文章编号:1674-7070(2015)04-0289-24

丁锋[1,2,3]　郭兰杰[1]

线性参数系统的多新息辨识方法

摘要

系统有线性和非线性之分.线性系统有统一的描述形式,非线性系统因类别无数,不可能有统一描述.线性参数系统是一类特殊的非线性系统,它在参数空间上呈现线性特征,介于线性系统与非线性系统之间.针对伪线性参数系统,讨论了基于辅助模型的多新息辨识方法、基于滤波的辅助模型多新息辨识方法、基于模型分解的辅助模型多新息辨识方法、基于滤波的分解多新息辨识方法,并给出了几个典型辨识算法的计算量、计算步骤和流程图.

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;伪线性回归模型;伪线性系统;线性参数系统

中图分类号 TP273
文献标志码 A

收稿日期 2015-08-01
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

线性系统的辨识方法很成熟,如最小二乘辨识方法、随机梯度辨识方法、辅助模型辨识方法、多新息辨识方法、递阶辨识方法、耦合辨识方法等,这些在最近出版的“系统辨识学术专著丛书”的《系统辨识新论》[1]和《系统辨识——辨识方法性能分析》[2]中有比较深入的介绍.如何将这些方法推广用于研究线性参数系统、双线性系统、双线性参数系统、块结构非线性系统,乃至非线性系统的辨识是当前控制领域的重要问题之一.

为了解决存在不可测变量的系统、存在数据损失的系统辨识问题,为了提高辨识精度,为了减小辨识算法的计算量,一些新的研究理念不断出现,如辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念等.这些辨识方法都在笔者等发表在《南京信息工程大学学报》2011—2012 年的连载论文中进行了详细介绍,而 2014 年至今的连载论文则分别讨论了多元系统的耦合多新息随机梯度辨识方法[3]和部分耦合多新息随机梯度辨识方法[4]、类多变量输出误差系统的耦合多新息辨识方法[5]、多变量方程误差类系统的部分耦合迭代辨识方法[6]、类多变量方程误差类系统的递阶多新息辨识方法[7]、规范状态空间系统辨识方法[8]、输入非线性方程误差自回归系统的多新息辨识方法[9]、输入非线性方程误差系统的递阶多新息辨识方法[10]以及输出非线性方程误差类系统递推最小二乘辨识方法[11].

近年来的研究焦点是非线性系统辨识.非线性系统解析数学表达式各种各样,结构也极其复杂,不可能有统一的辨识表达形式,也不可能取得深入有价值的研究结果.因此,研究工作主要集中在块结构非线性系统的辨识.所谓块结构非线性系统是由线性动态子系统与静态非线性环节串联或反馈等构成.当静态非线性环节串联于动态子系统之前,就是输入非线性系统;当非线性环节是一个多项式或已知基函数未知系数的线性组合,就称为 Hammerstein 非线性系统.对于 Hammerstein 非线性系统,丁锋等[12]发表在国际期刊《Digital Signal Processing》上的 24 页长文,综述了一些非线性系统辨识方法的研究现状,详细阐述了 Hammerstein 非线性系统的一些辨识方法,如投影辨识方法、随机梯度辨识方法、牛顿递推辨识方法、牛顿迭代辨识方法等(该文入选“2011 年中国百篇最具影响国际学术论文”).研究论

文“使用递阶辨识原理的 Hammerstein 非线性系统的梯度迭代辨识方法和最小二乘迭代辨识方法”[13]入选国际期刊《IET Control Theory and Applications》2015 年最佳论文(Premium Award for Best Paper).

当静态非线性环节串联于动态子系统之后,就是输出非线性系统;当非线性环节是一个多项式或已知基函数未知系数的线性组合,就称为 Wiener 非线性系统.对于 Wiener 非线性系统,文献[14]提出了一个牛顿迭代辨识方法.当两个静态非线性环节串联于动态子系统的两端,就是输入输出非线性系统;当非线性环节是一个多项式或已知基函数未知系数的线性组合,就称为 Hammerstein-Wiener 非线性系统.对于 Hammerstein-Wiener 非线性系统,文献[15]假设输出端静态非线性环节的逆存在,且是一个已知基函数未知系数的线性组合,利用递阶辨识原理,提出一个递阶最小二乘辨识方法,即一个三阶段递推最小二乘辨识方法.当两个动态子系统串联于一个静态非线性环节的两端,便称为 Wiener-Hammerstein 系统.当静态非线性环节位于反馈通道,前向通道是一个线性动态子系统时,或反之,就称为反馈非线性系统.当然,非线性环节也可以是硬非线性(hard nonlinearity),如饱和非线性、死区非线性、继电器非线性等.

当然,也可以这样定义非线性系统:输入非线性系统定义为系统输出是输入的非线性函数,是过去输出的线性函数[9-10,12,16-19];输出非线性系统定义为系统输出是过去输出的非线性函数,是输入的线性函数[11,20-21];输入输出非线性系统定义为系统输出既是输入的非线性函数,又是输入的非线性函数[22].对于这样的输出非线性系统,文献[23-24]提出了基于分解的牛顿迭代辨识方法,文献[20]提出了基于分解的梯度迭代辨识方法和最小二乘迭代辨识方法(此文入选“2011 年中国百篇最具影响国际学术论文”和 2014 年欧洲信号处理协会 3 篇最佳论文奖之一).对于这样的输入输出非线性系统,文献[21]提出了一个增广随机梯度辨识方法.对于反馈非线性系统,文献[23-24]基于递阶辨识原理提出了反馈非线性系统辅助模型递推最小二乘辨识方法和多阶段最小二乘迭代辨识方法,还有一类特殊的非线性系统,称为线性参数系统,介于线性系统与非线性系统之间,其输出是参数的线性函数,是输入和输出的非线性函数.本文基于辅助模型辨识思想、基于数据滤波技术、基于辨识模型分解技术,研究和提出这类线性参数系统的梯度辨识方法、多新息随机梯度辨识方法、递推最小二乘辨识方法和多新息最小二乘辨识方法,并讨论了几个典型辨识算法的计算量,给出了计算步骤和计算流程图.

1 基于辅助模型的多新息辨识方法

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 $zy(t)=y(t+1)$,$A(z)$,$B(z)$,$C(z)$,$D(z)$ 和 $F(z)$ 是算子 z^{-1} 的常系数时不变多项式,定义如下:

$$A(z):=1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a},\quad a_i\in\mathbf{R},$$
$$B(z):=b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b},\quad b_i\in\mathbf{R},$$
$$C(z):=1+c_1z^{-1}+c_2z^{-2}+\cdots+c_{n_c}z^{-n_c},\quad c_i\in\mathbf{R},$$
$$D(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d},\quad d_i\in\mathbf{R},$$
$$F(z):=1+f_1z^{-1}+f_2z^{-2}+\cdots+f_{n_f}z^{-n_f},\quad f_i\in\mathbf{R}.$$

多项式系数 a_i,b_i,c_i,d_i 和 f_i 为模型参数,设阶次 n_a,n_b,n_c,n_d 和 n_f 已知,根据移位算子的性质,有

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

伪线性系统是指系统经过参数化,得到伪线性回归模型,其信息向量或信息矩阵包含未知变量(未知内部变量或未知噪声项)的一类线性系统.

下列线性系统都是伪线性系统.

1) 自回归输出误差系统(AR-OE 系统):

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

2) 自回归输出误差滑动平均系统(AR-OEMA 系统):

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

3) 自回归输出误差自回归系统(AR-OEAR 系统):

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

4) 自回归输出误差自回归滑动平均系统(AR-OEARMA 系统),即自回归 Box-Jenkins 系统(AR-Box-Jenkins 系统):

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

5) 受控自回归自回归滑动平均系统(CARARMA 系统),即方程误差自回归滑动平均系统(EEARMA 系统):

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

上述系统经过参数化,都可以写为下列伪线性回归形式(伪线性回归辨识模型):

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

例如,对于 CARARMA 系统,令

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

则 CARARMA 系统的参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 分别为

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

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

信息向量 $\boldsymbol{\varphi}(t)$ 包含了系统的输入输出数据 $u(t-i)$ 和 $y(t-i)$,还包含了未知噪声项 $w(t-i)$ 和 $v(t-i)$,且输出 $y(t)$ 是 $u(t-i)$,$y(t-i)$,$w(t-i)$ 和 $v(t-i)$ 的线性函数,也是参数向量 $\boldsymbol{\theta}$ 的线性函数,故上述系统是伪线性系统.这类伪线性回归系统是一类线性系统,涵盖标量线性系统、多变量线性系统和多元线性系统,可以是时不变线性系统,也可以是时变线性系统.

另一类是伪线性参数系统,参数化后得到伪线性参数回归模型(伪线性参数辨识模型).顾名思义,伪线性参数系统是参数的线性函数,其输出 $y(t)$ 是 $u(t-i)$,$y(t-i)$,$w(t-i)$ 和 $v(t-i)$ 等的非线性函数.因此,伪线性参数系统是一类特殊的非线性系统,它包含了作为特例的伪线性系统,它是输入输出变量的非线性函数,是系统参数的线性函数,可以是标量非线性系统、多变量非线性系统和多元非线性系统,可以是时不变非线性系统,也可以是时变非线性系统.

1.1 系统描述与辨识模型

考虑下列一类伪线性参数系统(pseudo-linear-parameter system)的辨识问题:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t),\tag{1}$$

其示意如图 1 所示,$\boldsymbol{\theta}\in\mathbf{R}^m$ 是系统的部分参数向量,信息向量 $\boldsymbol{\phi}(t)\in\mathbf{R}^m$ 是由系统的观测数据构成的.

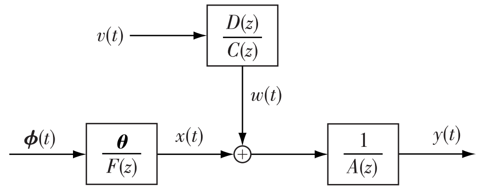

图 1 伪线性参数系统
Fig. 1 A pseudo-linear-parameter system

辨识目标是利用系统的观测数据 $\{y(t),\boldsymbol{\phi}(t)\}$,研究和提出辨识方法来估计系统参数,即估计系统的参数向量 $\boldsymbol{\theta}$ 和多项式 $A(z)$,$F(z)$,$C(z)$ 和 $D(z)$ 的系数 a_i,f_i,c_i 和 d_i.

设 $t\leqslant 0$ 时,$\boldsymbol{\phi}(t)=\mathbf{0}$,$y(t)=0$,$x(t)=0$,$w(t)=0$,$v(t)=0$.记 $n_0:=n_a+n_f+m$,$n:=n_a+n_f+m+n_c+n_d$.设 $\hat{X}(t)$ 为 X 在时刻 t 的估计.

注 1 当信息向量 $\boldsymbol{\phi}(t)$ 关于系统观测变量是线性的,式(1)就是一个伪线性回归系统.例如,当 $\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}=b_1u(t)+b_2u(t-1)+\cdots+b_{n_b}u(t-n_b)$ 时,式(1)就是一个伪线性系统(自回归 Box-Jenkins 系统).

注 2 当信息向量 $\boldsymbol{\phi}(t)$ 关于系统观测变量是非线性的,式(1)就是一个伪线性参数系统.例如,当 $\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}=b_1u^2(t)+b_2u^3(t-1)+\cdots+b_m\sin u(t-m)$ 时,式(1)就是一个伪线性参数系统(非线性系统).

定义中间变量 $x(t)$ 和相关噪声 $w(t)$:

$$x(t):=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}}{F(z)}\in\mathbf{R},\tag{2}$$

$$w(t):=\frac{D(z)}{C(z)}v(t)\in\mathbf{R}.\tag{3}$$

定义参数向量:

$$\boldsymbol{\vartheta}:=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{f}^{\mathrm{T}},\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{n}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n},$$

$$\boldsymbol{\theta}_{\mathrm{n}}:=[\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$

$$\boldsymbol{a}:=[a_1,a_2,\cdots,a_{n_a}]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$

$$\boldsymbol{f}:=[f_1,f_2,\cdots,f_{n_f}]^{\mathrm{T}}\in\mathbf{R}^{n_f},$$

$$\boldsymbol{c}:=[c_1,c_2,\cdots,c_{n_c}]^{\mathrm{T}}\in\mathbf{R}^{n_c},$$

$$\boldsymbol{d}:=[d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_d}$$

和信息向量:

$$\boldsymbol{\varphi}(t):=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\boldsymbol{\varphi}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n},$$

$$\boldsymbol{\varphi}_{\mathrm{n}}(t):=[\boldsymbol{\varphi}_w^{\mathrm{T}}(t),\boldsymbol{\varphi}_v^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$

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

$$\boldsymbol{\varphi}_x(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f},$$

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

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

设 $\boldsymbol{a},\boldsymbol{f}$ 和 $\boldsymbol{c},\boldsymbol{d}$ 在时刻 t 的估计分别为 $\hat{\boldsymbol{a}}(t),\hat{\boldsymbol{f}}(t)$ 和

$$\hat{\boldsymbol{c}}(t):=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_c},$$

$$\hat{\boldsymbol{d}}(t):=[\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_d},$$

设 $\hat{\boldsymbol{\vartheta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n}$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t):=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}\in\mathbf{R}^{n_c+n_d}$

分别为 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 在时刻 t 的估计.

借助于上述定义,式(2)和(3)可以写为

$$x(t)=[1-F(z)]x(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}=\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta},\tag{4}$$

$$w(t)=[1-C(z)]w(t)+D(z)v(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\tag{5}$$

于是伪线性参数系统(1)可以写为下列辨识模型(identification model):

$$\begin{aligned}y(t)&=[1-A(z)]y(t)+x(t)+w(t)\\&=\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}+x(t)+w(t)\qquad(6)\\&=\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}+\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t)\\&=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}+v(t).\qquad(7)\end{aligned}$$

观察辨识模型(7),可知信息向量 $\boldsymbol{\varphi}(t)$ 包含观测数据 $y(t-i)$ 和 $\boldsymbol{\phi}(t-i)$,还包括了未知中间变量 $x(t-i)$ 以及不可测噪声项 $w(t-i)$ 和 $v(t-i)$,这是辨识的困难所在.解决方式是在推导辨识方法时,利用辅助模型辨识思想[25],这些未知量用其对应的估计值代替.

对于不可测内部变量 $x(t-i)$ 的处理,根据 $x(t)$ 定义的结构式,建立一个辅助模型:

$$x_a(t)=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_a}{F_a(z)},\tag{8}$$

参考式(2)写成式(4)的方法,式(8)可以写为下列形式:

$$x_a(t)=\boldsymbol{\varphi}_{xa}^{\mathrm{T}}(t)\boldsymbol{f}_a+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_a,\tag{9}$$

其中 $\boldsymbol{f}_a$ 和 $\boldsymbol{\theta}_a$ 是辅助模型的参数向量,$\boldsymbol{\varphi}_{xa}(t)$ 和 $\boldsymbol{\phi}(t)$ 是辅助模型在时刻 t 的信息向量.注意,因为辅助模型是根据系统的可测信息构造的,$\boldsymbol{\phi}(t)$ 是已知的观测向量,故使用在辅助模型中.

辅助模型有很多选择方法[26-28],这里使用参数向量 $\boldsymbol{f}$ 和 $\boldsymbol{\theta}$ 的估计 $\hat{\boldsymbol{f}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$ 作为辅助模型的参数向量 $\boldsymbol{f}_a$ 和 $\boldsymbol{\theta}_a$,即 $\boldsymbol{f}_a=\hat{\boldsymbol{f}}(t)$ 和 $\boldsymbol{\theta}_a=\hat{\boldsymbol{\theta}}(t)$,信息向量 $\boldsymbol{\varphi}_x(t)$ 的估计 $\hat{\boldsymbol{\varphi}}_x(t)$ 作为辅助模型的信息向量 $\boldsymbol{\varphi}_{xa}(t)$,即 $\boldsymbol{\varphi}_{xa}(t)=\hat{\boldsymbol{\varphi}}_x(t)$,那么辅助模型可以表示为

$$x_a(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).\tag{10}$$

有了辅助模型,在推导辨识算法时,信息向量中包含的未知中间变量 $x(t-i)$ 就用辅助模型的输出 $x_a(t-i)$ 代替,不可测噪声项 $w(t-i)$ 和 $v(t-i)$ 就用其估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 代替.基于这种思想推导出的辨识算法就是辅助模型辨识算法,它可以用于线性和非线性输出误差类系统,以及伪线性参数输出误差系统的辨识.

在辅助模型辨识方法中,未知变量可以用其估计代替.因此,用辅助模型的输出 $x_a(t-i)$,不可测噪声项 $w(t-i)$ 和 $v(t-i)$ 的估计 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$ 定义估计的信息向量:

$$\hat{\boldsymbol{\varphi}}(t):=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n},$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t):=[\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$

$$\hat{\boldsymbol{\varphi}}_a(t):=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f},$$

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

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

1.2 基于辅助模型的广义增广随机梯度辨识算法

随机梯度辨识算法是指沿着准则函数负梯度方向进行搜索的参数估计方法.根据辨识模型(7),定义准则函数:

$$J_1(\boldsymbol{\vartheta}):=\frac{1}{2}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}]^2.$$

令 $\hat{\boldsymbol{\vartheta}}(t)$ 表示参数向量 $\boldsymbol{\vartheta}$ 在时刻 t 的估计.

极小化准则函数 $J_1(\boldsymbol{\vartheta})$ 得到下列递推关系:

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\tag{11}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1.\tag{12}$$

这个算法不可能实现,因为信息向量 $\boldsymbol{\varphi}(t)$ 是未知的,不仅包含了不可测真实输出 $x(t-i)$,而且还包含了未知噪声 $w(t-i)$ 和 $v(t-i)$.解决方法是借助于辅助模型辨识思想,未知变量用其估计代替,具体思路如下:

由式(6)可得 $w(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}-x(t)$.将其中未知参数向量 $\boldsymbol{a}$ 和未知中间变量 $x(t)$ 分别用其估计 $\hat{\boldsymbol{a}}(t)$ 和辅助模型的输出 $x_a(t)$ 代替,那么 $w(t)$ 的估计 $\hat{w}(t)$ 可以通过下式计算:

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t).$$

由式(7)可得 $v(t)=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}$.将其中未知信息向量 $\boldsymbol{\varphi}(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}(t)$ 代替,未知参数向量 $\boldsymbol{\vartheta}$ 用其估计 $\hat{\boldsymbol{\vartheta}}(t)$ 代替,那么 $v(t)$ 的估计 $\hat{v}(t)$ 可以通过下式计算:

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

式(11)—(12)中未知向量 $\boldsymbol{\varphi}(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}(t)$ 代替,定义新息 $e(t):=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的辅助模型广义增广随机梯度算法(Auxiliary Model based Generalized Extended Stochastic Gradient algorithm,AM-GESG 算法),其基本方程总结为

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t), \tag{13}$$

$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{14}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2, \tag{15}$$

$$x_a(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{16}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t), \tag{17}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{18}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{19}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{20}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{21}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_f)]^{\mathrm{T}}, \tag{22}$$

$$\hat{\boldsymbol{\varphi}}_w(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}}, \tag{23}$$

$$\hat{\boldsymbol{\varphi}}_v(t)=[\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{24}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{25}$$

AM-GESG 算法(13)—(25)的计算步骤如下:

1) 初始化:令 $t=1$,给定参数估计精度 ε,置参数估计初值 $\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_n/p_0$,设置 $r(0)=1$,$x_a(i)=1/p_0$,$\hat{w}(i)=1/p_0$,$\hat{v}(i)=1/p_0$,$i\leqslant 0$,$p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,用式(21)—(24)构成信息向量 $\boldsymbol{\varphi}_y(t)$,$\hat{\boldsymbol{\varphi}}_x(t)$,$\hat{\boldsymbol{\varphi}}_w(t)$ 和 $\hat{\boldsymbol{\varphi}}_v(t)$,用式(20)和(19)分别构成信息向量 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 和 $\hat{\boldsymbol{\varphi}}(t)$.

3) 用式(14)—(15)分别计算新息 $e(t)$ 和 $r(t)$.

4) 用式(13)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)$.

5) 从式(25)的 $\hat{\boldsymbol{\vartheta}}(t)$ 中读取 $\hat{\boldsymbol{a}}(t)$,$\hat{\boldsymbol{f}}(t)$,$\hat{\boldsymbol{\theta}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$,用式(16)—(18)计算 $x_a(t)$,$\hat{w}(t)$ 和 $\hat{v}(t)$.

6) 比较 $\hat{\boldsymbol{\vartheta}}(t)$ 和 $\hat{\boldsymbol{\vartheta}}(t-1)$,若 $\|\hat{\boldsymbol{\vartheta}}(t)-\hat{\boldsymbol{\vartheta}}(t-1)\|\leqslant\varepsilon$,结束计算,得到参数估计 $\hat{\boldsymbol{\vartheta}}(t)$,否则 t 增加 1 转到第 2)步,继续进行递推计算.

AM-GESG 算法计算参数估计 $\hat{\boldsymbol{\vartheta}}(t)$ 的流程如图 2 所示.

注 3 随机梯度算法具有计算量小的优点,但是收敛速度慢,参数估计精度不高.为了提高收敛速度,可引入遗忘因子 λ,将式(15)中 $r(t)$ 修改为

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

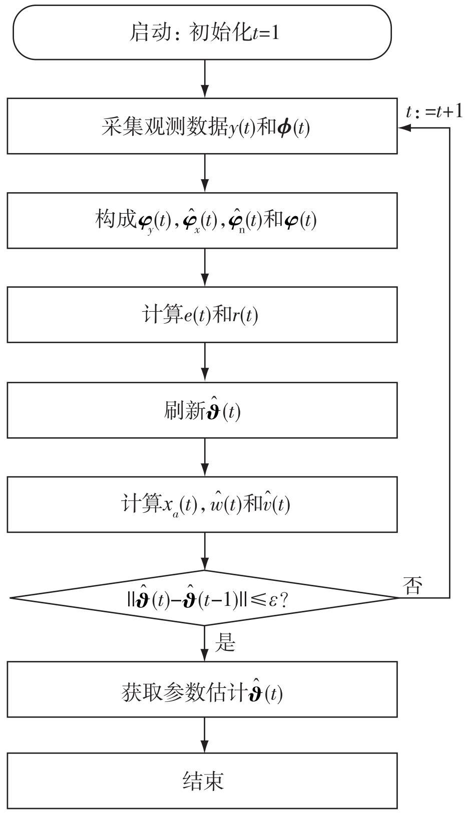

图 2 AM-GESG 算法计算参数估计的流程

Fig. 2 The flowchart of the AM-GESG algorithm

注 4 为了提高随机梯度算法的暂态收敛速度和稳态性能,可在随机梯度辨识算法中引入收敛指数(convergence index)ε,就得到修正随机梯度算法(Modified Stochastic Gradient algorithm,M-SG 算法),或称为 ε 随机梯度算法(Epsilon Stochastic Gradient algorithm,ε-SG 算法)[1,29-30].在式(13)中引入收敛指数 ε:

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

就得到辅助模型修正广义增广随机梯度(AM-M-GESG)算法.这里 $1/r^{\varepsilon}(t)$ 是收敛因子或步长.修正随机梯度算法是本文第一作者最近提出的,它比随机梯度算法具有更快的收敛速度,其性能优于遗忘梯度算法[29-30].

本文要研究的多新息随机梯度算法、基于滤波的随机梯度算法、基于滤波的多新息随机梯度算法、基于模型分解的随机梯度算法、基于滤波的多新息随机梯度算法等,都可以引入遗忘因子或收敛指数来提高算法的暂态和稳态性能.

1.3 基于辅助模型的多新息广义增广随机梯度辨识算法

辅助模型广义增广随机梯度算法(13)—(25)收敛速度慢,尽管可以通过引入遗忘因子或收敛指数改善收敛速度,但本节通过充分利用数据信息和扩展新息长度来提高算法的参数估计精度.

式(14)中 $e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$ 是标量新息(scalar innovation).下面借助多新息辨识理论[1,31-32],通过扩展新息 $e(t)$ 维数,推导基于辅助模型的多新息广义增广随机梯度算法,从而提高算法的估计精度.

将标量新息 $e(t)\in\mathbf{R}$ 扩展为向量新息(即多新息):

$$\boldsymbol{E}(p,t):=\begin{bmatrix}e(t)\\e(t-1)\\\vdots\\e(t-p+1)\end{bmatrix}=\begin{bmatrix}y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\\y(t-1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\vartheta}}(t-2)\\\vdots\\y(t-p+1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-p)\end{bmatrix}\in\mathbf{R}^p,$$

其中正整数 p 表示新息长度.一般情况下,可以认为时刻$(t-1)$的估计 $\hat{\boldsymbol{\vartheta}}(t-1)$ 比时刻$(t-i)$ $(i>2)$的估计 $\hat{\boldsymbol{\vartheta}}(t-i)$ 更接近真值 $\boldsymbol{\vartheta}$,所以多新息向量可修改为

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

定义堆积向量 $\boldsymbol{Y}(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$ 如下:

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

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

于是多新息向量可表达为

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

因为 $\boldsymbol{Y}(1,t)=y(t)$, $\hat{\boldsymbol{\Phi}}(1,t)=\hat{\boldsymbol{\varphi}}(t)$, $\boldsymbol{E}(1,t)=e(t)$,所以式(13)可以等价表达为

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

上式中新息长度 $p=1$.基于式(13)—(25),推广上式中的新息长度 1 为 p,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的,新息长度为 p 的辅助模型多新息广义增广随机梯度算法(Auxiliary Model based Multi-Innovation Generalized Extended Stochastic Gradient algorithm, AM-MI-GESG 算法):

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

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

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2,\tag{28}$$

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

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

$$x_a(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{31}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t),\tag{32}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\tag{33}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{34}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\tag{35}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_f)]^{\mathrm{T}},\tag{36}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{37}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{f}}(t)\\\hat{\boldsymbol{\theta}}(t)\\\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{38}$$

AM-MI-GESG 算法(26)—(38)的计算步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p,数据长度 L,置参数估计初值 $\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_n/p_0$,设置 $r(0)=1$, $x_a(i)=1/p_0$, $\hat{w}(i)=1/p_0$, $\hat{v}(i)=1/p_0$, $i\leqslant 0$, $p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,构成信息向量 $\boldsymbol{\phi}(t)$,由式(35)—(37)和(34)构成信息向量 $\boldsymbol{\varphi}_y(t)$, $\hat{\boldsymbol{\varphi}}_x(t)$, $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 和 $\hat{\boldsymbol{\varphi}}(t)$,用式(29)—(30)构造堆积输出向量 $\boldsymbol{Y}(p,t)$ 和堆积信息向量 $\hat{\boldsymbol{\Phi}}(p,t)$.

3) 用式(27)—(28)计算新息向量 $\boldsymbol{E}(p,t)$ 和 $r(t)$.

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

5) 从式(38)的 $\hat{\boldsymbol{\vartheta}}(t)$ 中读取 $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{f}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$,用式(31)—(33)计算 $x_a(t)$, $\hat{w}(t)$ 和 $\hat{v}(t)$.

6) 如果 $t<L$, t 增加 1 转到第 2)步,否则获得参数估计向量 $\hat{\boldsymbol{\vartheta}}(L)$,结束计算过程.

AM-MI-GESG 算法计算参数估计的流程如图 3 所示.

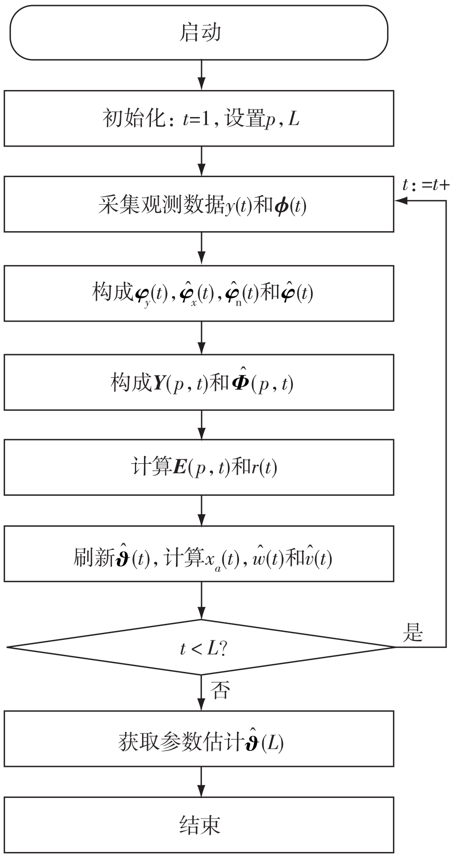

图 3　AM-MI-GESG 算法计算参数估计的流程
Fig. 3　The flowchart of the AM-MI-GESG algorithm

注 5　为了加快 AM-MI-GESG 算法的收敛速度,可以使算法增益向量$\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}$大一些,可以将式(28)中 $r(t)$ 修改为

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

当然,也可以引入遗忘因子或收敛指数.

注 6　通过扩展新息长度,多新息随机梯度算法能提高参数估计精度.随着新息长度的增加,多新息随机梯度算法的收敛性逼近于最小二乘算法.与最小二乘算法相比,随机梯度算法仍存在收敛速度慢的缺点,下面介绍收敛速度更快的最小二乘算法.

1.4　基于辅助模型的递推广义增广最小二乘辨识算法

对于辨识模型(7),极小化最小二乘准则函数(least squares criterion function):

$$J_2(\boldsymbol{\vartheta}):=\sum_{j=1}^{t}[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\vartheta}]^2,$$

运用辅助模型辨识思想,参照辅助模型递推最小二乘算法的推导[1],未知变量用其估计代替,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的辅助模型递推广义增广最小二乘算法(Auxiliary Model based Recursive Generalized Extended Least Squares algorithm,AM-RGELS 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\quad(39)$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1},\quad(40)$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1),\quad(41)$$

$$x_a(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad(42)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t),\quad(43)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\quad(44)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}},\quad(45)$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\quad(46)$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_f)]^{\mathrm{T}},\quad(47)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\quad(48)$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\quad(49)$$

AM-RGELS 算法(39)—(49)的计算步骤如下:

1) 初始化:给定参数估计精度 ε,令 $t=1$,置参数估计初值 $\hat{\boldsymbol{\vartheta}}(0)=\boldsymbol{1}_n/p_0$ 和协方差阵初值 $\boldsymbol{P}(0)=p_0\boldsymbol{I}_n$,设置 $x_a(i)=1/p_0$,$\hat{w}(i)=1/p_0$,$\hat{v}(i)=1/p_0$,$i\leqslant 0$,$p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,构成信息向量 $\boldsymbol{\phi}(t)$,由式(46)—(48)和(45)构成信息向量$\boldsymbol{\varphi}_y(t)$,$\hat{\boldsymbol{\varphi}}_x(t)$,$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 和 $\hat{\boldsymbol{\varphi}}(t)$.

3) 通过式(40)和(41)计算增益向量 $\boldsymbol{L}(t)$ 和协方差矩阵 $\boldsymbol{P}(t)$.

4) 通过式(39)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)$.

5) 从式(49)的 $\hat{\boldsymbol{\vartheta}}(t)$ 中读取 $\hat{\boldsymbol{a}}(t)$,$\hat{\boldsymbol{f}}(t)$,$\hat{\boldsymbol{\theta}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$,用式(42)—(44)计算 $x_a(t)$,$\hat{w}(t)$ 和 $\hat{v}(t)$.

6) 比较 $\hat{\boldsymbol{\vartheta}}(t)$ 和 $\hat{\boldsymbol{\vartheta}}(t-1)$,若 $\|\hat{\boldsymbol{\vartheta}}(t)-\hat{\boldsymbol{\vartheta}}(t-1)\|\leqslant\varepsilon$,结束计算,得到参数估计 $\hat{\boldsymbol{\vartheta}}(t)$,否则 t 增加 1 转到第 2)步,继续进行递推计算.

AM-RGELS 算法计算参数估计的流程如图 4 所示.

注 7　评价算法优劣的指标有算法收敛速度、收敛精度和计算量.辨识算法的计算量可用其乘法运算次数和加法运算次数的和表示[33].一次加法运算为一个 flop,一次乘法运算也为一个 flop.除法作为乘

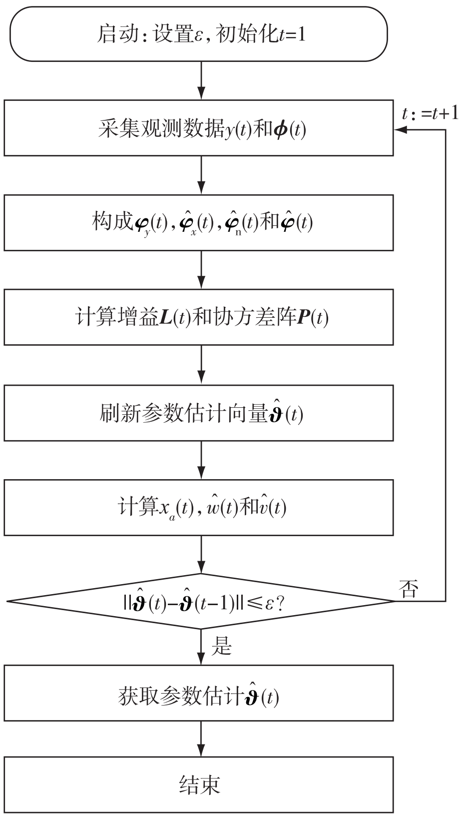

图 4　AM-RGELS 算法计算参数估计的流程

Fig. 4　The flowchart of the AM-RGELS algorithm

法对待，减法作为加法对待，就可以用 flop 数，即浮点运算数来表示计算量的大小.表 1 给出了 AM-RGELS 算法计算(39)—(49)每递推计算一步的计算量($n:=n_a+n_f+m+n_c+n_d$).

1.5　基于辅助模型的多新息广义增广最小二乘辨识算法

借鉴 AM-MI-GESG 算法(26)—(38)的推导，基于 AM-RGELS 算法(39)—(49)，将输出 $y(t)$ 和信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)\in\mathbf{R}^p$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)\in\mathbf{R}^{n\times p}$，将式(39)中标量新息 $y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$ 扩展为新息向量 $\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\cdot\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}^p$，可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的辅助模型多新息广义增广最小二乘算法(Auxiliary Model based Multi-Innovation Generalized Extended Least Squares algorithm，AM-MI-GELS 算法)：

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)],\tag{50}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1},\tag{51}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_n-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)]\boldsymbol{P}(t-1),\tag{52}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{53}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{54}$$

$$x_a(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{55}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t),\tag{56}$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t),\tag{57}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{58}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\tag{59}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_f)]^{\mathrm{T}},\tag{60}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{61}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{62}$$

当新息长度 $p=1$ 时，AM-MI-GELS 算法(50)—(62)退化为 AM-RGELS 算法(39)—(49).

注 8　在 AM-GESG 算法、AM-MI-GESG 算法、AM-RGELS 算法、AM-MI-GELS 算法中，如果将辅助

表 1　AM-RGELS 算法的计算量

Table 1　The computational efficiency of the AM-RGELS algorithm

变量	表达式	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)e(t)\in\mathbf{R}^n$	n	n
	$e(t):=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$	n	n
$\boldsymbol{L}(t)$	$\boldsymbol{L}(t)=\boldsymbol{\zeta}(t)/[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^n$	$2n$	n
	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)\in\mathbf{R}^n$	n^2	$n(n-1)$
$\boldsymbol{P}(t)$	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	n^2	n^2
$\hat{x}(t)$	$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)\in\mathbf{R}$	n_a+m	n_a+m-1
$\hat{w}(t)$	$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-x_a(t)\in\mathbf{R}$	n_f	n_f+1
$\hat{v}(t)$	$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\in\mathbf{R}$	n_c+n_d	n_c+n_d
总数		$2n^2+5n$	$2n^2+3n$
总 flop 数		$N_1:=4n^2+8n$	

模型的输出 $x_a(t)$ 作为 $x(t)$ 的估计 $\hat{x}(t)$,即 $\hat{x}(t)=x_a(t)$,则 $\boldsymbol{\varphi}_x(t)$ 的估计可以表示为

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f}.$$

注 9 新息长度的引入使得算法对观测数据和辨识新息的利用率提高了,多新息随机梯度算法提高参数估计精度效果是明显的,但是递推最小二乘算法本身的收敛速度很快,在最小二乘算法中引入新息长度的效果没有多新息随机梯度算法显著.当遇到数据缺失情况,采用变递推间隔方法时,多新息最小二乘算法的优越性就显现出来[1,34-36].

2 基于滤波的辅助模型多新息辨识方法

考虑式(1)描述的一类伪线性参数自回归滑动平均系统,重写如下:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t), \tag{63}$$

其中各变量定义同上.

用噪声模型传递函数 $H(z):=\frac{D(z)}{C(z)}$ 作为线性滤波器,对系统观测数据进行滤波处理,得到一个白噪声干扰下的输出误差模型.基于数据滤波技术,研究伪线性参数系统的辅助模型随机梯度算法、辅助模型多新息随机梯度算法、辅助模型递推最小二乘算法以及辅助模型多新息递推最小二乘算法.

2.1 基于数据滤波的辨识模型

定义中间变量 $x(t)$ 和相关噪声 $w(t)$:

$$x(t):=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}}{F(z)}=\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}, \tag{64}$$

$$w(t):=\frac{D(z)}{C(z)}v(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t). \tag{65}$$

定义滤波输出 $y_f(t)$,滤波中间变量 $x_f(t)$ 和滤波信息向量 $\boldsymbol{\phi}_f(t)$ 分别为

$$\begin{aligned}y_f(t):=&\frac{C(z)}{D(z)}y(t)=\\&y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\boldsymbol{c}+\\&[-y_f(t-1),-y_f(t-2),\cdots,-y_f(t-n_d)]\boldsymbol{d},\end{aligned} \tag{66}$$

$$\begin{aligned}x_f(t):=&\frac{C(z)}{D(z)}x(t)=\frac{C(z)}{D(z)}\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}}{F(z)}=\frac{\boldsymbol{\phi}_f^{\mathrm{T}}(t)\boldsymbol{\theta}}{F(z)}=\\&\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}_f^{\mathrm{T}}(t)\boldsymbol{\theta},\end{aligned} \tag{67}$$

$$\begin{aligned}\boldsymbol{\phi}_f(t):=&\frac{C(z)}{D(z)}\boldsymbol{\phi}(t)=\\&\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\boldsymbol{c}+\\&[-\boldsymbol{\phi}_f(t-1),-\boldsymbol{\phi}_f(t-2),\cdots,-\boldsymbol{\phi}_f(t-n_d)]\boldsymbol{d},\end{aligned} \tag{68}$$

其中定义参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 如下:

$$\boldsymbol{\vartheta}:=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{f}^{\mathrm{T}},\boldsymbol{\theta}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_0},\quad n_0:=n_a+n_f+m,$$

$$\boldsymbol{\theta}_{\mathrm{n}}:=[\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},$$

定义信息向量定义如下:

$$\boldsymbol{\varphi}(t):=[\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t),\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t),\boldsymbol{\phi}_f^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_0},$$

$$\boldsymbol{\varphi}_{yf}(t):=[-y_f(t-1),-y_f(t-2),\cdots,-y_f(t-n_a)]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$

$$\boldsymbol{\varphi}_{xf}(t):=[-x_f(t-1),-x_f(t-2),\cdots,-x_f(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f},$$

$$\boldsymbol{\varphi}_{\mathrm{n}}(t):=[\boldsymbol{\varphi}_w^{\mathrm{T}}(t),\boldsymbol{\varphi}_v^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{m}.$$

式(63)两边同乘 $\frac{C(z)}{D(z)}$ 得到

$$A(z)\frac{C(z)}{D(z)}y(t)=\frac{\frac{C(z)}{D(z)}\boldsymbol{\phi}^{\mathrm{T}}(t)}{F(z)}\boldsymbol{\theta}+v(t),$$

或

$$A(z)y_f(t)=x_f(t)+v(t),$$

或

$$\begin{aligned}y_f(t)=&[1-A(z)]y_f(t)+x_f(t)+v(t)\\=&\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t)\boldsymbol{a}+x_f(t)+v(t) &(69)\\=&\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t)\boldsymbol{a}+\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}_f^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t)\\=&\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}+v(t). &(70)\end{aligned}$$

式(65)和(70)是经过数据滤波后得到的两个辨识模型,它们包含了系统的所有参数.

2.2 基于滤波的辅助模型随机梯度辨识算法

根据辨识模型(70)和(65),定义两个准则函数:

$$J_3(\boldsymbol{\vartheta}):=\frac{1}{2}[y_f(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}]^2,$$

$$J_4(\boldsymbol{\theta}_{\mathrm{n}}):=\frac{1}{2}[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}]^2.$$

设 $\hat{\boldsymbol{\vartheta}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 为参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 在时刻 t 的估计,$\mu_1(t)$ 和 $\mu_2(t)$ 为步长.使用负梯度搜索,极小化准则函数 $J_3(\boldsymbol{\vartheta})$ 和 $J_4(\boldsymbol{\theta}_{\mathrm{n}})$ 可以得到下列递推关系式:

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(t)=&\hat{\boldsymbol{\vartheta}}(t-1)-\mu_1(t)\operatorname{grad}[J_3(\hat{\boldsymbol{\vartheta}}(t-1))]=\\&\hat{\boldsymbol{\vartheta}}(t-1)+\mu_1(t)\boldsymbol{\varphi}(t)[y_f(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\end{aligned} \tag{71}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=&\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)-\mu_2(t)\operatorname{grad}[J_4(\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1))]=\\&\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\mu_2(t)\boldsymbol{\varphi}_{\mathrm{n}}(t)[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)].\end{aligned} \tag{72}$$

定义新息 $e_1(t):=y_f(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$ 和 $e_2(t):=w(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$.取步长 $\mu_1(t):=1/r_1(t)$ 和 $\mu_2(t):=1/r_2(t)$,其中 $r_1(t):=r_1(t-1)+\|\boldsymbol{\varphi}(t)\|^2$ 和 $r_2(t):=r_2(t-1)+\|\boldsymbol{\varphi}_{\mathrm{n}}(t)\|^2$,由上两式可以得到

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r_1(t)}e_1(t),\tag{73}$$

$$e_1(t)=y_f(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{74}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r_1(0)=1,\tag{75}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\boldsymbol{\varphi}_{\mathrm{n}}(t)}{r_2(t)}e_2(t),\tag{76}$$

$$e_2(t)=w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\tag{77}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\varphi}_{\mathrm{n}}(t)\|^2,\quad r_2(0)=1.\tag{78}$$

算法(73)—(78)是不可实现的,因为信息向量$\boldsymbol{\varphi}(t)$中的$y_f(t-i)$,$x_f(t-i)$,$\boldsymbol{\phi}_f(t)$和$\boldsymbol{\varphi}_{\mathrm{n}}(t)$是未知的.若要计算这些未知变量,需要知道噪声模型的参数c_i和d_i,而这些参数是要估计的,只能采用其估计值进行计算.故在每次递推计算时,需要先辨识噪声模型的参数c_i和d_i,用其估计值$\hat{c}_i(t)$和$\hat{d}_i(t)$构造多项式$\hat{C}(t,z)$和$\hat{D}(t,z)$,用估计的传递函数$\hat{C}(t,z)/\hat{D}(t,z)$进行滤波,得到未知变量$y_f(t-i)$,$x_f(t-i)$和$\boldsymbol{\phi}_f(t)$的估计$\hat{y}_f(t-i)$,$\hat{x}_f(t-i)$和$\hat{\boldsymbol{\phi}}_f(t)$,进而推导和实现提出的算法.以上就是基于数据滤波辨识算法的基本原理.下面构造这些未知变量的估计.

定义信息向量$\hat{\boldsymbol{\varphi}}(t)$和$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$的估计如下:

$$\hat{\boldsymbol{\varphi}}(t):=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}},$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t):=[\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)]^{\mathrm{T}},$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t):=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t):=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f},$$

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

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

由式(70)得到$v(t)=y_f(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}$.未知变量$y_f(t)$,$\boldsymbol{\varphi}(t)$和$\boldsymbol{\vartheta}$用其估计$\hat{y}_f(t)$,$\hat{\boldsymbol{\varphi}}(t)$和$\hat{\boldsymbol{\vartheta}}(t)$代替,则$v(t)$的估计$\hat{v}(t)$可由下式计算:

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

或

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t).$$

用$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t):=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}$构造多项式$C(z)$和$D(z)$的估计:

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

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

由此可计算滤波后的变量$y_f(t)$,$x_f(t)$和向量$\boldsymbol{\phi}_f(t)$的估计:

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

$$\hat{x}_f(t):=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}\hat{x}(t)=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)}{\hat{F}(t,z)}=\frac{\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)}{\hat{F}(t,z)},$$

$$\hat{\boldsymbol{\phi}}_f(t):=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}\boldsymbol{\phi}(t).$$

由式(63)可以得到

$$w(t)=A(z)y(t)-x(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}-\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}-\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}.\tag{79}$$

将式(79)中未知的$\boldsymbol{a}$,$\boldsymbol{f}$,$\boldsymbol{\theta}$和$\boldsymbol{\varphi}_x(t)$分别用其估计$\hat{\boldsymbol{a}}(t-1)$,$\hat{\boldsymbol{f}}(t-1)$,$\hat{\boldsymbol{\theta}}(t-1)$和$\hat{\boldsymbol{\varphi}}_x(t)$代替,那么$w(t)$的估计$\hat{w}(t)$可以通过下式计算:

$$\hat{w}(t):=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1).$$

为使$\hat{\boldsymbol{\vartheta}}(t)$和$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$能递推计算,在$\hat{w}(t)$的表达式中采用了时刻$(t-1)$而非时刻$t$的估计.

根据式(64),$x(t)$的估计为

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),$$

因此,式(73)—(78)中涉及的未知变量$\boldsymbol{\varphi}(t)$,$y_f(t)$,$\boldsymbol{\varphi}_{\mathrm{n}}(t)$,$w(t)$等用其估计代替,可以得到辨识模型(70)和(65)参数向量$\boldsymbol{\vartheta}$和$\boldsymbol{\theta}_{\mathrm{n}}$的基于滤波的辅助模型随机梯度算法(Filtering based Auxiliary Model Stochastic Gradient algorithm,F-AM-SG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r_1(t)}e_1(t),\tag{80}$$

$$e_1(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{81}$$

$$r_1(t)=r_1(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\tag{82}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)}{r_2(t)}e_2(t),\tag{83}$$

$$e_2(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\tag{84}$$

$$r_2(t)=r_2(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\|^2,\tag{85}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{86}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t),\tag{87}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{88}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{89}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{90}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{91}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{92}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{93}$$

$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}$, (94)

$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}$, (95)

$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}}$, (96)

$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}}$, (97)

$\hat{\boldsymbol{\varphi}}_w(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}}$, (98)

$\hat{\boldsymbol{\varphi}}_v(t)=[\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}$, (99)

$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t),\hat{\boldsymbol{f}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)]^{\mathrm{T}}$, (100)

$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=[\hat{\boldsymbol{c}}^{\mathrm{T}}(t),\hat{\boldsymbol{d}}^{\mathrm{T}}(t)]^{\mathrm{T}}$. (101)

F-AM-SG 算法(80)—(101)的计算步骤如下:

1) 初始化:给定参数估计精度 ε,令 $t=1$,置参数估计初值 $\hat{\boldsymbol{\vartheta}}(0)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(0),\hat{\boldsymbol{f}}^{\mathrm{T}}(0),\hat{\boldsymbol{\theta}}^{\mathrm{T}}(0)]^{\mathrm{T}}=\mathbf{1}_{n_0}/p_0$, $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\mathbf{1}_{n_c+n_d}/p_0$,设置 $r_1(0)=1$, $r_2(0)=1$, $\hat{x}(i)=1/p_0$, $\hat{w}(i)=1/p_0$, $\hat{v}(i)=1/p_0$, $\hat{y}_f(i)=1/p_0$, $\hat{x}_f(i)=1/p_0$, $\hat{\boldsymbol{\phi}}_f(i)=\mathbf{1}_m/p_0$, $i=0,-1,-2,\cdots,1-\max[n_a,n_f,n_c,n_d]$, $p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,根据式(94)—(99)构造 $\boldsymbol{\varphi}_y(t)$, $\hat{\boldsymbol{\varphi}}_x(t)$, $\hat{\boldsymbol{\varphi}}_{yf}(t)$, $\hat{\boldsymbol{\varphi}}_{xf}(t)$, $\hat{\boldsymbol{\varphi}}_w(t)$ 和 $\hat{\boldsymbol{\varphi}}_v(t)$,根据式(93)构造 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,通过式(86)计算 $\hat{w}(t)$.

3) 根据式(84)—(85)计算 $e_2(t)$ 和 $r_2(t)$.

4) 通过式(83)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.

5) 从式(101)的 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 中读出 $\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$,根据式(90)—(91)计算 $\hat{y}_f(t)$ 和 $\hat{\boldsymbol{\phi}}_f(t)$,根据式(92)构造 $\hat{\boldsymbol{\varphi}}(t)$.

6) 通过式(81)—(82)计算 $e_1(t)$ 和 $r_1(t)$.

7) 通过式(80)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)$.

8) 从式(100)的 $\hat{\boldsymbol{\vartheta}}(t)$ 中读出 $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{f}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$.通过式(87)—(89)计算 $\hat{v}(t)$, $\hat{x}(t)$ 和 $\hat{x}_f(t)$.

9) 比较 $\hat{\boldsymbol{\vartheta}}(t)$ 和 $\hat{\boldsymbol{\vartheta}}(t-1)$, $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$,若 $\|\hat{\boldsymbol{\vartheta}}(t)-\hat{\boldsymbol{\vartheta}}(t-1)\|+\|\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)-\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\|\leqslant\varepsilon$,结束计算,得到参数估计 $\hat{\boldsymbol{\vartheta}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$,否则 t 增加 1 转到第 2)步,继续进行递推计算.

F-AM-SG 算法计算参数估计的流程如图 5 所示.

注 10 基于数据滤波的辅助模型随机梯度算法虽然能提高参数估计精度,但提高程度极其有限,为提高收敛速度,可以引入遗忘因子或收敛指数.

2.3 基于滤波的辅助模型多新息随机梯度辨识算法

根据多新息辨识理论,基于 F-AM-SG 算法(80)—(101),将信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,滤波输出 $\hat{y}_f(t)$ 和噪声项 $\hat{w}(t)$ 扩展为堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$ 和

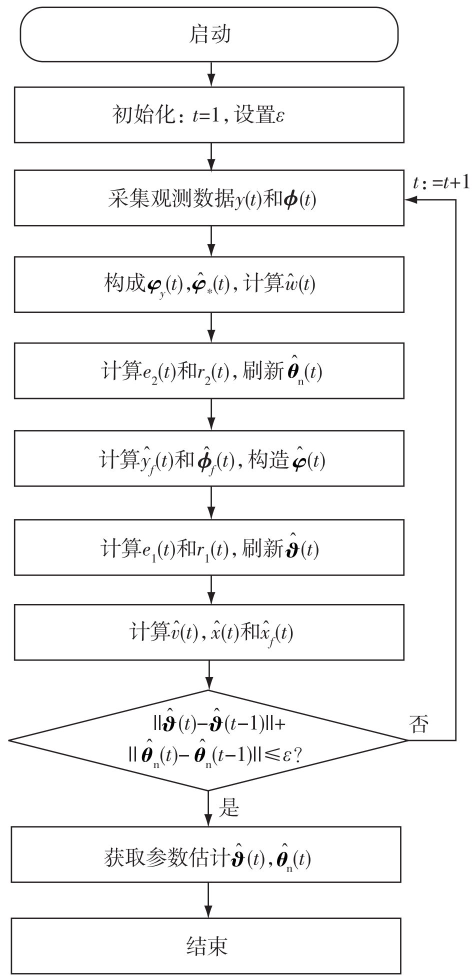

图 5 F-AM-SG 算法计算参数估计的流程

Fig. 5 The flowchart of the F-AM-SG algorithm

$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$,堆积滤波输出向量 $\hat{\boldsymbol{Y}}_f(p,t)$ 和堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$:

$$\hat{\boldsymbol{\Phi}}(p,t):=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)]\in\mathbf{R}^{n_0\times p},$$

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

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

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

将标量新息 $e_1(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$ 和 $e_2(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$ 扩展为新息向量:

$$\boldsymbol{E}_1(p,t):=\begin{bmatrix}\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\\ \hat{y}_f(t-1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\vartheta}}(t-1)\\ \vdots\\ \hat{y}_f(t-p+1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-1)\end{bmatrix}=$$

$$\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1),$$

$$\boldsymbol{E}_2(p,t):=\begin{bmatrix}\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\\ \hat{w}(t-1)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t-1)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\\ \vdots\\ \hat{w}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t-p+1)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\end{bmatrix}=$$

$$\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1),$$

可以得到辨识模型(70)和(65)参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\rm n}$ 的基于滤波的辅助模型多新息随机梯度算法(Filtering based Auxiliary Model Multi-Innovation Stochastic Gradient algorithm, F-AM-MISG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t),\tag{102}$$

$$\boldsymbol{E}_1(p,t)=\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}^{\rm T}(p,t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{103}$$

$$r_1(t)=r_1(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2,\tag{104}$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=\hat{\boldsymbol{\theta}}_{\rm n}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t),\tag{105}$$

$$\boldsymbol{E}_2(p,t)=\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1),\tag{106}$$

$$r_2(t)=r_2(t-1)+\|\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)\|^2,\tag{107}$$

$$\hat{\boldsymbol{Y}}_f(p,t)=[\hat{y}_f(t),\hat{y}_f(t-1),\cdots,\hat{y}_f(t-p+1)]^{\rm T},\tag{108}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{109}$$

$$\hat{\boldsymbol{W}}(p,t)=[\hat{w}(t),\hat{w}(t-1),\hat{w}(t-p+1)]^{\rm T},\tag{110}$$

$$\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)=[\hat{\boldsymbol{\varphi}}_{\rm n}(t),\hat{\boldsymbol{\varphi}}_{\rm n}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\rm n}(t-p+1)],\tag{111}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\rm T}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{112}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\rm T}(t)\hat{\boldsymbol{\vartheta}}(t),\tag{113}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{114}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{115}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{116}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{117}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\rm T}(t),\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t),\hat{\boldsymbol{\phi}}_f^{\rm T}(t)]^{\rm T},\tag{118}$$

$$\hat{\boldsymbol{\varphi}}_{\rm n}(t)=[\hat{\boldsymbol{\varphi}}_w^{\rm T}(t),\hat{\boldsymbol{\varphi}}_v^{\rm T}(t)]^{\rm T},\tag{119}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\rm T},\tag{120}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\rm T},\tag{121}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\rm T},\tag{122}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\rm T},\tag{123}$$

$$\hat{\boldsymbol{\varphi}}_w(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\rm T},\tag{124}$$

$$\hat{\boldsymbol{\varphi}}_v(t)=[\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\rm T},\tag{125}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{a}}^{\rm T}(t),\hat{\boldsymbol{f}}^{\rm T}(t),\hat{\boldsymbol{\theta}}^{\rm T}(t)]^{\rm T},\tag{126}$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=[\hat{\boldsymbol{c}}^{\rm T}(t),\hat{\boldsymbol{d}}^{\rm T}(t)]^{\rm T}.\tag{127}$$

F-AM-MISG 算法的计算步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p,选择数据长度 L,置参数估计初值 $\hat{\boldsymbol{\vartheta}}(0)=[\hat{\boldsymbol{a}}^{\rm T}(0),\hat{\boldsymbol{f}}^{\rm T}(0),\hat{\boldsymbol{\theta}}^{\rm T}(0)]^{\rm T}=\mathbf{1}_{n_0}/p_0$,$\hat{\boldsymbol{\theta}}_{\rm n}(0)=\mathbf{1}_{n_c+n_d}/p_0$,设置 $r_1(0)=1$,$r_2(0)=1$,$\hat{x}(i)=1/p_0$,$\hat{w}(i)=1/p_0$,$\hat{v}(i)=1/p_0$,$\hat{y}_f(i)=1/p_0$,$\hat{x}_f(i)=1/p_0$,$\hat{\boldsymbol{\phi}}_f(i)=\mathbf{1}_m/p_0$,$i=0,-1,-2,\cdots,1-\max[n_a,n_f,n_c,n_d]$,$p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,根据式(120)—(125)构造 $\boldsymbol{\varphi}_y(t)$,$\hat{\boldsymbol{\varphi}}_x(t)$,$\hat{\boldsymbol{\varphi}}_{yf}(t)$,$\hat{\boldsymbol{\varphi}}_{xf}(t)$,$\hat{\boldsymbol{\varphi}}_w(t)$ 和 $\hat{\boldsymbol{\varphi}}_v(t)$,根据式(119)构造 $\hat{\boldsymbol{\varphi}}_{\rm n}(t)$,通过式(112)计算 $\hat{w}(t)$,由式(110)—(111)构造堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)$.

3) 根据式(106)和(107)计算 $\boldsymbol{E}_2(p,t)$ 和 $r_2(t)$.

4) 通过式(105)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_{\rm n}(t)$.

5) 从式(127)的 $\hat{\boldsymbol{\theta}}_{\rm n}(t)$ 中读出 $\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$,根据式(116)—(117)计算 $\hat{y}_f(t)$ 和 $\hat{\boldsymbol{\phi}}_f(t)$,根据式(118)构造 $\hat{\boldsymbol{\varphi}}(t)$,由式(108)和(109)构造堆积滤波输出向量 $\hat{\boldsymbol{Y}}_f(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$.

6) 通过式(103)—(104)计算 $\boldsymbol{E}_1(p,t)$ 和 $r_1(t)$.

7) 通过式(102)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)$.

8) 从式(126)的 $\hat{\boldsymbol{\vartheta}}(t)$ 中读出 $\hat{\boldsymbol{a}}(t)$,$\hat{\boldsymbol{f}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$. 通过式(113)—(115)计算 $\hat{v}(t)$,$\hat{x}(t)$ 和 $\hat{x}_f(t)$.

9) 如果 $t<L$,t 增加 1 转到第 2)步,继续进行递推计算,否则获得参数估计向量 $\hat{\boldsymbol{\vartheta}}(L)$ 和 $\hat{\boldsymbol{\theta}}_{\rm n}(L)$.

注 11 为加快收敛速度,式(104)和(107)可修改为

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

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

当然也可引入收敛因子 λ 或收敛指数进一步提高算法的收敛速度.

2.4 基于滤波的辅助模型递推最小二乘辨识算法

对于辨识模型(70)和(65),极小化下列准则函数:

$$J_5(\boldsymbol{\vartheta}):=\sum_{j=1}^{t}[y_f(j)-\boldsymbol{\varphi}^{\rm T}(j)\boldsymbol{\vartheta}]^2,$$

$$J_6(\boldsymbol{\theta}_{\rm n}):=\sum_{j=1}^{t}[w(j)-\boldsymbol{\varphi}_{\rm n}^{\rm T}(j)\boldsymbol{\theta}_{\rm n}]^2.$$

仿照 AM-REGLS 辨识算法和 F-AM-SG 辨识算法的推导,涉及的未知变量用其估计代替,可以得到辨识模型(70)和(65)参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\rm n}$ 的基于滤波的辅助模型递推最小二乘算法(Filtering based Auxiliary Model Recursive Least Squares algorithm, F-AM-

RLS 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}(t)[\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (128)$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \quad (129)$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_{n_0}-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1), \quad (130)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)[\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)], \quad (131)$$

$$\boldsymbol{L}_{\mathrm{n}}(t)=\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)[1+\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_{\mathrm{n}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)]^{-1}, \quad (132)$$

$$\boldsymbol{P}_{\mathrm{n}}(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_{\mathrm{n}}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_{\mathrm{n}}(t-1), \quad (133)$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \quad (134)$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \quad (135)$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}_f(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \quad (136)$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \quad (137)$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \quad (138)$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \quad (139)$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}}, \quad (140)$$

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

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \quad (142)$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \quad (143)$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}}, \quad (144)$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}}, \quad (145)$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t),\hat{\boldsymbol{f}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \quad (146)$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=[\hat{\boldsymbol{c}}^{\mathrm{T}}(t),\hat{\boldsymbol{d}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \quad (147)$$

表 2 给出了 F-AM-RLS 算法每递推计算一步的计算量($n_0:=n_a+n_f+m, n:=n_0+n_c+n_d$).算法的总 flop 数为

$$\begin{aligned}N_2:=&4(n_a+n_f+m)^2+4(n_c+n_d)^2+2m(n_c+n_d)+12n-2n_a-1=\\&4(n_a+n_f+m+n_c+n_d)^2-8(n_a+n_f+m)(n_c+n_d)+\\&2m(n_c+n_d)+12n-2n_a-1=\\&4n^2-8(n_a+n_f+m+n_c+n_d)(n_c+n_d)+\\&8(n_c+n_d)^2+2m(n_c+n_d)+12n-2n_a-1=\\&4n^2-8(n_c+n_d)n+8(n_c+n_d)^2+\\&2m(n_c+n_d)+12n-2n_a-1=\end{aligned}$$

表 2 F-AM-RLS 算法的计算量

Table 1 The computational efficiency of the F-AM-RLS algorithm

变量	表达式	乘法次数	加法次数
$\hat{\boldsymbol{\vartheta}}(t)$	$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}(t)e(t)\in\mathbf{R}^{n_a+n_f+m}$	n_a+n_f+m	n_a+n_f+m
	$e(t):=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R}$	n_a+n_f+m	n_a+n_f+m
$\boldsymbol{L}(t)$	$\boldsymbol{L}(t)=\boldsymbol{\zeta}(t)/[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^{n_a+n_f+m}$	$2(n_a+n_f+m)$	n_a+n_f+m
	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)\in\mathbf{R}^{n_a+n_f+m}$	$(n_a+n_f+m)^2$	$(n_a+n_f+m-1)(n_a+n_f+m)$
$\boldsymbol{P}(t)$	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{(n_a+n_f+m)\times(n_a+n_f+m)}$	$(n_a+n_f+m)^2$	$(n_a+n_f+m)^2$
$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$	$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)e_{\mathrm{n}}(t)\in\mathbf{R}^{n_c+n_d}$	n_c+n_d	n_c+n_d
	$e_{\mathrm{n}}(t):=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$	n_c+n_d	n_c+n_d
$\boldsymbol{L}_{\mathrm{n}}(t)$	$\boldsymbol{L}_{\mathrm{n}}(t)=\boldsymbol{\zeta}_{\mathrm{n}}(t)/[1+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\zeta}_{\mathrm{n}}(t)]\in\mathbf{R}^{n_c+n_d}$	$2(n_c+n_d)$	n_c+n_d
	$\boldsymbol{\zeta}_{\mathrm{n}}(t):=\boldsymbol{P}_{\mathrm{n}}(t-1)\boldsymbol{\varphi}_{\mathrm{n}}(t)\in\mathbf{R}^{n_c+n_d}$	$(n_c+n_d)^2$	$(n_c+n_d)(n_c+n_d-1)$
$\boldsymbol{P}_{\mathrm{n}}(t)$	$\boldsymbol{P}_{\mathrm{n}}(t)=\boldsymbol{P}_{\mathrm{n}}(t-1)-\boldsymbol{L}_{\mathrm{n}}(t)\boldsymbol{\zeta}_{\mathrm{n}}^{\mathrm{T}}(t)\in\mathbf{R}^{(n_c+n_d)\times(n_c+n_d)}$	$(n_c+n_d)^2$	$(n_c+n_d)^2$
$\hat{w}(t)$	$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}$	n	n
$\hat{x}(t)$	$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)\in\mathbf{R}$	n_f+m	n_f+m-1
$\hat{v}(t)$	$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\in\mathbf{R}$	$n_a+n_c+n_d$	$n_a+n_c+n_d+1$
$\hat{x}_f(t)$	$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)\in\mathbf{R}$	n_f+m	n_f+m-1
$\hat{y}_f(t)$	$\hat{y}_f(t)=y(t)+[y(t-1),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t)\in\mathbf{R}$	n_c+n_d	n_c+n_d
$\hat{\boldsymbol{\phi}}_f(t)$	$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t)\in\mathbf{R}$	$m(n_c+n_d)$	$m(n_c+n_d)$
总数		$2(n_a+n_f+m)^2+2(n_c+n_d)^2+m(n_c+n_d)+7n-n_a$	$2(n_a+n_f+m)^2+2(n_c+n_d)^2+m(n_c+n_d)+5n-n_a-1$
总 flop 数		$N_2:=4(n_a+n_f+m)^2+4(n_c+n_d)^2+2m(n_c+n_d)+12n-2n_a-1$	

$4n^2+8(n_c+n_d)^2-(8n-2m)(n_c+n_d)+$
$12n-2n_a-1.$

AM-RGELS 算法与 F-AM-RLS 算法的计算量之差为

$$\begin{aligned}N_1-N_2=&(4n^2+8n)-[4n^2+8(n_c+n_d)^2-\\&(8n-2m)(n_c+n_d)+12n-2n_a-1]=\\&-8(n_c+n_d)^2-4n+(8n-2m)(n_c+n_d)+2n_a+1=\\&(8n-8n_c-8n_d-2m)(n_c+n_d)-4n+2n_a+1=\\&(8n_a+8n_f+6m)(n_c+n_d)-4n+2n_a+1=\\&4(n_a+n_f+m)(n_c+n_d)+(4n_a+4n_f+2m)(n_c+n_d)-\\&4(n_a+n_f+m+n_c+n_d)+2n_a+1=\\&4(n_a+n_f+m)(n_c+n_d-1)+\\&(4n_a+4n_f+2m-4)(n_c+n_d)+2n_a+1.\end{aligned}$$

当 $n_c+n_d\geqslant 1$，n_a 或 $n_f\geqslant 1$，或 $m\geqslant 2$ 时，$N_1-N_2\geqslant 0$，所以 F-AM-RLS 算法比 AM-RGELS 算法计算量小.

2.5 基于滤波的辅助模型多新息最小二乘辨识算法

针对辨识模型(70)和(65)，下面利用系统的观测数据$\{y(j),\boldsymbol{\phi}(j):0\leqslant j\leqslant t\}$，研究多新息最小二乘算法，对系统的未知参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\rm n}$进行估计.

定义堆积信息矩阵 $\boldsymbol{\Phi}(p,t)$ 和 $\boldsymbol{\Phi}_{\rm n}(p,t)$，堆积滤波输出向量 $\boldsymbol{Y}_f(p,t)$，堆积噪声向量 $\boldsymbol{W}(p,t)$ 和 $\boldsymbol{V}(p,t)$分别为

$$\boldsymbol{\Phi}(p,t):=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]\in\mathbf{R}^{n_0\times p},$$
$$\boldsymbol{\Phi}_{\rm n}(p,t):=[\boldsymbol{\varphi}_{\rm n}(t),\boldsymbol{\varphi}_{\rm n}(t-1),\cdots,\boldsymbol{\varphi}_{\rm n}(t-p+1)]\in\mathbf{R}^{(n_c+n_d)\times p},$$
$$\boldsymbol{Y}_f(p,t):=[y_f(t),y_f(t-1),\cdots,y_f(t-p+1)]^{\rm T}\in\mathbf{R}^p,$$
$$\boldsymbol{W}(p,t):=[w(t),w(t-1),\cdots,w(t-p+1)]^{\rm T}\in\mathbf{R}^p,$$
$$\boldsymbol{V}(p,t):=[v(t),v(t-1),\cdots,v(t-p+1)]^{\rm T}\in\mathbf{R}^p,$$

则由式(70)和(65)得到矩阵方程：

$$\boldsymbol{Y}_f(p,t)=\boldsymbol{\Phi}^{\rm T}(p,t)\boldsymbol{\vartheta}+\boldsymbol{V}(p,t),\tag{148}$$
$$\boldsymbol{W}(p,t)=\boldsymbol{\Phi}_{\rm n}^{\rm T}(p,t)\boldsymbol{\theta}_{\rm n}+\boldsymbol{V}(p,t).\tag{149}$$

上述两式称为基于数据滤波的多新息辨识模型.取准则函数为

$$J_7(\boldsymbol{\vartheta}):=\sum_{j=1}^{t}[\boldsymbol{Y}_f(p,j)-\boldsymbol{\Phi}^{\rm T}(p,j)\boldsymbol{\vartheta}]^{\rm T}[\boldsymbol{Y}_f(p,j)-\boldsymbol{\Phi}^{\rm T}(p,j)\boldsymbol{\vartheta}]=\sum_{j=1}^{t}\|\boldsymbol{Y}_f(p,j)-\boldsymbol{\Phi}^{\rm T}(p,j)\boldsymbol{\vartheta}\|^2,$$

$$J_8(\boldsymbol{\theta}_{\rm n}):=\sum_{j=1}^{t}[\boldsymbol{W}(p,j)-\boldsymbol{\Phi}_{\rm n}^{\rm T}(p,j)\boldsymbol{\theta}_{\rm n}]^{\rm T}[\boldsymbol{W}(p,j)-\boldsymbol{\Phi}_{\rm n}^{\rm T}(p,j)\boldsymbol{\theta}_{\rm n}]=\sum_{j=1}^{t}\|\boldsymbol{W}(p,j)-\boldsymbol{\Phi}_{\rm n}^{\rm T}(p,j)\boldsymbol{\theta}_{\rm n}\|^2.$$

令其对 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\rm n}$ 的偏导数为零，涉及的未知变量 $\boldsymbol{Y}_f(p,t)$，$\boldsymbol{\Phi}(p,t)$，$\boldsymbol{W}(p,t)$ 和 $\boldsymbol{\Phi}_{\rm n}(p,t)$ 分别用其估计 $\hat{\boldsymbol{Y}}_f(p,t)$，$\hat{\boldsymbol{\Phi}}(p,t)$，$\hat{\boldsymbol{W}}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)$ 代替，得到多新息最小二乘估计：

$$\hat{\boldsymbol{\vartheta}}(t)=\left[\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}(p,j)\hat{\boldsymbol{\Phi}}^{\rm T}(p,j)\right]^{-1}\left[\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}(p,j)\hat{\boldsymbol{Y}}_f(p,j)\right],$$
$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=\left[\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}_{\rm n}(p,j)\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,j)\right]^{-1}\left[\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}_{\rm n}(p,j)\hat{\boldsymbol{W}}(p,j)\right].$$

定义协方差阵：

$$\boldsymbol{P}^{-1}(t)=\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}(p,j)\hat{\boldsymbol{\Phi}}^{\rm T}(p,j)\in\mathbf{R}^{(n_a+n_f+m)\times(n_a+n_f+m)},$$
$$\boldsymbol{P}_{\rm n}^{-1}(t)=\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}_{\rm n}(p,j)\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,j)\in\mathbf{R}^{(n_c+n_d)\times(n_c+n_d)},$$

则有递推计算关系：

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\Phi}}(p,t)\hat{\boldsymbol{\Phi}}^{\rm T}(p,t),$$
$$\boldsymbol{P}_{\rm n}^{-1}(t)=\boldsymbol{P}_{\rm n}^{-1}(t-1)+\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t).$$

参照多新息最小二乘算法的推导[1]，并且应用矩阵求逆引理：

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

可以得到辨识模型(70)和(65)参数向量 $\boldsymbol{\vartheta}$ 和 $\boldsymbol{\theta}_{\rm n}$的基于滤波的辅助模型多新息最小二乘算法(Filtering based Auxiliary Model Muli-Innovation Least Squares algorithm，F-AM-MILS 算法)：

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}(t)[\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}^{\rm T}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)],\tag{150}$$
$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}^{\rm T}(p,t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(p,t)]^{-1},\tag{151}$$
$$\boldsymbol{P}(t)=[\boldsymbol{I}_{n_0}-\boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\rm T}(p,t)]\boldsymbol{P}(t-1),\tag{152}$$
$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=\hat{\boldsymbol{\theta}}_{\rm n}(t-1)+\boldsymbol{L}_{\rm n}(t)[\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)],\tag{153}$$
$$\boldsymbol{L}_{\rm n}(t)=\boldsymbol{P}_{\rm n}(t-1)\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)[\boldsymbol{I}_p+\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)\boldsymbol{P}_{\rm n}(t-1)\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)]^{-1},\tag{154}$$
$$\boldsymbol{P}_{\rm n}(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_{\rm n}(t)\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)]\boldsymbol{P}_{\rm n}(t-1),\tag{155}$$
$$\hat{\boldsymbol{Y}}_f(p,t)=[\hat{y}_f(t),\hat{y}_f(t-1),\cdots,\hat{y}_f(t-p+1)]^{\rm T},\tag{156}$$
$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{157}$$
$$\hat{\boldsymbol{W}}(p,t)=[\hat{w}(t),\hat{w}(t-1),\hat{w}(t-p+1)]^{\rm T},\tag{158}$$
$$\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)=[\hat{\boldsymbol{\varphi}}_{\rm n}(t),\hat{\boldsymbol{\varphi}}_{\rm n}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\rm n}(t-p+1)],\tag{159}$$
$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\rm T}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{160}$$
$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}^{\rm T}(t)\hat{\boldsymbol{\vartheta}}(t),\tag{161}$$
$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{162}$$
$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{163}$$
$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{164}$$
$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{165}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_{f}^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{166}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{167}$$

$$\boldsymbol{\varphi}_{y}(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\tag{168}$$

$$\hat{\boldsymbol{\varphi}}_{x}(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}},\tag{169}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}},\tag{170}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}},\tag{171}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t),\hat{\boldsymbol{f}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{172}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=[\hat{\boldsymbol{c}}^{\mathrm{T}}(t),\hat{\boldsymbol{d}}^{\mathrm{T}}(t)]^{\mathrm{T}}.\tag{173}$$

3 基于模型分解的辅助模型多新息辨识方法

考虑式(63)描述的一类伪线性参数自回归滑动平均系统,重写如下:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t),\tag{174}$$

其中各变量定义同上.

当系统(174)的参数数目很多时,最小二乘算法协方差矩阵维数大,其计算量也大.与前述计算量很大的 AM-RGELS 算法相比,尽管基于观测数据滤波的辅助模型递推最小二乘(F-AM-RLS)算法的计算量有所降低,但降低程度是有限的.下面利用辨识模型分解,研究基于模型分解的随机梯度算法、基于模型分解的多新息随机梯度算法、基于模型分解的递推最小二乘算法、基于模型分解的多新息最小二乘算法.特别是基于模型分解的递推最小二乘算法比 F-AM-RLS 算法的计算量还要小(是指同类算法间的比较).

3.1 基于模型分解的辨识模型

定义中间变量 $x(t)$ 和相关噪声 $w(t)$:

$$x(t):=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}}{F(z)}=\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta},$$

$$w(t):=\frac{D(z)}{C(z)}v(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\tag{175}$$

式(174)可以写为

$$y(t)=[1-A(z)]y(t)+x(t)+w(t)$$

$$=\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}+x(t)+w(t)\tag{176}$$

$$=\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}+\boldsymbol{\varphi}_x^{\mathrm{T}}(t)\boldsymbol{f}+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\tag{177}$$

这个辨识模型包含了 4 个子信息向量 $\boldsymbol{\varphi}_y(t)$,$\boldsymbol{\varphi}_x(t)$,$\boldsymbol{\phi}(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$,其中 $\boldsymbol{\varphi}_y(t)$ 和 $\boldsymbol{\phi}(t)$ 是由已知观测数据构成的,$\boldsymbol{\varphi}_x(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 是未知的.将信息向量分为已知的和未知的两组,定义如下:

$$\boldsymbol{\varphi}_2(t):=[\boldsymbol{\varphi}_y^{\mathrm{T}}(t),\boldsymbol{\phi}^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_a+m},$$

$$\boldsymbol{\varphi}_3(t):=[\boldsymbol{\varphi}_x^{\mathrm{T}}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_f+n_c+n_d}.$$

对应的参数向量也分为两组:

$$\boldsymbol{\theta}_2:=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{\theta}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_a+m},$$

$$\boldsymbol{\theta}_3:=[\boldsymbol{f}^{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{n}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_f+n_c+n_d}.$$

借助于上述定义,式(177)可以写为

$$y(t)=\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\theta}_2+\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\boldsymbol{\theta}_3+v(t).\tag{178}$$

引入两个中间变量:

$$y_2(t):=y(t)-\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\boldsymbol{\theta}_3,$$

$$y_3(t):=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\theta}_2.$$

系统(178)可以被分解为两个虚拟子系统:

$$y_2(t)=\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\theta}_2+v(t),\tag{179}$$

$$y_3(t)=\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\boldsymbol{\theta}_3+v(t).\tag{180}$$

式(179)和(180)为分解后的辨识模型,如图 6 所示.

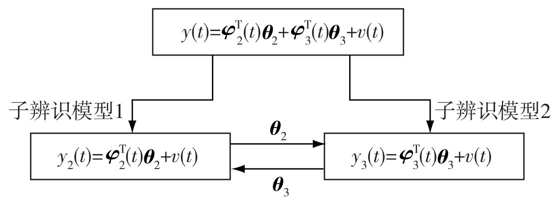

图 6 辨识模型分解为子辨识模型的递阶结构

Fig. 6 The hierarchical structure of the identification model

3.2 基于模型分解的随机梯度辨识算法

根据辨识模型(179)和(180),定义两个准则函数:

$$J_9(\boldsymbol{\theta}_2):=\frac{1}{2}[y_2(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\theta}_2]^2,$$

$$J_{10}(\boldsymbol{\theta}_3):=\frac{1}{2}[y_3(t)-\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\boldsymbol{\theta}_3]^2.$$

设 $\hat{\boldsymbol{\theta}}_2(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{\theta}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+m}$ 和 $\hat{\boldsymbol{\theta}}_3(t):=\begin{bmatrix}\hat{\boldsymbol{f}}(t)\\\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_f+n_c+n_d}$ 分别是 $\boldsymbol{\theta}_2=\begin{bmatrix}\boldsymbol{a}\\\boldsymbol{\theta}\end{bmatrix}\in\mathbf{R}^{n_a+m}$ 和 $\boldsymbol{\theta}_3=\begin{bmatrix}\boldsymbol{f}\\\boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^{n_f+n_c+n_d}$ 在时刻 t 的估计.使用负梯度搜索,极小化准则函数 $J_9(\boldsymbol{\theta}_2)$ 和 $J_{10}(\boldsymbol{\theta}_3)$ 可以得到下列递推关系:

$$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)-\frac{1}{r_2(t)}\mathrm{grad}[J_9(\hat{\boldsymbol{\theta}}_2(t-1))=$$

$$\hat{\boldsymbol{\theta}}_2(t-1)+\frac{\boldsymbol{\varphi}_2(t)}{r_2(t)}[y_2(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)]=$$

$$\hat{\boldsymbol{\theta}}_2(t-1)+\frac{\boldsymbol{\varphi}_2(t)}{r_2(t)}[y(t)-\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\boldsymbol{\theta}_3-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)],\tag{181}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\varphi}_2(t)\|^2,\quad r_2(0)=1,\tag{182}$$

$$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)-\frac{1}{r_3(t)}\text{grad}[J_{10}(\hat{\boldsymbol{\theta}}_3(t-1))=\hat{\boldsymbol{\theta}}_3(t-1)+\frac{\boldsymbol{\varphi}_3(t)}{r_3(t)}[y_3(t)-\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)]=\hat{\boldsymbol{\theta}}_3(t-1)+\frac{\boldsymbol{\varphi}_3(t)}{r_3(t)}[y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\theta}_2-\boldsymbol{\varphi}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)], \tag{183}$$

$$r_3(t)=r_3(t-1)+\|\boldsymbol{\varphi}_3(t)\|^2. \tag{184}$$

式(181)—(184)右边包含的未知参数向量 $\boldsymbol{\theta}_3$ 和 $\boldsymbol{\theta}_2$,未知信息向量 $\boldsymbol{\varphi}_3(t)$ 分别用其估计 $\hat{\boldsymbol{\theta}}_3(t-1)$,$\hat{\boldsymbol{\theta}}_2(t-1)$ 和 $\hat{\boldsymbol{\varphi}}_3(t)$ 代替,定义新息

$$e(t):=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R},$$

则可以得到估计参数向量 $\boldsymbol{\theta}_2$ 和 $\boldsymbol{\theta}_3$ 的基于模型分解的随机梯度算法(model Decomposition based Stochastic Gradient algorithm, D-SG 算法):

$$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)+\frac{\boldsymbol{\varphi}_2(t)}{r_2(t)}e(t), \tag{185}$$

$$e(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1), \tag{186}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\varphi}_2(t)\|^2, \tag{187}$$

$$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)+\frac{\hat{\boldsymbol{\varphi}}_3(t)}{r_3(t)}e(t), \tag{188}$$

$$r_3(t)=r_3(t-1)+\|\hat{\boldsymbol{\varphi}}_3(t)\|^2, \tag{189}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{190}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t), \tag{191}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t), \tag{192}$$

$$\boldsymbol{\varphi}_2(t)=\begin{bmatrix}\boldsymbol{\varphi}_y(t)\\ \boldsymbol{\phi}(t)\end{bmatrix},\quad \hat{\boldsymbol{\varphi}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_x(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}, \tag{193}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{194}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \tag{195}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{196}$$

$$\hat{\boldsymbol{\theta}}_2(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{197}$$

D-SG 算法(185)—(197)的计算步骤如下:

1) 初始化:给定参数估计精度 ε,令 $t=1$,置参数估计初值 $\hat{\boldsymbol{\theta}}_2(0)=\mathbf{1}_{n_a+m}/p_0$,$\hat{\boldsymbol{\theta}}_3(0)=\mathbf{1}_{n_f+n_c+n_d}/p_0$,设置 $r_2(0)=1$,$r_3(0)=1$,$\hat{x}(-i)=1/p_0$,$\hat{w}(-i)=1/p_0$,$\hat{v}(-i)=1/p_0$,$i=0,1,2,\cdots,\max[n_a,n_f,n_c,n_d]$,$p_0=10^6$.

2) 采集观测数据 $y(t)$ 和 $\boldsymbol{\phi}(t)$,由式(194)—(196)构成信息向量 $\boldsymbol{\varphi}_y(t)$,$\hat{\boldsymbol{\varphi}}_x(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,由式(193)构造信息向量 $\boldsymbol{\varphi}_2(t)$ 和 $\hat{\boldsymbol{\varphi}}_3(t)$.

3) 用式(186)计算新息 $e(t)$,用式(187)计算 $r_2(t)$,用式(185)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_2(t)$.

4) 用式(189)计算 $r_3(t)$,用式(188)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_3(t)$.

5) 从式(197)的 $\hat{\boldsymbol{\theta}}_2(t)$ 中读取 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{\theta}}(t)$,从式(197)的 $\hat{\boldsymbol{\theta}}_3(t)$ 中读取 $\hat{\boldsymbol{f}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$.用式(190)—(192)计算 $\hat{x}(t)$,$\hat{w}(t)$ 和 $\hat{v}(t)$.

6) 比较 $\hat{\boldsymbol{\theta}}_2(t)$ 和 $\hat{\boldsymbol{\theta}}_2(t-1)$,$\hat{\boldsymbol{\theta}}_3(t)$ 和 $\hat{\boldsymbol{\theta}}_3(t-1)$,若 $\|\hat{\boldsymbol{\theta}}_2(t)-\hat{\boldsymbol{\theta}}_2(t-1)\|+\|\hat{\boldsymbol{\theta}}_3(t)-\hat{\boldsymbol{\theta}}_3(t-1)\|\leqslant\varepsilon$,结束计算,得到参数估计 $\hat{\boldsymbol{\theta}}_2(t)$ 和 $\hat{\boldsymbol{\theta}}_3(t)$,否则 t 增加 1 转到第 2)步,继续进行递推计算.

注 12 同样可以引入遗忘因子 λ 来提高算法的收敛速度,将(188)和(189)修改为

$$r_2(t)=\lambda r_2(t-1)+\|\boldsymbol{\varphi}_2(t)\|^2,\quad 0\leqslant\lambda\leqslant1,\quad r_2(0)=1,$$

$$r_3(t)=\lambda r_3(t-1)+\|\hat{\boldsymbol{\varphi}}_3(t)\|^2,\quad r_3(0)=1.$$

3.3 基于模型分解的多新息随机梯度辨识算法

根据多新息辨识理论,基于 D-SG 算法(185)—(197),将系统输出 $y(t)$,信息向量 $\boldsymbol{\varphi}_2(t)$ 和 $\hat{\boldsymbol{\varphi}}_3(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_2(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_3(p,t)$:

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

$$\boldsymbol{\Phi}_2(p,t):=[\boldsymbol{\varphi}_2(t),\boldsymbol{\varphi}_2(t-1),\cdots,\boldsymbol{\varphi}_2(t-p+1)]\in\mathbf{R}^{(n_a+m)\times p},$$

$$\hat{\boldsymbol{\Phi}}_3(p,t):=[\hat{\boldsymbol{\varphi}}_3(t),\hat{\boldsymbol{\varphi}}_3(t-1),\cdots,\hat{\boldsymbol{\varphi}}_3(t-p+1)]\in\mathbf{R}^{(n_f+n_c+n_d)\times p},$$

将标量新息

$$e(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R}$$

扩展为新息向量

$$\boldsymbol{E}(p,t):=\begin{bmatrix}y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)\\ y(t-1)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}_3(t-1)\\ \vdots\\ y(t-p+1)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}_3(t-1)\end{bmatrix}=$$

$$\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_2^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\Phi}}_3^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R}^p, \tag{198}$$

可以得到估计参数向量 $\boldsymbol{\theta}_2$ 和 $\boldsymbol{\theta}_3$ 的基于模型分解的多新息随机梯度算法(model Decomposition based Multi-Innovation Stochastic Gradient algorithm, D-MISG 算法):

$$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)+\frac{\boldsymbol{\Phi}_2(p,t)}{r_2(t)}\boldsymbol{E}(p,t), \tag{199}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_2^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_2(t-1)-$$

$$\hat{\boldsymbol{\Phi}}_3^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_3(t-1), \tag{200}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\Phi}_2(p,t)\|^2, \tag{201}$$

$$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)+\frac{\hat{\boldsymbol{\Phi}}_3(p,t)}{r_3(t)}\boldsymbol{E}(p,t), \tag{202}$$

$$r_3(t)=r_3(t-1)+\|\hat{\boldsymbol{\Phi}}_3(p,t)\|^2, \tag{203}$$

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

$$\boldsymbol{\Phi}_2(p,t)=[\boldsymbol{\varphi}_2(t),\boldsymbol{\varphi}_2(t-1),\cdots,\boldsymbol{\varphi}_2(t-p+1)], \tag{205}$$

$$\hat{\boldsymbol{\Phi}}_3(p,t)=[\hat{\boldsymbol{\varphi}}_3(t),\hat{\boldsymbol{\varphi}}_3(t-1),\cdots,\hat{\boldsymbol{\varphi}}_3(t-p+1)], \tag{206}$$

$$\boldsymbol{\varphi}_2(t)=\begin{bmatrix}\boldsymbol{\varphi}_y(t)\\ \boldsymbol{\phi}(t)\end{bmatrix},\quad \hat{\boldsymbol{\varphi}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_x(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}, \tag{207}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{208}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t), \tag{209}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t), \tag{210}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{211}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \tag{212}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{213}$$

$$\hat{\boldsymbol{\theta}}_2(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{214}$$

注 13　为改进随机梯度辨识算法的收敛速度,引入新息长度推导出多新息随机梯度算法,多新息随机梯度算法是在收敛速度和计算量之间找到的折中,较最小二乘算法避免了协方差的计算,较随机梯度算法提高了算法的收敛速度.式(201)和(203)可以修改为

$$r_2(t)=\lambda\, r_2(t-1)+\|\boldsymbol{\varphi}_2(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r_2(0)=1,$$

$$r_3(t)=\lambda\, r_3(t-1)+\|\hat{\boldsymbol{\varphi}}_3(t)\|^2,\quad r_3(0)=1.$$

3.4　基于模型分解的递推最小二乘辨识算法

根据辨识模型(179)和(180),定义和极小化最小二乘准则函数:

$$J_{11}(\boldsymbol{\theta}_2):=\sum_{j=1}^{t}[y_2(j)-\boldsymbol{\varphi}_2^{\mathrm{T}}(j)\boldsymbol{\theta}_2]^2,$$

$$J_{12}(\boldsymbol{\theta}_3):=\sum_{j=1}^{t}[y_3(j)-\boldsymbol{\varphi}_3^{\mathrm{T}}(j)\boldsymbol{\theta}_3]^2.$$

仿照递推最小二乘辨识算法和基于模型分解的随机梯度算法的推导,涉及的未知变量用其估计代替,可以得到辨识模型(178)参数向量 $\boldsymbol{\theta}_2$ 和 $\boldsymbol{\theta}_3$ 的基于模型分解的递推最小二乘算法(model Decomposition based Recursive Least Squares algorithm,D-RLS 算法):

$$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)+\boldsymbol{L}_2(t)e(t), \tag{215}$$

$$e(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1), \tag{216}$$

$$\boldsymbol{L}_2(t)=\boldsymbol{P}_2(t-1)\boldsymbol{\varphi}_2(t)[1+\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{P}_2(t-1)\boldsymbol{\varphi}_2(t)]^{-1}, \tag{217}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{n_a+m}-\boldsymbol{L}_2(t)\boldsymbol{\varphi}_2^{\mathrm{T}}(t)]\boldsymbol{P}_2(t-1), \tag{218}$$

$$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)+\boldsymbol{L}_3(t)e(t), \tag{219}$$

$$\boldsymbol{L}_3(t)=\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_3(t)[1+\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_3(t)]^{-1}, \tag{220}$$

$$\boldsymbol{P}_3(t)=[\boldsymbol{I}_{n_f+n_c+n_d}-\boldsymbol{L}_3(t)\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)]\boldsymbol{P}_3(t-1), \tag{221}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{222}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t), \tag{223}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{224}$$

$$\boldsymbol{\varphi}_2(t)=\begin{bmatrix}\boldsymbol{\varphi}_y(t)\\ \boldsymbol{\phi}(t)\end{bmatrix},\quad \hat{\boldsymbol{\varphi}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_x(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}, \tag{225}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{226}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \tag{227}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\ \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{228}$$

$$\hat{\boldsymbol{\theta}}_2(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}. \tag{229}$$

表 3 列出 D-RLS 算法(215)—(229)每步的计算量($n:=n_a+m+n_f+n_c+n_d$).算法的总 flop 数为

$$N_3:=4(n_a+m)^2+4(n_f+n_c+n_d)^2+8n=4n^2-8(n_a+m)(n_f+n_c+n_d)+8n.$$

AM-RGELS 算法与 D-RLS 算法的计算量之差为

$$N_1-N_3=(4n^2+8n)-[4n^2-8(n_a+m)(n_f+n_c+n_d)+8n]=8(n_a+m)(n_f+n_c+n_d)>0.$$

因此,D-RLS 算法比 AM-RGELS 算法计算量小.当 $n_a=n_f=m=n_c=n_d=10$ 时,$N_1=10\ 400$ flops,$N_2=5\ 600$ flops,$N_1-N_3=4\ 800$ flops,$\frac{N_1-N_3}{N_1}=\frac{4\ 800}{10\ 400}\approx 46.15\%$,对于这个例子,D-RLS 算法比 AM-RGELS 算法计算量降低了约 46%.

3.5　基于模型分解的多新息最小二乘辨识算法

借鉴 D-MISG 算法的推导(199)—(214),基于 D-RLS 算法(215)—(229),将系统输出 $y(t)$,信息向量 $\boldsymbol{\varphi}_2(t)$ 和 $\hat{\boldsymbol{\varphi}}_3(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}_2(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_3(p,t)$,将标量新息

$$e(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R}$$

扩展为新息向量

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_2^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\Phi}}_3^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R}^p,$$

表 3 D-RLS 算法的计算量

Table 3 The computational efficiency of the D-RLS algorithm

变量	表达式	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}_2(t)$	$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)+\boldsymbol{L}_2(t)e(t)\in\mathbf{R}^{n_a+m}$	n_a+m	n_a+m
	$e(t)=\hat{y}_2(t)-\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t-1)\in\mathbf{R}$	n	n
$\boldsymbol{L}_2(t)$	$\boldsymbol{L}_2(t)=\boldsymbol{\zeta}_2(t)/[1+\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\boldsymbol{\zeta}_2(t)]\in\mathbf{R}^{n_a+m}$	$2(n_a+m)$	n_a+m
	$\boldsymbol{\zeta}_2(t):=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)\in\mathbf{R}^{n_a+m}$	$(n_a+m)^2$	$(n_a+m)(n_a+m-1)$
$\boldsymbol{P}_2(t)$	$\boldsymbol{P}_2(t)=\boldsymbol{P}_2(t-1)-\boldsymbol{L}_2(t)\boldsymbol{\zeta}_2^{\mathrm{T}}(t)\in\mathbf{R}^{(n_a+m)\times(n_a+m)}$	$(n_a+m)^2$	$(n_a+m)^2$
$\hat{\boldsymbol{\theta}}_3(t)$	$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)+\boldsymbol{L}_3(t)e(t)\in\mathbf{R}^{n_f+n_c+n_d}$	$n_f+n_c+n_d$	$n_f+n_c+n_d$
$\boldsymbol{L}_3(t)$	$\boldsymbol{L}_3(t)=\boldsymbol{\zeta}_3(t)/[1+\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\boldsymbol{\zeta}_3(t)]\in\mathbf{R}^{n_f+n_c+n_d}$	$2(n_f+n_c+n_d)$	$n_f+n_c+n_d$
	$\boldsymbol{\zeta}_3(t):=\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_3(t)\in\mathbf{R}^{n_f+n_c+n_d}$	$(n_f+n_c+n_d)^2$	$(n_f+n_c+n_d-1)(n_f+n_c+n_d)$
$\boldsymbol{P}_3(t)$	$\boldsymbol{P}_3(t)=\boldsymbol{P}_3(t-1)-\boldsymbol{L}_3(t)\boldsymbol{\zeta}_3^{\mathrm{T}}(t)\in\mathbf{R}^{(n_f+n_c+n_d)\times(n_f+n_c+n_d)}$	$(n_f+n_c+n_d)^2$	$(n_f+n_c+n_d)^2$
$\hat{x}(t)$	$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)\in\mathbf{R}$	n_f+m	n_f+m-1
$\hat{w}(t)$	$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)\in\mathbf{R}$	n_a	n_a+1
$\hat{v}(t)$	$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\in\mathbf{R}$	n_c+n_d	n_c+n_d
总数		$2(n_a+m)^2+2(n_f+n_c+n_d)^2+5n$	$2(n_a+m)^2+2(n_f+n_c+n_d)^2+3n$
总 flop 数		$N_3:=4(n_a+m)^2+4(n_f+n_c+n_d)^2+8n$	

可以得到估计参数向量 $\boldsymbol{\theta}_2$ 和 $\boldsymbol{\theta}_3$ 的基于模型分解的多新息最小二乘算法(model Decomposition based Multi-Innovation Least Squares algorithm, D-MILS 算法):

$$\hat{\boldsymbol{\theta}}_2(t)=\hat{\boldsymbol{\theta}}_2(t-1)+\boldsymbol{P}_2(t)\boldsymbol{\Phi}_2(p,t)\boldsymbol{E}(p,t),\tag{230}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Phi}_2^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_2(t-1)-\hat{\boldsymbol{\Phi}}_3^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_3(t-1),\tag{231}$$

$$\boldsymbol{P}_2^{-1}(t)=\boldsymbol{P}_2^{-1}(t-1)+\boldsymbol{\Phi}_2(p,t)\boldsymbol{\Phi}_2^{\mathrm{T}}(p,t),\tag{232}$$

$$\hat{\boldsymbol{\theta}}_3(t)=\hat{\boldsymbol{\theta}}_3(t-1)+\boldsymbol{P}_3(t)\hat{\boldsymbol{\Phi}}_3(p,t)\boldsymbol{E}(p,t),\tag{233}$$

$$\boldsymbol{P}_3^{-1}(t)=\boldsymbol{P}_3^{-1}(t-1)+\hat{\boldsymbol{\Phi}}_3(p,t)\hat{\boldsymbol{\Phi}}_3^{\mathrm{T}}(p,t),\tag{234}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{235}$$

$$\boldsymbol{\Phi}_2(p,t)=[\boldsymbol{\varphi}_2(t),\boldsymbol{\varphi}_2(t-1),\cdots,\boldsymbol{\varphi}_2(t-p+1)],\tag{236}$$

$$\hat{\boldsymbol{\Phi}}_3(p,t)=[\hat{\boldsymbol{\varphi}}_3(t),\hat{\boldsymbol{\varphi}}_3(t-1),\cdots,\hat{\boldsymbol{\varphi}}_3(t-p+1)],\tag{237}$$

$$\boldsymbol{\varphi}_2(t)=\begin{bmatrix}\boldsymbol{\varphi}_y(t)\\ \boldsymbol{\phi}(t)\end{bmatrix},\quad \hat{\boldsymbol{\varphi}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{\varphi}}_x(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix},\tag{238}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{239}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t),\tag{240}$$

$$\hat{v}(t)=y(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_2(t)-\hat{\boldsymbol{\varphi}}_3^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_3(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\tag{241}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\tag{242}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}},\tag{243}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{244}$$

$$\hat{\boldsymbol{\theta}}_2(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_3(t)=\begin{bmatrix}\hat{\boldsymbol{f}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}.\tag{245}$$

注 14 对于最小二乘算法而言,将分解思想运用到辨识算法,主要减小了算法中协方差矩阵的维数,从而减少了计算量,提高了计算效率.

4 基于滤波的分解多新息辨识方法

考虑式(63)描述的一类伪线性参数自回归滑动平均系统,重写如下:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\mathrm{T}}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t),\tag{246}$$

其中各变量定义同上.

为进一步减小计算量,这里利用观测数据滤波技术和辨识模型分解技术,推导基于滤波的分解随机梯度辨识算法、基于滤波的分解多新息随机梯度辨识算法、基于滤波的分解递推最小二乘辨识算法、基于滤波的分解多新息最小二乘辨识算法.思路是参照第 2 节的信息滤波方法和第 3 节的辨识模型分解方法.

4.1 基于滤波的分解辨识模型

采用式(64)—(70)的定义和推导.重写滤波辨识模型(70)和(65)如下:

$$y_f(t)=\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t)\boldsymbol{a}+\boldsymbol{\phi}_f^{\mathrm{T}}(t)\boldsymbol{\theta}+\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}+v(t)=[\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t),\boldsymbol{\phi}_f^{\mathrm{T}}(t)]\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{\theta}\end{bmatrix}+\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}+v(t),\tag{247}$$

$$w(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\tag{248}$$

定义参数向量：

$$\boldsymbol{\vartheta}:=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{\theta}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_a+m},$$

$$\boldsymbol{f}:=[f_1,f_2,\cdots,f_{n_f}]^{\mathrm{T}}\in\mathbf{R}^{n_f}$$

和信息向量：

$$\boldsymbol{\varphi}_4(t):=[\boldsymbol{\varphi}_{yf}^{\mathrm{T}}(t),\boldsymbol{\phi}_f^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_a+m},$$

$$\boldsymbol{\varphi}_{yf}(t):=[-y_f(t-1),-y_f(t-2),\cdots,-y_f(t-n_a)]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$

$$\boldsymbol{\varphi}_{xf}(t):=[-x_f(t-1),-x_f(t-2),\cdots,-x_f(t-n_f)]^{\mathrm{T}}\in\mathbf{R}^{n_f}.$$

基于递阶辨识原理，将滤波后的系统模型(247)分解成两个子模型：一个包含了参数向量 $\boldsymbol{\vartheta}$，另一个包含参数向量 $\boldsymbol{f}$.定义中间变量：

$$y_4(t):=y_f(t)-\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f},$$

$$y_5(t):=y_f(t)-\boldsymbol{\varphi}_4^{\mathrm{T}}(t)\boldsymbol{\vartheta},$$

则式(247)可以分解下列两个辨识子系统：

$$\mathrm{S}_1:\quad y_4(t)=\boldsymbol{\varphi}_4^{\mathrm{T}}(t)\boldsymbol{\vartheta}+v(t),\tag{249}$$

$$\mathrm{S}_2:\quad y_5(t)=\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}+v(t).\tag{250}$$

式(249)—(250)和(248)构成了基于滤波的分解辨识模型，它们包含了系统的所有参数 $\boldsymbol{\vartheta}=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{\theta}^{\mathrm{T}}]^{\mathrm{T}}$，$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}=[\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}]^{\mathrm{T}}$.

4.2 基于滤波的分解随机梯度辨识算法

根据辨识模型式(249)—(250)和(248)，定义三个准则函数：

$$J_{13}(\boldsymbol{\vartheta}):=\frac{1}{2}[y_4(t)-\boldsymbol{\varphi}_4^{\mathrm{T}}(t)\boldsymbol{\vartheta}]^2,$$

$$J_{14}(\boldsymbol{f}):=\frac{1}{2}[y_5(t)-\boldsymbol{\varphi}_{xf}^{\mathrm{T}}(t)\boldsymbol{f}]^2,$$

$$J_{15}(\boldsymbol{\theta}_{\mathrm{n}}):=\frac{1}{2}[w(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}]^2.$$

设 $\hat{\boldsymbol{\vartheta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{\theta}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+m}$，$\hat{\boldsymbol{f}}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t):=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\\hat{\boldsymbol{d}}(t)\end{bmatrix}$ 分别为参数向量 $\boldsymbol{\vartheta}=\begin{bmatrix}\boldsymbol{a}\\\boldsymbol{\theta}\end{bmatrix}$，$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}=\begin{bmatrix}\boldsymbol{c}\\\boldsymbol{d}\end{bmatrix}$ 在时刻 t 的估计.使用负梯度搜索，未知量用其估计代替，极小化 $J_{13}(\boldsymbol{\vartheta})$，$J_{14}(\boldsymbol{f})$ 和 $J_{15}(\boldsymbol{\theta}_{\mathrm{n}})$，可以得到辨识参数向量 $\boldsymbol{\vartheta}$，$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的分解随机梯度算法(Filtering based Decomposition Stochastic Gradient algorithm，F-D-SG 算法)：

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_4(t)}{r_4(t)}e_4(t),\tag{251}$$

$$e_4(t)=\hat{y}_4(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{252}$$

$$\hat{y}_4(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1),\tag{253}$$

$$r_4(t)=r_4(t-1)+\|\hat{\boldsymbol{\varphi}}_4(t)\|^2,\tag{254}$$

$$\hat{\boldsymbol{f}}(t)=\hat{\boldsymbol{f}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{xf}(t)}{r_5(t)}e_5(t),\tag{255}$$

$$e_5(t)=\hat{y}_5(t)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1),\tag{256}$$

$$\hat{y}_5(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{257}$$

$$r_5(t)=r_5(t-1)+\|\hat{\boldsymbol{\varphi}}_{xf}(t)\|^2,\tag{258}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)}{r_6(t)}e_6(t),\tag{259}$$

$$e_6(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1),\tag{260}$$

$$r_6(t)=r_6(t-1)+\|\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\|^2,\tag{261}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{262}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{263}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}_f(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t),\tag{264}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\tag{265}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{266}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{267}$$

$$\hat{\boldsymbol{\varphi}}_4(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{268}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},\tag{269}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}},\tag{270}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}},\tag{271}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}},\tag{272}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}},\tag{273}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\\hat{\boldsymbol{d}}(t)\end{bmatrix}.\tag{274}$$

注 15 细心的读者可能会发现 F-D-SG 算法(251)—(274)中 $e_4(t)$ 与 $e_5(t)$ 是相等的，都为

$$e(t):=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1),$$

因此式(251)—(258)可以等价写为

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_4(t)}{r_4(t)}e(t),\tag{275}$$

$$e(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1),\tag{276}$$

$$r_4(t)=r_4(t-1)+\|\hat{\boldsymbol{\varphi}}_4(t)\|^2,\tag{277}$$

$$\hat{\boldsymbol{f}}(t)=\hat{\boldsymbol{f}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}_{xf}(t)}{r_5(t)}e(t),\tag{278}$$

$$r_5(t)=r_5(t-1)+\|\hat{\boldsymbol{\varphi}}_{xf}(t)\|^2.\tag{279}$$

4.3 基于滤波的分解多新息随机梯度辨识算法

根据多新息辨识理论，基于 F-D-SG 算法

(259)—(279),滤波输出 $\hat{y}_f(t)$ 和噪声项 $\hat{w}(t)$,信息向量 $\hat{\boldsymbol{\varphi}}_4(t)$,$\hat{\boldsymbol{\varphi}}_{xf}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\rm n}(t)$ 扩展为堆积滤波输出向量 $\hat{\boldsymbol{Y}}_f(p,t)$ 和堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$,堆积信息矩阵 $\hat{\boldsymbol{\Phi}}_4(p,t)$,$\hat{\boldsymbol{\Phi}}_{xf}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)$:

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

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

$$\hat{\boldsymbol{\Phi}}_4(p,t):=[\hat{\boldsymbol{\varphi}}_4(t),\hat{\boldsymbol{\varphi}}_4(t-1),\cdots,\hat{\boldsymbol{\varphi}}_4(t-p+1)]\in\mathbf{R}^{(n_a+m)\times p},$$

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

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

将标量新息

$$e(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\rm T}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t)\hat{\boldsymbol{f}}(t-1)\in\mathbf{R},$$

$$e_6(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\in\mathbf{R}$$

扩展为新息向量:

$$\boldsymbol{E}(p,t):=\begin{bmatrix}\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\rm T}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t)\hat{\boldsymbol{f}}(t-1)\\ \hat{y}_f(t-1)-\hat{\boldsymbol{\varphi}}_4^{\rm T}(t-1)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t-1)\hat{\boldsymbol{f}}(t-1)\\ \vdots\\ \hat{y}_f(t-p+1)-\hat{\boldsymbol{\varphi}}_4^{\rm T}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t-p+1)\hat{\boldsymbol{f}}(t-1)\end{bmatrix}=$$

$$\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}_4^{\rm T}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\Phi}}_{xf}^{\rm T}(p,t)\hat{\boldsymbol{f}}(t-1)\in\mathbf{R}^p,$$

$$\boldsymbol{E}_6(p,t):=\begin{bmatrix}\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\\ \hat{w}(t-1)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t-1)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\\ \vdots\\ \hat{w}(t-p+1)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t-p+1)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)\end{bmatrix}=$$

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

可以得到辨识参数向量 $\boldsymbol{\vartheta}$,$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\rm n}$ 的基于滤波的分解多新息随机梯度算法(Filtering based Decomposition Multi-Innovation Stochastic Gradient algorithm, F-D-MISG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_4(p,t)}{r_4(t)}\boldsymbol{E}_4(p,t),\tag{280}$$

$$\boldsymbol{E}(p,t)=\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}_4^{\rm T}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\Phi}}_{xf}^{\rm T}(p,t)\hat{\boldsymbol{f}}(t-1),\tag{281}$$

$$r_4(t)=r_4(t-1)+\|\hat{\boldsymbol{\varphi}}_4(t)\|^2,\tag{282}$$

$$\hat{\boldsymbol{f}}(t)=\hat{\boldsymbol{f}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{xf}(p,t)}{r_5(t)}\boldsymbol{E}(p,t),\tag{283}$$

$$r_5(t)=r_5(t-1)+\|\hat{\boldsymbol{\varphi}}_{xf}(t)\|^2,\tag{284}$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=\hat{\boldsymbol{\theta}}_{\rm n}(t-1)+\frac{\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)}{r_6(t)}\boldsymbol{E}_6(p,t),\tag{285}$$

$$\boldsymbol{E}_6(p,t)=\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\rm n}^{\rm T}(p,t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1),\tag{286}$$

$$r_6(t)=r_6(t-1)+\|\hat{\boldsymbol{\varphi}}_{\rm n}(t)\|^2,\tag{287}$$

$$\hat{\boldsymbol{Y}}_f(p,t)=[\hat{y}_f(t),\hat{y}_f(t-1),\cdots,\hat{y}_f(t-p+1)]^{\rm T},\tag{288}$$

$$\hat{\boldsymbol{\Phi}}_4(p,t)=[\hat{\boldsymbol{\varphi}}_4(t),\hat{\boldsymbol{\varphi}}_4(t-1),\cdots,\hat{\boldsymbol{\varphi}}_4(t-p+1)],\tag{289}$$

$$\hat{\boldsymbol{\Phi}}_{xf}(p,t)=[\hat{\boldsymbol{\varphi}}_{xf}(t),\hat{\boldsymbol{\varphi}}_{xf}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{xf}(t-p+1)],\tag{290}$$

$$\hat{\boldsymbol{W}}(p,t)=[\hat{w}(t),\hat{w}(t-1),\cdots,\hat{w}(t-p+1)]^{\rm T},\tag{291}$$

$$\hat{\boldsymbol{\Phi}}_{\rm n}(p,t)=[\hat{\boldsymbol{\varphi}}_{\rm n}(t),\hat{\boldsymbol{\varphi}}_{\rm n}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\rm n}(t-p+1)],\tag{292}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\rm T}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{293}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{294}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{yf}^{\rm T}(t)\hat{\boldsymbol{a}}(t)-\hat{x}_f(t)=y(t)-\boldsymbol{\varphi}_y^{\rm T}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t),\tag{295}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\rm T}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\rm T}(t)\hat{\boldsymbol{\theta}}(t),\tag{296}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{297}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t),\tag{298}$$

$$\hat{\boldsymbol{\varphi}}_4(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\rm T}(t),\hat{\boldsymbol{\phi}}_f^{\rm T}(t)]^{\rm T},\tag{299}$$

$$\hat{\boldsymbol{\varphi}}_{\rm n}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\rm T},\tag{300}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\rm T},\tag{301}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\rm T},\tag{302}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\rm T},\tag{303}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\rm T},\tag{304}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_{\rm n}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}.\tag{305}$$

4.4 基于滤波的分解递推最小二乘辨识算法

根据辨识模型式(249)—(250)和(248),定义三个准则函数:

$$J_{16}(\boldsymbol{\vartheta}):=\sum_{j=1}^{t}[y_4(j)-\boldsymbol{\varphi}_4^{\rm T}(j)\boldsymbol{\vartheta}]^2,$$

$$J_{17}(\boldsymbol{f}):=\sum_{j=1}^{t}[y_5(j)-\boldsymbol{\varphi}_{xf}^{\rm T}(j)\boldsymbol{f}]^2,$$

$$J_{18}(\boldsymbol{\theta}_{\rm n}):=\sum_{j=1}^{t}[w(j)-\boldsymbol{\varphi}_{\rm n}^{\rm T}(j)\boldsymbol{\theta}_{\rm n}]^2.$$

仿照递推最小二乘辨识算法和基于 F-D-SG 算法(251)—(274)的推导,涉及的未知变量用其估计代替,可以得到辨识参数向量 $\boldsymbol{\vartheta}$,$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\rm n}$ 的基于滤波的分解递推最小二乘算法(Filtering based Decomposition Recursive Least Squares algorithm, F-D-RLS 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_4(t)e(t), \tag{306}$$

$$e(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1), \tag{307}$$

$$\boldsymbol{L}_4(t)=\boldsymbol{P}_4(t-1)\hat{\boldsymbol{\varphi}}_4(t)[1+\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\boldsymbol{P}_4(t-1)\hat{\boldsymbol{\varphi}}_4(t)]^{-1}, \tag{308}$$

$$\boldsymbol{P}_4(t)=[\boldsymbol{I}_{n_a+m}-\boldsymbol{L}_4(t)\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)]\boldsymbol{P}_4(t-1), \tag{309}$$

$$\hat{\boldsymbol{f}}(t)=\hat{\boldsymbol{f}}(t-1)+\boldsymbol{L}_5(t)e(t), \tag{310}$$

$$\boldsymbol{L}_5(t)=\boldsymbol{P}_5(t-1)\hat{\boldsymbol{\varphi}}_{xf}(t)[1+\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\boldsymbol{P}_5(t-1)\hat{\boldsymbol{\varphi}}_{xf}(t)]^{-1}, \tag{311}$$

$$\boldsymbol{P}_5(t)=[\boldsymbol{I}_{n_f}-\boldsymbol{L}_5(t)\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)]\boldsymbol{P}_5(t-1), \tag{312}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{L}_{\mathrm{n}}(t)[\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)], \tag{313}$$

$$\boldsymbol{L}_6(t)=\boldsymbol{P}_6(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)[1+\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{P}_6(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)]^{-1}, \tag{314}$$

$$\boldsymbol{P}_6(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_6(t)\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]\boldsymbol{P}_6(t-1), \tag{315}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{316}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{317}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}_f(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{318}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{319}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{320}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{321}$$

$$\hat{\boldsymbol{\varphi}}_4(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{322}$$

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

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{324}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \tag{325}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}}, \tag{326}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}}, \tag{327}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}. \tag{328}$$

4.5 基于滤波的分解多新息最小二乘辨识算法

仿照 F-D-MISG 算法(280)—(305)的推导,基于 F-D-RLS 算法(306)—(328),滤波输出 $\hat{y}_f(t)$ 和噪声项 $\hat{w}(t)$,信息向量 $\hat{\boldsymbol{\varphi}}_4(t)$,$\hat{\boldsymbol{\varphi}}_{xf}(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 扩展为堆积滤波输出向量 $\hat{\boldsymbol{Y}}_f(p,t)$ 和堆积噪声向量 $\hat{\boldsymbol{W}}(p,t)$,堆积信息矩阵 $\hat{\boldsymbol{\Phi}}_4(p,t)$,$\hat{\boldsymbol{\Phi}}_{xf}(p,t)$ 和 $\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)$,将标量新息

$$e(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_4^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)\in\mathbf{R},$$

$$e_6(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}$$

扩展为新息向量:

$$\boldsymbol{E}(p,t):=\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}_4^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\Phi}}_{xf}^{\mathrm{T}}(p,t)\hat{\boldsymbol{f}}(t-1)\in\mathbf{R}^p,$$

$$\boldsymbol{E}_6(p,t):=\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)\in\mathbf{R}^p,$$

可以得到辨识参数向量 $\boldsymbol{\vartheta}$,$\boldsymbol{f}$ 和 $\boldsymbol{\theta}_{\mathrm{n}}$ 的基于滤波的分解多新息最小二乘算法(Filtering based Decomposition Multi-Innovation Least Squares algorithm,F-D-MILS 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{P}_4(t)\hat{\boldsymbol{\Phi}}_4(p,t)\boldsymbol{E}(p,t), \tag{329}$$

$$\boldsymbol{E}(p,t)=\hat{\boldsymbol{Y}}_f(p,t)-\hat{\boldsymbol{\Phi}}_4^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)-\hat{\boldsymbol{\Phi}}_{xf}^{\mathrm{T}}(p,t)\hat{\boldsymbol{f}}(t-1), \tag{330}$$

$$\boldsymbol{P}_4^{-1}(t)=\boldsymbol{P}_4^{-1}(t-1)+\hat{\boldsymbol{\Phi}}_4(p,t)\hat{\boldsymbol{\Phi}}_4^{\mathrm{T}}(p,t), \tag{331}$$

$$\hat{\boldsymbol{f}}(t)=\hat{\boldsymbol{f}}(t-1)+\boldsymbol{P}_5(t)\hat{\boldsymbol{\Phi}}_{xf}(p,t)\boldsymbol{E}(p,t), \tag{332}$$

$$\boldsymbol{P}_5^{-1}(t)=\boldsymbol{P}_5^{-1}(t-1)+\hat{\boldsymbol{\Phi}}_{xf}(p,t)\hat{\boldsymbol{\Phi}}_{xf}^{\mathrm{T}}(p,t), \tag{333}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)=\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)+\boldsymbol{P}_6(t)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)\boldsymbol{E}_6(p,t), \tag{334}$$

$$\boldsymbol{E}_6(p,t)=\hat{\boldsymbol{W}}(p,t)-\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1), \tag{335}$$

$$\boldsymbol{P}_6^{-1}(t)=\boldsymbol{P}_6^{-1}(t-1)+\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)\hat{\boldsymbol{\Phi}}_{\mathrm{n}}^{\mathrm{T}}(p,t), \tag{336}$$

$$\hat{\boldsymbol{Y}}_f(p,t)=[\hat{y}_f(t),\hat{y}_f(t-1),\cdots,\hat{y}_f(t-p+1)]^{\mathrm{T}}, \tag{337}$$

$$\hat{\boldsymbol{\Phi}}_4(p,t)=[\hat{\boldsymbol{\varphi}}_4(t),\hat{\boldsymbol{\varphi}}_4(t-1),\cdots,\hat{\boldsymbol{\varphi}}_4(t-p+1)], \tag{338}$$

$$\hat{\boldsymbol{\Phi}}_{xf}(p,t)=[\hat{\boldsymbol{\varphi}}_{xf}(t),\hat{\boldsymbol{\varphi}}_{xf}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{xf}(t-p+1)], \tag{339}$$

$$\hat{\boldsymbol{W}}(p,t)=[\hat{w}(t),\hat{w}(t-1),\cdots,\hat{w}(t-p+1)]^{\mathrm{T}}, \tag{340}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{n}}(p,t)=[\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t-p+1)], \tag{341}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t-1)-\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{342}$$

$$\hat{x}(t)=\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{343}$$

$$\hat{v}(t)=\hat{y}_f(t)-\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}_f(t)=y(t)-\boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)-\hat{x}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{344}$$

$$\hat{x}_f(t)=\hat{\boldsymbol{\varphi}}_{xf}^{\mathrm{T}}(t)\hat{\boldsymbol{f}}(t)+\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{345}$$

$$\hat{y}_f(t)=y(t)+[y(t-1),y(t-2),\cdots,y(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{346}$$

$$\hat{\boldsymbol{\phi}}_f(t)=\boldsymbol{\phi}(t)+[\boldsymbol{\phi}(t-1),\boldsymbol{\phi}(t-2),\cdots,\boldsymbol{\phi}(t-n_c)]\hat{\boldsymbol{c}}(t)+[-\hat{\boldsymbol{\phi}}_f(t-1),-\hat{\boldsymbol{\phi}}_f(t-2),\cdots,-\hat{\boldsymbol{\phi}}_f(t-n_d)]\hat{\boldsymbol{d}}(t), \tag{347}$$

$$\hat{\boldsymbol{\varphi}}_4(t)=[\hat{\boldsymbol{\varphi}}_{yf}^{\mathrm{T}}(t),\hat{\boldsymbol{\phi}}_f^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{348}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{349}$$

$$\boldsymbol{\varphi}_y(t)=[-y(t-1),-y(t-2),\cdots,-y(t-n_a)]^{\mathrm{T}}, \tag{350}$$

$$\hat{\boldsymbol{\varphi}}_x(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n_f)]^{\mathrm{T}}, \tag{351}$$

$$\hat{\boldsymbol{\varphi}}_{yf}(t)=[-\hat{y}_f(t-1),-\hat{y}_f(t-2),\cdots,-\hat{y}_f(t-n_a)]^{\mathrm{T}}, \tag{352}$$

$$\hat{\boldsymbol{\varphi}}_{xf}(t)=[-\hat{x}_f(t-1),-\hat{x}_f(t-2),\cdots,-\hat{x}_f(t-n_f)]^{\mathrm{T}}, \tag{353}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\theta}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\theta}}_{\rm n}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix}. \tag{354}$$

5 结语

本文基于观测数据滤波、基于辨识模型分解,研究了一类伪线性参数系统的辅助模型随机梯度辨识算法、辅助模型多新息随机梯度辨识方法、辅助模型递推最小二乘辨识算法、辅助模型多新息最小二乘辨识方法.本文的伪线性参数系统具有下列形式:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t),$$

全称为自回归伪线性参数输出误差自回归滑动平均系统,包含了下列伪线性参数输出误差自回归滑动平均系统(有时不致于混淆时,可去掉“伪”字):

$$y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{A(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t),$$

还包括下列系统作为特例:

1)(伪)线性参数输出误差系统:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{F(z)}\boldsymbol{\theta}+v(t).$$

2)(伪)线性参数输出误差自回归系统:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{F(z)}\boldsymbol{\theta}+\frac{1}{C(z)}v(t).$$

3)(伪)线性参数输出误差滑动平均系统:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{F(z)}\boldsymbol{\theta}+D(z)v(t).$$

4)(伪)线性参数输出误差自回归滑动平均系统:

$$A(z)y(t)=\frac{\boldsymbol{\phi}^{\rm T}(t)}{F(z)}\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t).$$

5)AR-Box-Jenkins 系统:

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

6)(伪)参数线性方程误差自回归滑动平均系统:

$$A(z)y(t)=\boldsymbol{\phi}^{\rm T}(t)\boldsymbol{\theta}+\frac{D(z)}{C(z)}v(t).$$

本文提出的辨识算法的收敛性和仿真模拟都是需要进一步研究的课题.

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,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

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] Ding F,Liu X P,Liu G.Identification methods for Hammerstein nonlinear systems[J].Digital Signal Processing,2011,21(2):215-238

[13] Ding F,Liu X G,Chu J.Gradient based and least squares based iterative algorithms for Hammerstein systems using the hierarchical identification principle[J].IET Control Theory and Applications,2013,7(2):176-184

[14] Liu M M,Xiao Y S,Ding R F.Iterative identification algorithm for Wiener nonlinear systems using the Newton method[J].Applied Mathematical Modelling,2013,37(9):6584-6591.

[15] Wang D Q,Ding F.Hierarchical least squares estimation algorithm for Hammerstein-Wiener systems[J].IEEE Signal Processing Letters,2012,19(12):825-828

[16] Ding F,Chen T.Identification of Hammerstein nonlinear ARMAX systems[J].Automatica,2005,41(9):1479-1489

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

[18] Ding F,Shi Y,Chen T.Auxiliary model based least-squares identification methods for Hammerstein output-error systems[J].Systems and Control Letters,2007,56(5):373-380

[19] 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

[20] Ding F,Ma J X,Xiao Y S.Newton iterative identification for a class of output nonlinear systems with moving average noises[J].Nonlinear Dynamics,2013,74(1/2):21-30

[21] Wang D Q,Ding F.Least squares based and gradient based iterative identification for Wiener nonlinear systems[J].Signal Processing,2011,91(5):1182-1189

[22] Wang D Q,Ding F.Extended stochastic gradient identification algorithms for Hammerstein-Wiener ARMAX systems[J].Computers and Mathematics with Applications,2008,56(12):3157-3164

[23] Hu P P,Ding F,Shen J.Auxiliary model based least squares parameter estimation algorithm for feedback nonlinear systems using the hierarchical identification principle[J].Journal of the Franklin Institute:Engineering and Applied Mathematics,2013,350(10),3248-3259

[24] Hu P P,Ding F.Multistage least squares based iterative estimation for feedback nonlinear systems with moving average noises using the hierarchical identification principle[J].Nonlinear Dynamics,2013,73(1/2),583-592

[25] 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

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

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

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

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

[30] 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

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

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

[33] Golub G H,Van Loan C F.Matrix Computations[M],3rd Ed.Baltimore,MD:Johns Hopkins University Press,1996

[34] 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

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

[36] Ding F.Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling[J].Applied Mathematical Modelling,2013,37(4):1694-1704

Multi-innovation identification methods for linear-parameter systems

DING Feng[1,2,3] GUO Lanjie[1]
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 Systems have two categories, one is linear and the other is nonlinear. The linear systems have uniform descriptions and the nonlinear systems have countless categories and have no uniform descriptions. The linear-parameter systems are a special class of nonlinear systems and are linear on the parameter space. For pseudo-linear-parameter systems, this paper studies and presents the auxiliary model based multi-innovation (MI) identification methods, the data filtering based auxiliary model MI identification methods, the model decomposition based auxiliary model MI identification methods, and the filtering based decomposition MI identification methods. Finally, the computational efficiency, the computational steps and the flowcharts of several typical identification algorithms are discussed.

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