdots,-y_{\mathrm{f}}(t-n_a),$$
$$u_{\mathrm{f}}(t-1),u_{\mathrm{f}}(t-2),\cdots,u_{\mathrm{f}}(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b}.$$

式(145)两边乘以$L(z)=\dfrac{C(z)}{D(z)}$得到

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

或

$$A(z)y_{\mathrm{f}}(t)=B(z)u_{\mathrm{f}}(t)+v(t).$$

这个滤波后的模型是一个方程误差模型(即 CAR 模型).将它写为向量形式为

$$y_{\mathrm{f}}(t)=[1-A(z)]y_{\mathrm{f}}(t)+B(z)u_{\mathrm{f}}(t)+v(t)=$$
$$\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+v(t). \tag{151}$$

注 14 对于两个辨识模型(151)和(147),由于多项式$C(z)$和$D(z)$是未知的(即$L(z)$是未知的),是要辨识的,故$u_{\mathrm{f}}(t)$和$y_{\mathrm{f}}(t)$是未知的,滤波信息向量$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$是未知的;由噪声$w(t-i)$和$v(t-i)$构成的信息向量$\boldsymbol{\varphi}_{\mathrm{n}}(t)$也是未知的,故需要采用估计值代替未知变量来推导基于数据滤波的辨识方法.

4.2 基于滤波的广义增广随机梯度算法

对于 CARMAMA 系统的辨识模型(151)和(147),定义两个梯度准则函数:

$$J_{10}(\boldsymbol{\theta}_{\mathrm{s}}):=\frac{1}{2}[y_{\mathrm{f}}(t)-\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}]^2,$$

$$J_{11}(\boldsymbol{\theta}_{\mathrm{n}}):=\frac{1}{2}[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}]^2.$$

这两个准则函数是相等的:$J_{10}(\boldsymbol{\theta}_{\mathrm{s}})=J_{11}(\boldsymbol{\theta}_{\mathrm{n}})$.即每个

准则函数推导一个参数向量的辨识算法.

设 $\hat{\boldsymbol{\theta}}(t):=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}$,$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 分别是 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 在时刻 t 的估计,$1/r_{\mathrm{s}}(t)$ 和 $1/r_{\mathrm{n}}(t)$ 为收敛因子或步长.使用负梯度搜索,极小化 $J_{10}(\boldsymbol{\theta}_{\mathrm{s}})$ 和 $J_{11}(\boldsymbol{\theta}_{\mathrm{n}})$,可以得到下列递推关系:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\frac{1}{r_{\mathrm{s}}(t)}\mathrm{grad}[J_{10}(\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1))]=$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\boldsymbol{\varphi}_{\mathrm{sf}}(t)}{r_{\mathrm{s}}(t)}[y_{\mathrm{f}}(t)-\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)],\tag{152}$$

$$r_{\mathrm{s}}(t)=r_{\mathrm{s}}(t-1)+\|\boldsymbol{\varphi}_{\mathrm{sf}}(t)\|^{2},\tag{153}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)-\frac{1}{r_{\mathrm{n}}(t)}\mathrm{grad}[J_{11}(\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1))]=$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\boldsymbol{\varphi}_{\mathrm{n}}(t)}{r_{\mathrm{n}}(t)}[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)],\tag{154}$$

$$r_{\mathrm{n}}(t)=r_{\mathrm{n}}(t-1)+\|\boldsymbol{\varphi}_{\mathrm{n}}(t)\|^{2}.\tag{155}$$

这些递推关系无法实现,因为式(152)—(155)右边 $y_{\mathrm{f}}(t)$,$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$,$w(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 都是未知的,所以在推导辨识算法前,需要先构造它们的估计.

用 $w(t-i)$ 和 $v(t-i)$ 的估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 构造 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 的估计:

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

用 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 定义 $\boldsymbol{\varphi}(t)$ 的估计:

$$\hat{\boldsymbol{\varphi}}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n}.$$

由式(148)可得

$$w(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}},$$

用估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$ 代替上式右边未知的 $\boldsymbol{\theta}_{\mathrm{s}}$,可计算出 $w(t)$ 的估计:

$$\hat{w}(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1).$$

由式(150)可得

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

用 $\hat{\boldsymbol{\varphi}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$ 分别代替上式中 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{\theta}$,可计算出 $v(t)$ 的估计:

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

$$y(t)-[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}=$$

$$y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t).$$

由噪声模型参数估计

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),$$
$$\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}$$

构造 $C(z)$ 和 $D(z)$ 的估计:

$$\hat{C}(t,z):=1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c},$$
$$\hat{D}(t,z):=1+\hat{d}_1(t)z^{-1}+\hat{d}_2(t)z^{-2}+\cdots+\hat{d}_{n_d}(t)z^{-n_d}.$$

用动态估计的滤波器 $\hat{L}(t,z):=\dfrac{\hat{C}(t,z)}{\hat{D}(t,z)}$ 对 $u(t)$ 和 $y(t)$ 进行滤波,分别得到 $u_{\mathrm{f}}(t)$ 和 $y_{\mathrm{f}}(t)$ 的估计:

$$\hat{u}_{\mathrm{f}}(t):=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}u(t),$$

$$\hat{y}_{\mathrm{f}}(t):=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}y(t),$$

或

$$\hat{D}(t,z)\hat{u}_{\mathrm{f}}(t)=\hat{C}(t,z)u(t),$$

$$\hat{D}(t,z)\hat{y}_{\mathrm{f}}(t)=\hat{C}(t,z)y(t).$$

$\hat{u}_{\mathrm{f}}(t)$ 和 $\hat{y}_{\mathrm{f}}(t)$ 也可按照下列递推式计算:

$$\hat{u}_{\mathrm{f}}(t)=[1-\hat{D}(t,z)]\hat{u}_{\mathrm{f}}(t)+\hat{C}(t,z)u(t)=$$
$$[-\hat{d}_1(t)z^{-1}-\hat{d}_2(t)z^{-2}-\cdots-\hat{d}_{n_d}(t)z^{-n_d}]\hat{u}_{\mathrm{f}}(t)+$$
$$[1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c}]u(t)=$$
$$-\hat{d}_1(t)\hat{u}_{\mathrm{f}}(t-1)-\hat{d}_2(t)\hat{u}_{\mathrm{f}}(t-2)-\cdots-$$
$$\hat{d}_{n_d}(t)\hat{u}_{\mathrm{f}}(t-n_d)+u(t)+\hat{c}_1(t)u(t-1)+$$
$$\hat{c}_2(t)u(t-2)+\cdots+\hat{c}_{n_c}(t)u(t-n_c)=$$
$$u(t)+[u(t-1),u(t-2),\cdots,u(t-n_c)]\hat{\boldsymbol{c}}(t)+$$
$$[\hat{u}_{\mathrm{f}}(t-1),\hat{u}_{\mathrm{f}}(t-2),\cdots,\hat{u}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t),$$

$$\hat{y}_{\mathrm{f}}(t)=[1-\hat{D}(t,z)]\hat{y}_{\mathrm{f}}(t)+\hat{C}(t,z)y(t)=$$
$$[-\hat{d}_1(t)z^{-1}-\hat{d}_2(t)z^{-2}-\cdots-\hat{d}_{n_d}(t)z^{-n_d}]\hat{y}_{\mathrm{f}}(t)+$$
$$[1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c}]y(t)=$$
$$-\hat{d}_1(t)\hat{y}_{\mathrm{f}}(t-1)-\hat{d}_2(t)\hat{y}_{\mathrm{f}}(t-2)-\cdots-$$
$$\hat{d}_{n_d}(t)\hat{y}_{\mathrm{f}}(t-n_d)+y(t)+\hat{c}_1(t)y(t-1)+$$
$$\hat{c}_2(t)y(t-2)+\cdots+\hat{c}_{n_c}(t)y(t-n_c)=$$
$$y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+$$
$$[\hat{y}_{\mathrm{f}}(t-1),\hat{y}_{\mathrm{f}}(t-2),\cdots,\hat{y}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t).$$

用 $\hat{y}_{\mathrm{f}}(t-i)$ 和 $\hat{u}_{\mathrm{f}}(t-i)$ 构造 $\boldsymbol{\varphi}_{\mathrm{sf}}(t)$ 的估计:

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

于是,式(152)—(155)右边未知的 $y_{\mathrm{f}}(t)$,$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$,$w(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 分别用其估计 $\hat{y}_{\mathrm{f}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$,$\hat{w}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 代替,可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的广义增广随机梯度算法(Filtering based GESG algorithm,F-GESG 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)}{r_{\mathrm{s}}(t)}[\hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)],\tag{156}$$

$$r_{\mathrm{s}}(t)=r_{\mathrm{s}}(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)\|^2,\tag{157}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)}{r_{\mathrm{n}}(t)}[\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)],\tag{158}$$

$$r_{\mathrm{n}}(t)=r_{\mathrm{n}}(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\|^2,\tag{159}$$

$$\hat{\boldsymbol{\varphi}}(t)=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix},\tag{160}$$

$$\begin{aligned}\boldsymbol{\varphi}_{\mathrm{s}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\end{aligned}\tag{161}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=&[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned}\tag{162}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)=&[-\hat{y}_{\mathrm{f}}(t-1),-\hat{y}_{\mathrm{f}}(t-2),\cdots,-\hat{y}_{\mathrm{f}}(t-n_a),\\&\hat{u}_{\mathrm{f}}(t-1),\hat{u}_{\mathrm{f}}(t-2),\cdots,\hat{u}_{\mathrm{f}}(t-n_b)]^{\mathrm{T}},\end{aligned}\tag{163}$$

$$\begin{aligned}\hat{u}_{\mathrm{f}}(t)=&-\hat{d}_1(t)\hat{u}_{\mathrm{f}}(t-1)-\hat{d}_2(t)\hat{u}_{\mathrm{f}}(t-2)-\cdots-\\&\hat{d}_{n_d}(t)\hat{u}_{\mathrm{f}}(t-n_d)+u(t)+\hat{c}_1(t)u(t-1)+\\&\hat{c}_2(t)u(t-2)+\cdots+\hat{c}_{n_c}(t)u(t-n_c),\end{aligned}\tag{164}$$

$$\begin{aligned}\hat{y}_{\mathrm{f}}(t)=&-\hat{d}_1(t)\hat{y}_{\mathrm{f}}(t-1)-\hat{d}_2(t)\hat{y}_{\mathrm{f}}(t-2)-\cdots-\\&\hat{d}_{n_d}(t)\hat{y}_{\mathrm{f}}(t-n_d)+y(t)+\hat{c}_1(t)y(t-1)+\\&\hat{c}_2(t)y(t-2)+\cdots+\hat{c}_{n_c}(t)y(t-n_c),\end{aligned}\tag{165}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1),\tag{166}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{167}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=&[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),\hat{d}_1(t),\\&\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}},\end{aligned}\tag{168}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{169}$$

F-GESG 算法(156)—(169)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置参数估计初值 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\mathbf{1}_{n_c+n_d}/p_0$,$r_{\mathrm{s}}(0)=1$,$r_{\mathrm{n}}(0)=1$,$\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$\hat{y}_{\mathrm{f}}(t-i)=1/p_0$,$\hat{u}_{\mathrm{f}}(t-i)=1/p_0$,$i=1,2,\cdots,\max[n_a,n_b,n_c,n_d]$,$p_0=10^6$.给定小正数 ε.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(161)—(163)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$,由式(160)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$.

3) 由式(166)计算 $\hat{w}(t)$,由式(159)计算新息 $r_{\mathrm{n}}(t)$.

4) 根据式(158)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

5) 从式(168)$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 中读取 $\hat{c}_i(t)$ 和 $\hat{d}_i(t)$,由式(164)—(165)计算 $\hat{u}_{\mathrm{f}}(t)$ 和 $\hat{y}_{\mathrm{f}}(t)$.

6) 由式(157)计算 $r_{\mathrm{s}}(t)$.

7) 根据式(156)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

8) 由式(167)计算 $\hat{v}(t)$.

9) 将 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 与 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$ 进行比较,将 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 与 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$ 进行比较,如果它们充分接近,满足

$$\|\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\|<\varepsilon \text{ 和 } \|\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)-\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\|<\varepsilon,$$

则终止递推计算过程,得到满意精度的参数估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$;否则 t 增 1 转到第 2)步,进行递推计算.

4.3 基于滤波的多新息广义增广随机梯度算法

为改进基于 F-GESG 算法的收敛速度,下面借助于多新息辨识理论,通过扩展新息维数,导出基于滤波的多新息广义增广随机梯度算法.

设正整数 p 表示新息长度.基于 F-GESG 算法(156)—(169),将滤波输出 $\hat{y}_{\mathrm{f}}(t)$,滤波信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$,噪声估计 $\hat{w}(t)$ 和噪声信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 扩展为堆积滤波输出向量 $\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)$,堆积滤波信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)$,堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$ 和堆积噪声信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$:

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

$$\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t):=[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-p+1)]\in\mathbf{R}^{(n_a+n_b)\times p},$$

$$\hat{\boldsymbol{W}}(p,t):=[\hat{w}(t),\hat{w}(t-1),\cdots,\hat{w}(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^p,$$

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

将式(156)和(158)中标量新息

$$e_{\mathrm{s}}(t):=\hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\in\mathbf{R},$$

$$e_{\mathrm{n}}(t):=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$$

扩展为新息向量

$$\boldsymbol{E}_{\mathrm{s}}(p,t):=\begin{bmatrix}e_{\mathrm{s}}(t)\\ e_{\mathrm{s}}(t-1)\\ \vdots\\ e_{\mathrm{s}}(t-p+1)\end{bmatrix}=$$

$$\begin{bmatrix}\hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\\ \hat{y}_{\mathrm{f}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-2)\\ \vdots\\ \hat{y}_{\mathrm{f}}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-p)\end{bmatrix}\in\mathbf{R}^p,$$

$$\boldsymbol{E}_{\mathrm{n}}(p,t):=\begin{bmatrix}e_{\mathrm{n}}(t)\\ e_{\mathrm{n}}(t-1)\\ \vdots\\ e_{\mathrm{n}}(t-p+1)\end{bmatrix}=$$

$$\begin{bmatrix} \hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1) \\ \hat{w}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-2) \\ \vdots \\ \hat{w}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-p) \end{bmatrix} \in \mathbf{R}^{p}.$$

通常可认为时刻$(t-1)$的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$比时刻$(t-i)$ $(i\geqslant 2)$的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-i)$更接近真值 $\boldsymbol{\theta}_{\mathrm{s}}$,估计$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$比时刻$(t-i)$ $(i\geqslant 2)$的估计 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-i)$更接近真值 $\boldsymbol{\theta}_{\mathrm{n}}$.因此,将 $\boldsymbol{E}_{\mathrm{s}}(p,t)$表达式中 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-i)$ $(i\geqslant 2)$换为 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$,将 $\boldsymbol{E}_{\mathrm{n}}(p,t)$表达式中 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-i)$ $(i\geqslant 2)$换为 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$,多新息向量可合理修改为

$$\boldsymbol{E}_{\mathrm{s}}(p,t):=\begin{bmatrix} \hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1) \\ \hat{y}_{\mathrm{f}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1) \\ \vdots \\ \hat{y}_{\mathrm{f}}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1) \end{bmatrix}=$$

$$\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{f}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\in\mathbf{R}^{p},$$

$$\boldsymbol{E}_{\mathrm{n}}(p,t):=\begin{bmatrix} \hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1) \\ \hat{w}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1) \\ \vdots \\ \hat{w}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1) \end{bmatrix}=$$

$$\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}^{p}.$$

注意到

$$\hat{\boldsymbol{Y}}_{\mathrm{f}}(1,t)=\hat{y}_{\mathrm{f}}(t),\boldsymbol{E}_{\mathrm{s}}(1,t)=e_{\mathrm{s}}(t),$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(1,t)=\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t),\hat{\boldsymbol{W}}(1,t)=\hat{w}(t),$$

$$\boldsymbol{E}_{\mathrm{n}}(1,t)=e_{\mathrm{n}}(t),\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)=\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),$$

式(156)和(158)可以分别等价表达为

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(1,t)}{r_{\mathrm{s}}(t)}\boldsymbol{E}_{\mathrm{s}}(1,t),$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)}{r_{\mathrm{n}}(t)}\boldsymbol{E}_{\mathrm{n}}(1,t).$$

这是新息长度 $p=1$ 的基于滤波的“多新息”随机梯度算法.将这两式中 $\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(1,t)$,和 $\boldsymbol{E}_{\mathrm{s}}(1,t)$,$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(1,t)$和 $\boldsymbol{E}_{\mathrm{n}}(1,t)$里的“1”换为 p,就得到新息长度为 p 的基于滤波的多新息随机梯度算法:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)}{r_{\mathrm{s}}(t)}\boldsymbol{E}_{\mathrm{s}}(p,t),$$

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

于是,可以总结出新息长度为 p 的,估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的多新息广义增广随机梯度算法(Filtering based Multi-Innovation GESG algorithm,F-MI-GESG 算法)的基本方程:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)}{r_{\mathrm{s}}(t)}\boldsymbol{E}_{\mathrm{s}}(p,t),\tag{170}$$

$$\boldsymbol{E}_{\mathrm{s}}(p,t)=\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{f}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1),\tag{171}$$

$$r_{\mathrm{s}}(t)=r_{\mathrm{s}}(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)\|^{2},\tag{172}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)}{r_{\mathrm{n}}(t)}\boldsymbol{E}_{\mathrm{n}}(p,t),\tag{173}$$

$$\boldsymbol{E}_{\mathrm{n}}(p,t)=\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\tag{174}$$

$$r_{\mathrm{n}}(t)=r_{\mathrm{n}}(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\|^{2},\tag{175}$$

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

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

$$\hat{\boldsymbol{W}}(p,t)=[\hat{w}(t),\hat{w}(t-1),\cdots,\hat{w}(t-p+1)]^{\mathrm{T}},\tag{178}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-p+1)],\tag{179}$$

$$\begin{aligned}\boldsymbol{\varphi}_{\mathrm{s}}(t)=&[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\\&u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\end{aligned}\tag{180}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=&[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned}\tag{181}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)=&[-\hat{y}_{\mathrm{f}}(t-1),-\hat{y}_{\mathrm{f}}(t-2),\cdots,-\hat{y}_{\mathrm{f}}(t-n_a),\\&\hat{u}_{\mathrm{f}}(t-1),\hat{u}_{\mathrm{f}}(t-2),\cdots,\hat{u}_{\mathrm{f}}(t-n_b)]^{\mathrm{T}},\end{aligned}\tag{182}$$

$$\begin{aligned}\hat{u}_{\mathrm{f}}(t)=&u(t)+[u(t-1),u(t-2),\cdots,u(t-n_c)]\hat{\boldsymbol{c}}(t)+\\&[\hat{u}_{\mathrm{f}}(t-1),\hat{u}_{\mathrm{f}}(t-2),\cdots,\hat{u}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t),\end{aligned}\tag{183}$$

$$\begin{aligned}\hat{y}_{\mathrm{f}}(t)=&y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+\\&[\hat{y}_{\mathrm{f}}(t-1),\hat{y}_{\mathrm{f}}(t-2),\cdots,\hat{y}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t),\end{aligned}\tag{184}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1),\tag{185}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\tag{186}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix},\tag{187}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{188}$$

F-MI-GESG 算法(170)—(188)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$的步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p.置参数估计初值 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\mathbf{1}_{n_c+n_d}/p_0$,$r_{\mathrm{s}}(0)=1$,$r_{\mathrm{n}}(0)=1$,$\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$\hat{y}_{\mathrm{f}}(t-i)=1/p_0$,$\hat{u}_{\mathrm{f}}(t-i)=1/p_0$,$i=1,2,\cdots,\max[n_a,n_b,n_c,n_d]$,$p_0=10^6$.给定小正数 ε.

2) 采集输入输出数据 $u(t)$和 $y(t)$,由式(180)—(182)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$和 $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$,

3) 由式(185)计算 $\hat{w}(t)$,由式(178)—(179)构

造堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$ 和堆积噪声信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$.

4) 由式(174)计算新息向量 $\boldsymbol{E}_{\mathrm{n}}(p,t)$,由式(175)计算 $r_{\mathrm{n}}(t)$.

5) 根据式(173)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

6) 从式(187)中 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 读取 $\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$,由式(183)—(184)计算 $\hat{u}_{\mathrm{f}}(t)$ 和 $\hat{y}_{\mathrm{f}}(t)$.

7) 由式(176)—(177)构造堆积滤波输出向量 $\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)$ 和堆积滤波信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)$.

8) 由式(171)计算新息向量 $\boldsymbol{E}_{\mathrm{s}}(p,t)$,由式(172)计算 $r_{\mathrm{s}}(t)$.

9) 根据式(170)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

10) 由式(186)计算 $\hat{v}(t)$.

11) 如果参数估计差满足

$\|\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\|<\varepsilon$ 和 $\|\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)-\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\|<\varepsilon$,则终止递推计算过程,得到满意精度的参数估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$;否则 t 增 1 转到第 2)步,进行递推计算.

4.4 基于滤波的递推广义增广最小二乘算法

对于 CARMAMA 系统的辨识模型(151)和(147),定义两个最小二乘准则函数:

$$J_{12}(\boldsymbol{\theta}_{\mathrm{s}}):=\sum_{j=1}^{t}[y_{\mathrm{f}}(j)-\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(j)\boldsymbol{\theta}_{\mathrm{s}}]^2,$$

$$J_{13}(\boldsymbol{\theta}_{\mathrm{n}}):=\sum_{j=1}^{t}[w(j)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(j)\boldsymbol{\theta}_{\mathrm{n}}]^2.$$

仿照 RLS 算法的推导,极小化 $J_{12}(\boldsymbol{\theta}_{\mathrm{s}})$ 和 $J_{13}(\boldsymbol{\theta}_{\mathrm{n}})$,可以得到下列最小二乘递推关系:

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_{\mathrm{s}}(t)[y_{\mathrm{f}}(t)-\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)],\quad(189)$$

$$\boldsymbol{L}_{\mathrm{s}}(t)=\boldsymbol{P}_{\mathrm{s}}(t-1)\boldsymbol{\varphi}_{\mathrm{sf}}(t)[1+\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{s}}(t-1)\boldsymbol{\varphi}_{\mathrm{sf}}(t)]^{-1},\quad(190)$$

$$\boldsymbol{P}_{\mathrm{s}}(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_{\mathrm{s}}(t)\boldsymbol{\varphi}_{\mathrm{sf}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{s}}(t-1),\quad(191)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)]=$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)[y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)],\quad(192)$$

$$\boldsymbol{L}_{\mathrm{n}}(t)=\boldsymbol{P}_{\mathrm{n}}(t-1)\boldsymbol{\varphi}_{\mathrm{n}}(t)[1+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{n}}(t-1)\boldsymbol{\varphi}_{\mathrm{n}}(t)]^{-1},\quad(193)$$

$$\boldsymbol{P}_{\mathrm{n}}(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_{\mathrm{n}}(t)\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{n}}(t-1).\quad(194)$$

这些递推关系无法实现,因为式(189)—(194)右边的 $y_{\mathrm{f}}(t)$,$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$,$\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 都是未知的,同样在辨识算法中使用它们的估计代替.

用 $w(t-i)$ 和 $v(t-i)$ 的估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 构造 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 的估计:

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

由式(148)可得

$$w(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}},$$

用估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 代替上式右边未知的 $\boldsymbol{\theta}_{\mathrm{s}}$,可计算出 $w(t)$ 的估计:

$$\hat{w}(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t).$$

由式(149)可得

$$v(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}},$$

上式中未知量 $\boldsymbol{\theta}_{\mathrm{s}}$,$\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 分别用其估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 代替,可计算出 $v(t)$ 的估计:

$$\hat{v}(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t).$$

由噪声模型参数估计

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),$$
$$\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}$$

构造 $C(z)$ 和 $D(z)$ 的估计:

$$\hat{C}(t,z):=1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c},$$

$$\hat{D}(t,z):=1+\hat{d}_1(t)z^{-1}+\hat{d}_2(t)z^{-2}+\cdots+\hat{d}_{n_d}(t)z^{-n_d}.$$

用动态估计的滤波器 $\hat{L}(t,z):=\dfrac{\hat{C}(t,z)}{\hat{D}(t,z)}$ 对 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 进行滤波,得到 $\boldsymbol{\varphi}_{\mathrm{sf}}(t)$ 的估计:

$$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t):=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}\boldsymbol{\varphi}_{\mathrm{s}}(t).$$

$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$ 也可按照下列递推式计算:

$$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)=[1-\hat{D}(t,z)]\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)+\hat{C}(t,z)\boldsymbol{\varphi}_{\mathrm{s}}(t)=$$
$$[-\hat{d}_1(t)z^{-1}-\hat{d}_2(t)z^{-2}-\cdots-\hat{d}_{n_d}(t)z^{-n_d}]\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)+$$
$$[1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c}]\boldsymbol{\varphi}_{\mathrm{s}}(t)=$$
$$-\hat{d}_1(t)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1)-\hat{d}_2(t)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-2)-\cdots-\hat{d}_{n_d}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-n_d)+$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)+\hat{c}_1(t)\boldsymbol{\varphi}_{\mathrm{s}}(t-1)+\hat{c}_2(t)\boldsymbol{\varphi}_{\mathrm{s}}(t-2)+\cdots+\hat{c}_{n_c}(t)\boldsymbol{\varphi}_{\mathrm{s}}(t-n_c)=$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t)+[\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\boldsymbol{\varphi}_{\mathrm{s}}(t-2),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-n_c)]\hat{\boldsymbol{c}}(t)+$$
$$[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-2),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-n_d)]\hat{\boldsymbol{d}}(t).$$

于是,式(189)—(194)右边未知的 $y_{\mathrm{f}}(t)$,$\boldsymbol{\varphi}_{\mathrm{sf}}(t)$,$\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 分别用其估计 $\hat{y}_{\mathrm{f}}(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$,$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 代替,就可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的递推广义增广最小二乘算法(Filtering based RGELS algorithm,F-RGELS 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_{\mathrm{s}}(t)[\hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)],\quad(195)$$

$$\boldsymbol{L}_{\mathrm{s}}(t)=\boldsymbol{P}_{\mathrm{s}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)[1+\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{s}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)]^{-1},\quad(196)$$

$$\boldsymbol{P}_{\mathrm{s}}(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_{\mathrm{s}}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{s}}(t-1),\quad(197)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)[y(t)-$$
$$\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)],\quad(198)$$

$$\boldsymbol{L}_{\mathrm{n}}(t)=\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)[1+\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)]^{-1},\quad(199)$$

$$\boldsymbol{P}_{\mathrm{n}}(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_{\mathrm{n}}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{n}}(t-1), \tag{200}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{201}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{202}$$

$$\hat{y}_{\mathrm{f}}(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[\hat{y}_{\mathrm{f}}(t-1),\hat{y}_{\mathrm{f}}(t-2),\cdots,\hat{y}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{203}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)=\boldsymbol{\varphi}_{\mathrm{s}}(t)+[\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\boldsymbol{\varphi}_{\mathrm{s}}(t-2),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-n_c)]\hat{\boldsymbol{c}}(t)+[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-2),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{204}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{205}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{206}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}, \tag{207}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{208}$$

F-RGELS 算法(195)—(208)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$. 置参数估计初值 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\boldsymbol{1}_{n_a+n_b}/p_0$, $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\boldsymbol{1}_{n_c+n_d}/p_0$, 和协方差阵初值 $\boldsymbol{P}_{\mathrm{s}}(0)=p_0\boldsymbol{I}_{n_a+n_b}$, $\boldsymbol{P}_{\mathrm{n}}(0)=p_0\boldsymbol{I}_{n_c+n_d}$. 置 $\hat{w}(t-i)=1/p_0$, $\hat{v}(t-i)=1/p_0$, $\hat{y}_{\mathrm{f}}(t-i)=1/p_0$, $\hat{u}_{\mathrm{f}}(t-i)=1/p_0$, $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-i)=1_{n_a+n_b}/p_0$, $i=1,2,\cdots,\max[n_a,n_b,n_c,n_d]$, $p_0=10^6$. 给定小正数 ε.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 由式(201)—(202)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,

3) 由式(199)—(200)计算增益向量 $\boldsymbol{L}_{\mathrm{n}}(t)$ 和协方差阵 $\boldsymbol{P}_{\mathrm{n}}(t)$.

4) 根据式(198)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

5) 从式(207)中 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 读取 $\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$, 由式(203)—(204)计算滤波输出 $\hat{y}_{\mathrm{f}}(t)$ 和滤波信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$.

6)由式(196)—(197)计算增益向量 $\boldsymbol{L}_{\mathrm{s}}(t)$ 和协方差阵 $\boldsymbol{P}_{\mathrm{s}}(t)$.

7) 根据式(195)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

8) 由式(205)—(206)计算 $\hat{w}(t)$ 和 $\hat{v}(t)$.

9) 如果参数估计差满足 $\|\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\theta}}(t-1)\|<\varepsilon$, 则终止递推计算过程,得到满意精度的参数估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$;否则 t 增 1 转到第 2)步,进行递推计算.

4.5 基于滤波的多新息广义增广最小二乘算法

基于 F-RGELS 算法(195)—(208),将滤波输出 $\hat{y}_{\mathrm{f}}(t)$、滤波信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$、系统输出 $y(t)$、信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和噪声信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 分别扩展为堆积滤波输出向量 $\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)$、堆积滤波信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)$、堆积输出向量 $\boldsymbol{Y}(p,t)$、堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和堆积噪声信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$:

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

$$\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-p+1)]\in\mathbf{R}^{(n_a+n_b)\times p},$$

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

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t):=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)]\in\mathbf{R}^{(n_a+n_b)\times p},$$

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

将式(195)和(198)中标量新息

$$e_{\mathrm{s}}(t):=\hat{y}_{\mathrm{f}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)\in\mathbf{R},$$

$$e_{\mathrm{n}}(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$$

扩展为新息向量(innovation vector)

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

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

可以得到估计 CARARMA 系统参数向量 $\boldsymbol{\theta}_{\mathrm{s}}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的多新息广义增广最小二乘算法(Filtering based MI-GELS algorithm, F-MI-GELS 算法):

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)+\boldsymbol{L}_{\mathrm{s}}(t)\boldsymbol{E}_{\mathrm{s}}(p,t), \tag{209}$$

$$\boldsymbol{E}_{\mathrm{s}}(p,t)=\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{f}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1), \tag{210}$$

$$\boldsymbol{L}_{\mathrm{s}}(t)=\boldsymbol{P}_{\mathrm{s}}(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{s}}(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(t)]^{-1}, \tag{211}$$

$$\boldsymbol{P}_{\mathrm{s}}(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_{\mathrm{s}}(t)\hat{\boldsymbol{\Phi}}_{\mathrm{f}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{s}}(t-1), \tag{212}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)\boldsymbol{E}_{\mathrm{n}}(p,t), \tag{213}$$

$$\boldsymbol{E}_{\mathrm{n}}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_{\mathrm{s}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1), \tag{214}$$

$$\boldsymbol{L}_{\mathrm{n}}(t)=\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(t)]^{-1}, \tag{215}$$

$$\boldsymbol{P}_{\mathrm{n}}(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_{\mathrm{n}}(t)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{n}}(t-1), \tag{216}$$

$$\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)=[\hat{y}_{\mathrm{f}}(t),\hat{y}_{\mathrm{f}}(t-1),\cdots,\hat{y}_{\mathrm{f}}(t-p+1)]^{\mathrm{T}}, \tag{217}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-p+1)], \tag{218}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}, \tag{219}$$

$$\boldsymbol{\Phi}_{\mathrm{s}}(p,t)=[\boldsymbol{\varphi}_{\mathrm{s}}(t),\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-p+1)], \tag{220}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-p+1)], \tag{221}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a),\ u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{222}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{223}$$

$$\hat{y}_{\mathrm{f}}(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+$$

$$[\hat{y}_{\mathrm{f}}(t-1),\hat{y}_{\mathrm{f}}(t-2),\cdots,\hat{y}_{\mathrm{f}}(t-n_d)]\hat{\boldsymbol{d}}(t), \quad (224)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)=\boldsymbol{\varphi}_{\mathrm{s}}(t)+[\boldsymbol{\varphi}_{\mathrm{s}}(t-1),\boldsymbol{\varphi}_{\mathrm{s}}(t-2),\cdots,\boldsymbol{\varphi}_{\mathrm{s}}(t-n_c)]\hat{\boldsymbol{c}}(t)+$$
$$[\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-1),\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-2),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-n_d)]\hat{\boldsymbol{d}}(t), \quad (225)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \quad (226)$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \quad (227)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}, \quad (228)$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \quad (229)$$

F-RGELS 算法(209)—(229)随 t 增加,计算参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 的步骤如下:

1) 初始化:令 $t=1$.置参数估计初值 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\mathbf{1}_{n_c+n_d}/p_0$,和协方差阵初值 $\boldsymbol{P}_{\mathrm{s}}(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_{\mathrm{n}}(0)=p_0\boldsymbol{I}_{n_c+n_d}$.置 $\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$\hat{y}_{\mathrm{f}}(t-i)=1/p_0$,$\hat{u}_f(t-i)=1/p_0$,$\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t-i)=1_{n_a+n_b}/p_0$,$i=1,2,\cdots,\max[n_a,n_b,n_c,n_d]$,$p_0=10^6$.给定小正数 ε.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,由式(222)—(223)构造信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$.

3) 由式(219)—(221)构造堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_{\mathrm{s}}(p,t)$ 和堆积噪声信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$.

4) 由式(214)计算新息向量 $\boldsymbol{E}_{\mathrm{n}}(p,t)$,由式(215)—(216)计算增益向量 $\boldsymbol{L}_{\mathrm{n}}(t)$ 和协方差阵 $\boldsymbol{P}_{\mathrm{n}}(t)$.

5) 根据式(213)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

6) 从式(228)中 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 读取 $\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$,由式(224)—(225)计算滤波输出 $\hat{y}_{\mathrm{f}}(t)$ 和滤波信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{sf}}(t)$.

7) 由式(217)—(218)构造堆积滤波输出向量 $\hat{\boldsymbol{Y}}_{\mathrm{f}}(p,t)$ 和堆积滤波信息矩阵 $\hat{\boldsymbol{\Phi}}_{\mathrm{f}}(p,t)$.

8) 由式(210)计算新息向量 $\boldsymbol{E}_{\mathrm{s}}(p,t)$,由式(211)—(212)计算增益向量 $\boldsymbol{L}_{\mathrm{s}}(t)$ 和协方差阵 $\boldsymbol{P}_{\mathrm{s}}(t)$.

9) 根据式(209)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$.

10) 由式(226)—(227)计算 $\hat{w}(t)$ 和 $\hat{v}(t)$.

11) 如果参数估计差满足 $\|\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\theta}}(t-1)\|<\varepsilon$,则终止递推计算过程,得到满意精度的参数估计 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$;否则 t 增 1 转到第 2)步,进行递推计算.

5 结语

本文讨论了方程误差类系统中 CARMA 模型的增广随机梯度(ESG)算法、多新息增广随机梯度(MI-ESG)算法、递推增广最小二乘(RELS)算法、多新息增广最小二乘(MI-ELS)算法.针对 CARARMA 系统,研究了基于模型分解的广义增广随机梯度(D-GESG)算法、基于模型分解的多新息广义增广随机梯度(D-MI-GESG)算法、基于模型分解的递推广义增广最小二乘(D-RGELS)算法和基于模型分解的多新息广义增广最小二乘(D-MI-GELS)算法,以及基于滤波的广义增广随机梯度(F-GESG)算法、基于滤波的多新息广义增广随机梯度(F-MI-GESG)算法、基于滤波的递推广义增广最小二乘(F-RGELS)算法和基于滤波的多新息广义增广最小二乘(F-MI-GELS)算法.这些方法可以推广到下列多输入单输出方程误差类系统、多输入多输出方程误差类系统、多变量方程误差类系统、多元方程误差类系统,以及非线性方程误差类系统[1-2].

1) 研究下列多输入方程误差系统(多输入 CAR 系统)的多输入多新息随机梯度辨识算法:

$$A(z)y(t)=B_1(z)u_1(t)+B_2(z)u_2(t)+\cdots+B_r(z)u_r(t)+v(t).$$

2) 研究下列多输入 CARMA 模型(多输入 ARMAX 模型)的多输入多新息增广随机梯度辨识算法:

$$A(z)y(t)=\sum_{j=1}^{r}B_j(z)u_j(t)+D(z)v(t).$$

3) 研究下列多输入 CARAR 模型(多输入 ARARX 模型)的多输入多新息广义随机梯度辨识算法:

$$A(z)y(t)=\sum_{j=1}^{r}B_j(z)u_j(t)+\frac{1}{C(z)}v(t).$$

4) 研究下列多输入 CARARMA 模型(多输入 ARARMAX 模型)的多输入多新息广义增广随机梯度辨识算法:

$$A(z)y(t)=\sum_{j=1}^{r}B_j(z)u_j(t)+\frac{D(z)}{C(z)}v(t).$$

5) 研究下列多变量方程误差 ARMA 系统(即多变量 CARARMA 系统)的多变量多新息广义增广随机梯度辨识算法:

$$\boldsymbol{A}(z)\boldsymbol{y}(t)=\boldsymbol{B}(z)\boldsymbol{u}(t)+\boldsymbol{C}^{-1}(z)\boldsymbol{D}(z)\boldsymbol{v}(t),$$

其中 $\boldsymbol{u}(t)=[u_1(t),u_2(t),\cdots,u_r(t)]^{\mathrm{T}}\in\mathbf{R}^r$ 为输入向量,$\boldsymbol{y}(t)=[y_1(t),y_2(t),\cdots,y_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为输出

向量,$\boldsymbol{v}(t)=[v_1(t),v_2(t),\cdots,v_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为零均值白噪声向量,$\boldsymbol{A}(z)$,$\boldsymbol{B}(z)$,$\boldsymbol{C}(z)$和$\boldsymbol{D}(z)$是单位后移算子z^{-1}的多项式矩阵:

$$\boldsymbol{A}(z):=\boldsymbol{I}+\boldsymbol{A}_1z^{-1}+\boldsymbol{A}_2z^{-2}+\cdots+\boldsymbol{A}_{n_a}z^{-n_a},\boldsymbol{A}_i\in\mathbf{R}^{m\times m},$$

$$\boldsymbol{B}(z):=\boldsymbol{B}_1z^{-1}+\boldsymbol{B}_2z^{-2}+\cdots+\boldsymbol{B}_{n_b}z^{-n_b},\boldsymbol{B}_i\in\mathbf{R}^{m\times r},$$

$$\boldsymbol{C}(z):=\boldsymbol{I}+\boldsymbol{C}_1z^{-1}+\boldsymbol{C}_2z^{-2}+\cdots+\boldsymbol{C}_{n_c}z^{-n_c},\boldsymbol{C}_i\in\mathbf{R}^{m\times m},$$

$$\boldsymbol{D}(z):=\boldsymbol{I}+\boldsymbol{D}_1z^{-1}+\boldsymbol{D}_2z^{-2}+\cdots+\boldsymbol{D}_{n_d}z^{-n_d},\boldsymbol{D}_i\in\mathbf{R}^{m\times m}.$$

对于这类多变量系统,笔者提出了递推广义增广最小二乘算法、有限数据长度下梯度迭代算法和最小二乘迭代算法,论文“Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems(多变量 CARARMA 系统的梯度迭代算法和最小二乘迭代估计算法)”[46]入选“2012 年中国百篇最具影响国际学术论文”.

6) 对于多变量 CARARMA 系统:

$$\boldsymbol{A}(z)\boldsymbol{y}(t)=\boldsymbol{B}(z)\boldsymbol{u}(t)+\boldsymbol{C}^{-1}(z)\boldsymbol{D}(z)\boldsymbol{v}(t).$$

基于分解技术,研究基于分解的多新息广义增广随机梯度辨识算法和基于分解的递推广义增广最小二乘辨识算法.

7) 研究下列多元伪线性回归 ARMA 模型描述的动态随机系统的多元多新息广义增广随机梯度辨识算法和基于滤波的多元多新息广义增广随机梯度辨识算法:

① $\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\theta}+\boldsymbol{C}^{-1}(z)\boldsymbol{D}(z)\boldsymbol{v}(t)$,

② $A(z)\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\theta}+\boldsymbol{C}^{-1}(z)\boldsymbol{D}(z)\boldsymbol{v}(t)$,

其中 $\boldsymbol{\Phi}\in\mathbf{R}^{m\times n}$是可测信息矩阵.

参考文献

References

[1] 丁锋.系统辨识新论[M].北京:科学出版社,2013
DING Feng. System identification: New theory and methods[M].Beijing:Science Press,2013

[2] 丁锋.系统辨识:辨识方法性能分析[M].北京:科学出版社,2014
DING Feng. System identification: Performance analysis for identification methods [M]. Beijing: Science Press,2014

[3] 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 and Technology:Natural Science Edition,2011,3(1):1-22

[4] 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 and Technology: Natural Science Edition,2011,3(2):97-117

[5] 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 and Technology: Natural Science Edition,2011,3(3):193-226

[6] 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 and Technology:Natural Science Edition,2011,3(4):289-318

[7] 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 and Technology: Natural Science Edition, 2011, 3 (6): 481-510

[8] 丁锋.系统辨识(6):多新息辨识理论与方法[J].南京信息工程大学学报:自然科学版,2012,4(1):1-28
DING Feng.System identification.Part F:Multi-innovation identification theory and methods[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2012,4(1):1-28

[9] 丁锋.系统辨识(7):递阶辨识原理与方法[J].南京信息工程大学学报:自然科学版,2012,4(2):97-124
DING Feng.System identification.Part G:Hierarchical identification principle and methods [J]. Journal of Nanjing University of Information Science and Technology:Natural Science Edition,2012,4(2):97-124

[10] 丁锋.系统辨识(8):耦合辨识概念与方法[J].南京信息工程大学学报:自然科学版,2012,4(3):193-212
DING Feng.System identification.Part H:Coupling identification concept and methods[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2012,4(3):193-212

[11] 丁锋.辨识方法的计算效率(1):递推算法[J].南京信息工程大学学报:自然科学版,2012,4(4):289-300
DING Feng.Computational efficiency of the identification methods.Part A:Recursive algorithms[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2012,4(4):289-300

[12] 丁锋.辨识方法的计算效率(2):迭代算法[J].南京信息工程大学学报:自然科学版,2012,4(5):385-401
DING Feng.Computational efficiency of the identification methods. Part B: Iterative algorithms [J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2012, 4 (5): 385-401

[13] 丁锋.辨识方法的计算效率(3):信息向量耦合算法[J].南京信息工程大学学报:自然科学版,2012,4(6):481-495
DING Feng.Computational efficiency of the identification

methods. Part C: Coupled information vector algorithms [J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2012, 4(6): 481-495.

[14] 丁锋,汪菲菲.多元系统耦合多新息随机梯度类辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(1):1-16
DING Feng, WANG Feifei. Coupled multi-innovation stochastic gradient type identification methods for multivariate systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(1):1-16

[15] 丁锋,汪菲菲,汪学海.多元伪线性回归系统部分耦合多新息随机梯度类辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(2):97-112
DING Feng, WANG Feifei, WANG Xuehai. Partially Coupled multi-innovation stochastic gradient type identification methods for multivariate pseudo-linear regressive systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(2):97-112

[16] 丁锋,汪菲菲,汪学海.类多变量输出误差系统的耦合多新息辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(3):193-210
DING Feng, WANG Feifei, WANG Xuehai. Coupled multi-innovation identification methods for multivariable output-error-like systems [J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(3):193-210

[17] 丁锋,汪菲菲,汪学海.多变量方程误差类系统的部分耦合迭代辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(4):289-305
DING Feng, WANG Feifei, WANG Xuehai. Partially coupled iterative identification methods for multivariable equation error type systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(4):289-305

[18] 丁锋,王艳娇.类多变量方程误差类系统的递阶多新息辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(5):385-404
DING Feng, WANG Yanjiao. Hierarchical multi-innovation identification methods for multivariable equation-error-like type systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(5):385-404

[19] 丁锋,马兴云.规范状态空间系统辨识方法[J].南京信息工程大学学报:自然科学版,2014,6(6):481-504
DING Feng, MA Xingyun. Identification methods for canonical state space systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2014,6(6):481-504

[20] 丁锋,毛亚文.输入非线性方程误差自回归系统的多新息辨识方法[J].南京信息工程大学学报:自然科学版,2015,7(1):1-23
DING Feng, MAO Yawen. Multi-innovation identification methods for input nonlinear equation-error autoregressive systems[J].Journal of Nanjing University of Information Science and Technology:Natural Science Edition,2015,7(1):1-23

[21] 丁锋,陈慧波.输入非线性方程误差系统的多新息辨识方法[J].南京信息工程大学学报:自然科学版,2015,7(2):97-124
DING Feng, CHEN Huibo. Multi-innovation identification methods for input nonlinear equation-error systems[J].Journal of Nanjing University of Information Science and Technology:Natural Science Edition,2015,7(2):97-124

[22] 丁锋,陈启佳.输出非线性方程误差类系统递推最小二乘辨识方法[J].南京信息工程大学学报:自然科学版,2015,7(3):193-213
DING Feng, CHEN Qijia. Recursive least squares identification methods for output nonlinear equation-error type systems.Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2015, 7(3):193-213

[23] 丁锋,郭兰杰.线性参数系统的多新息辨识方法[J].南京信息工程大学学报:自然科学版,2015,7(4):289-312
DING Feng, GUO Lanjie. Multi-innovation identification methods for linear-parameter systems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2015,7(4):289-312

[24] Ding F, Xiao Y S. A finite-data-window least squares algorithm with a forgetting factor for dynamical modeling [J]. Applied Mathematics and Computation, 2007, 186(1):184-192

[25] 丁锋,丁韬,萧德云,杨家本.时变系统有限数据窗最小二乘辨识算法的有界收敛性[J].自动化学报,2002,28(5):754-761
DING Feng, DING Tao, XIAO Deyun, YANG Jiaben. Bounded convergence of finite data window least squares identification for time-varying systems [J]. Acta Automatica Sinica,2002,28(5):754-761

[26] Xiao Y S, Wang D Q, Ding F. The residual based ESG algorithm and its performance analysis[J]. Journal of the Franklin Institute: Engineering and Applied Mathematics,2010,347(2):426-437.

[27] Ding F, Shi Y, Chen T. Gradient-based identification methods for Hammerstein nonlinear ARMAX models[J]. Nonlinear Dynamics,2006,45(1-2):31-43

[28] Chen J, Ding F. Modified stochastic gradient algorithms with fast convergence rates[J]. Journal of Vibration and Control,2011,17(9):1281-1286

[29] Ding J, Shi Y, Wang H G, Ding F. A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems[J]. Digital Signal Processing, 2010,20(4):1238-1247

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

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

[32] 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

[33] Liu Y J, Yu L, Ding F. Multi-innovation extended

stochastic gradient algorithm and its performance analysis [J]. Circuits, Systems and Signal Processing, 2010, 29 (4):649-667

[34] 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

[35] Ding F.Two-stage least squares based iterative estimation algorithm for CARARMA system modeling[J]. Applied Mathematical Modelling,2013,37(7):4798-4808

[36] Ding F,Liu X M,Chen H B,Yao G.Y.Hierarchical gradient based and hierarchical least squares based iterative parameter identification for CARARMA systems [J]. Signal Processing,2014,97:31-39

[37] 方崇智,萧德云.过程辨识[M].北京:清华大学出版社,1988
FANG Chongzhi, XIAO Deyun. Process Identification [M].Beijing:Tsinghua University Press,1988

[38] Ding F,Wang Y J,Ding J.Recursive least squares parameter identification for systems with colored noise using the filtering technique and the auxiliary model[J]. Digital Signal Processing,2015,37:100-108

[39] Wang D Q. Least squares-based recursive and iterative estimation for output error moving average systems using data filtering[J].IET Control theory and Applications, 2011,5(14):1648-1657

[40] Guo L J,Ding F.Least squares based iterative algorithm for pseudo-linear autoregressive moving average systems using the data filtering technique [J]. Journal of the Franklin Institute: Engineering and Applied Mathematics,2015,352(10):4339-4353

[41] Mao Y W,Ding F.Multi-innovation stochastic gradient identification for Hammerstein controlled autoregressive autoregressive systems based on the filtering technique [J].Nonlinear Dynamics,2015,79(3):1745-1755

[42] Mao Y W,Ding F. Data filtering-based multi-innovation stochastic gradient algorithm for nonlinear output error autoregressive systems [J]. Circuits, Systems and Signal Processing,2015.doi 10. 1007/s00034-015-0064-y

[43] Mao Y W,Ding F.A novel data filtering based multi-innovation stochastic gradient algorithm for Hammerstein nonlinear systems [J]. Digital Signal Processing, 2015. http://dx.doi.org/10. 1016/j.dsp.2015. 07. 002

[44] Mao Y W,Ding F.Parameter estimation for nonlinear systems by using the data filtering and the multi-innovation identification theory [J]. International Journal of Computer Mathematics 92 (x) (2015). http://dx.doi.org/10. 1080/00207160. 2015. 1077949

[45] Wang D Q,Ding F. Input-output data filtering based recursive least squares parameter estimation for CARARMA systems[J].Digital Signal Processing,2010, 20(4):991-999

[46] Ding F,Liu Y J,Bao B.Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems[J].Journal of Systems and Control Engineering,2012,226(1):43-55

Multi-innovation identification methods for equation-error systems

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 The multi-innovation methods can be applied to adaptive filtering,parameter estimation,self-tuning control,adaptive fault detection and diagnosis,etc.for linear systems and nonlinear systems.Linear systems include two basic types:equation-error type systems and output-error type systems.This paper studies the multi-innovation (MI) extended stochastic gradient algorithm and the MI extended least squares identification algorithm for equation-error moving average (EEMA) systems (namely, CARMA systems), and presents the decomposition based MI generalized extended stochastic gradient (GESG) algorithm and the MI generalized extended least squares (GELS) algorithm,and the filtering based MI-GESG algorithm and the filtering based MI-GELS algorithm for equation-error autoregressive moving average (EEARMA) systems (namely,CARARMA systems).

Key words parameter estimation;recursive identification;gradient search;least squares;filtering;decomposition;auxiliary model identification idea;multi-innovation identification theory;hierarchical identification principle;equation-error system;linear system