文章编号:1674-7070(2015)03-0193-21

丁锋[1,2,3]　陈启佳[1]

输出非线性方程误差类系统递推最小二乘辨识方法

摘要

随着控制技术的发展,控制对象的规模越来越大,使得辨识算法的计算量也越来越大.对于结构复杂的非线性系统,特别是包含未知参数乘积的非线性系统,使得过参数化辨识方法的参数数目大幅度增加,辨识算法的计算量也急剧增加,因此探索计算量小的参数估计方法势在必行.针对输出非线性方程误差类系统,讨论了基于过参数化模型的递推最小二乘类辨识方法;为减小过参数化辨识算法的计算量和提高辨识精度,分别利用分解技术和数据滤波技术,研究和提出了基于模型分解的递推最小二乘辨识方法和基于数据滤波的递推最小二乘辨识方法.最后给出了几个典型辨识算法的计算量、计算步骤、流程图.

关键词

参数估计;递推辨识;最小二乘;模型分解;数据滤波;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;输入非线性系统;输出非线性系统

中图分类号 TP273
文献标志码 A

收稿日期 2015-06-06
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

系统辨识就是在输入和输出数据的基础上,从给定的模型类中,确定一个与系统外特性最接近的模型[1-2].如果只注重系统的外部特性而忽略内部特性,这时可采用输入输出表达描述;如果还注重系统的内部特性,这时可采用状态空间模型描述.数学模型是研究一切事物运动规律的基础,系统辨识是研究建立系统数学模型的理论与方法.系统辨识作为一门基础学科得到了迅速的发展,迄今为止,已经诞生了各种不同类型的辨识方法,笔者等最近的连载论文[3-10]研究了多元系统、类多变量系统的耦合多新息辨识方法,多变量方程误差类系统的部分耦合迭代辨识方法,类多变量方程误差类系统的递阶多新息辨识方法,状态空间系统的多新息辨识方法,以及基于数据滤波技术和分解技术的输入非线性方程误差系统和输入非线性方程误差自回归系统的多新息辨识方法.本文分别基于过参数化模型、数据滤波技术,研究输出非线性方程误差类系统的递推最小二乘辨识方法.

非线性特性广泛存在于工业过程中,采取线性化方法对非线性系统进行处理难以取得令人满意的结果,因此,必须针对非线性系统的特殊结构,研究特别的辨识方法.如文献[11-12]提出了 Hammerstein 非线性 ARMAX 系统的最小二乘迭代方法与梯度迭代方法、递推增广最小二乘方法与增广随机梯度方法及其收敛性,文献[13]提出了 Hammerstein 输出误差系统的辅助模型递推最小二乘方法,文献[14]研究和提出了 Hammerstein 非线性系统的投影辨识方法、随机梯度辨识方法、牛顿递推辨识方法、牛顿迭代方法等,文献[15-16]针对输出非线性滑动平均系统提出了最小二乘迭代辨识方法与梯度迭代辨识方法、牛顿迭代辨识方法与基于分解的牛顿迭代辨识方法(文献[14-15]入选“2011 年中国百篇最具影响国际学术论文”,文献[15]入选 2014 年欧洲信号处理协会 3 篇最佳论文奖之一),文献[17]提出了 Hammerstein-Wiener ARMAX 系统的增广随机梯度辨识方法,文献[18]研究了 Hammerstein 非线性 FIR-ARMA 系统的分解牛顿迭代方法,文献[19]使用递阶辨识原理提出了 Hammerstein 非线性系统梯度迭代和最小二乘迭代方法,并使用分解技术提出了 Hammerstein 系统递阶多新息随机梯度辨识方法,文献[20]使用关键项分离和数据滤波提出了输入非线性 OEMA 系统递推最小二乘方法,文献[21-23]针对 Hammerstein 非线性受控自回归系

统提出了基于过参数化模型的递阶最小二乘辨识方法、基于关键项分离的分解递推最小二乘参数辨识方法和三阶段多新息随机梯度辨识方法.

本文考虑输出非线性方程误差类系统,包括输出非线性方程误差滑动平均系统、输出非线性方程误差自回归系统、输出非线性方程误差自回归滑动平均系统,讨论了基于过参数化模型的递推增广最小二乘辨识方法、递推广义最小二乘辨识方法、递推广义增广最小二乘辨识方法.为减小过参数化辨识算法的计算量和提高辨识精度,利用分解技术,研究和提出了基于模型分解的递推增广最小二乘辨识方法、递推广义最小二乘辨识方法、递推广义增广最小二乘辨识方法,将滤波技术与分解技术相结合,研究和提出了输出非线性方程误差系统的滤波递推最小二乘辨识方法.

1 非线性系统的基本类型

非线性系统模型结构是多种多样的,没有统一的数学表达形式.研究最多的是具有简单块结构特性的非线性系统.块结构非线性系统包括:1) 输入非线性系统(非线性块位于线性块之前:N—L);2) 输出非线性系统(非线性块位于线性块之后:L—N);3) 输入输出非线性系统(两个非线性块夹着一个线性块:N—L—N);4) 中间非线性系统(两端线性动态子系统夹着一个静态非线性环节:L—N—L);5) 反馈非线性系统(静态非线性环节可在前向通道,也可在反馈通道上)等.这里用 L 代表线性(linear)之意,N 代表非线性(nonlinear)之意.块结构非线性系统中的非线性通常指静态非线性.

当输入非线性块是一个静态多项式非线性环节,线性块是一个动态子系统时,这样的非线性系统也称为 Hammerstein 系统.当输出非线性块是一个静态多项式非线性环节,线性块是一个动态子系统时,这样的非线性系统也称为 Wiener 系统.非线性块可以是具有已知基函数、未知参数的线性组合,也可以是硬非线性,如死区非线性、饱和非线性、继电器非线性等.

以上是从线性块与非线性块的构成关系定义输入非线性系统、输出非线性系统的,其实,也可以从输入变量和输出变量在模型中的线性和非线性表达关系来定义输入非线性系统和输出非线性系统.按照这一规则,输入非线性系统定义为系统输出 $y(t)$ 是系统输入 $u(t-i)$ 的非线性函数,是过去输出 $y(t-i)$ 的线性函数,如典型的 Hammerstein 非线性系统[11,14].对应的输出非线性系统,定义为其输出 $y(t)$ 是系统过去输出 $y(t-i)$ 的非线性函数,是输入 $u(t-i)$ 的线性函数,如文献[15-16,24]中的非线性系统.输入输出非线性系统定义为其输出 $y(t)$ 既是系统输入 $u(t-i)$ 的非线性函数,又是过去输出 $y(t-i)$ 的非线性函数,如文献[15-16]中的非线性系统.

输入非线性系统包括输入非线性方程误差类系统和输入非线性输出误差类系统.输出非线性系统包括输出非线性方程误差类系统和输出非线性输出误差类系统.

按照干扰噪声性质区分,输出非线性方程误差类系统(ON-EET 系统)包含:

1) 基本输出非线性方程误差系统(ON-EE 系统),又称输出非线性受控自回归系统(ON-CAR 系统),它包含了输出非线性有限脉冲响应系统(ON-FIR 系统)作为特例;

2) 输出非线性方程误差滑动平均系统(ON-EEMA 系统),又称输出非线性受控自回归滑动平均系统(ON-CARMA 系统);

3) 输出非线性方程误差自回归系统(ON-EEAR 系统),又称输出非线性受控自回归自回归系统(ON-CARAR 系统),也称输出非线性动态调节系统;

4) 输出非线性方程误差自回归滑动平均系统(ON-EEARMA 系统),又称输出非线性受控自回归自回归滑动平均系统(ON-CARARMA 系统).

输出非线性输出误差类系统(ON-OET 系统)包含:

1) 基本输出非线性输出误差系统(ON-OE 系统);

2) 输出非线性输出误差滑动平均系统(ON-OEMA 系统);

3) 输出非线性输出误差自回归系统(ON-OEAR 系统);

4) 输出非线性输出误差自回归滑动平均系统(ON-OEARMA 系统),本文将其称为输出非线性 Box-Jenkins 系统(ON-BJ 系统).

2 基于过参数化模型的递推最小二乘方法

先定义一些符号.$\hat{\boldsymbol{\theta}}(t)$ 表示在时刻 t 参数向量 $\boldsymbol{\theta}$ 的估计;$\mathbf{1}_n$ 表示元都为 1 的 n 维列向量;$\boldsymbol{I}$ 表示适当维数的单位矩阵;$\boldsymbol{I}_n$ 表示 n 阶单位矩阵;“$X:=A$”和

“$A=:X$”表示 A 定义为 X;矩阵 $\boldsymbol{X}$ 的范数定义为 $\|\boldsymbol{X}\|^2:=\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$;$\{u(t)\}$ 和 $\{y(t)\}$ 分别表示系统输入和输出序列,$\{v(t)\}$ 是零均值、方差为 σ^2 的随机白噪声序列,z^{-1} 为单位后移算子[$z^{-1}y(t)=y(t-1)$,$zy(t)=y(t+1)$],$A(z)$,$B(z)$,$C(z)$ 和 $D(z)$ 是单位后移算子 z^{-1} 的常系数时不变多项式,定义为

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

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

设非线性函数 $f(y(t))$ 是已知非线性基函数 $\boldsymbol{f}=(f_1,f_2,\cdots,f_m)$,系数为 $(\gamma_1,\gamma_2,\cdots,\gamma_m)$ 的线性组合:

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}.$$

$\boldsymbol{f}(y(t)):=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}$ 是基函数构成的行向量,$\boldsymbol{\gamma}:=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m$ 是输出非线性部分的参数向量.

根据移位算子的性质,有:

$$A(z)\bar{y}(t)=A(z)f(y(t))=(a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a})f(y(t))=a_1f(y(t-1))+a_2f(y(t-2))+\cdots+a_{n_a}f(y(t-n_a)),$$
$$B(z)u(t)=(b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b})u(t)=b_1u(t-1)+b_2u(t-2)+\cdots+b_{n_b}u(t-n_b),$$
$$D(z)v(t)=(1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d})v(t)=v(t)+d_1v(t-1)+d_2v(t-2)+\cdots+d_{n_d}v(t-n_d).$$

参数向量定义如下:

$$\boldsymbol{a}:=[a_1,a_2,\cdots,a_{n_a}]^{\mathrm{T}}\in\mathbf{R}^{n_a},$$
$$\boldsymbol{b}:=[b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_b},$$
$$\boldsymbol{c}:=[c_1,c_2,\cdots,c_{n_c}]^{\mathrm{T}}\in\mathbf{R}^{n_c},$$
$$\boldsymbol{d}:=[d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_d},$$
$$\boldsymbol{\theta}_{\mathrm{s}}:=[\gamma_1\boldsymbol{a}^{\mathrm{T}},\gamma_2\boldsymbol{a}^{\mathrm{T}},\cdots,\gamma_m\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}^{\mathrm{T}}]^{\mathrm{T}}=[(\boldsymbol{\gamma}\otimes\boldsymbol{a})^{\mathrm{T}},\boldsymbol{b}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{n_1},\quad n_1:=mn_a+n_b,$$

定义信息矩阵 $\boldsymbol{F}(t)$,信息向量 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$,$\boldsymbol{\varphi}_u(t)$ 和 $\boldsymbol{\varphi}_v(t)$ 如下:

$$\boldsymbol{F}(t):=\begin{bmatrix} f_1(y(t-1)) & f_2(y(t-1)) & \cdots f_m(y(t-1)) \\ f_1(y(t-2)) & f_2(y(t-2)) & \cdots f_m(y(t-2)) \\ \vdots & \vdots & \vdots \\ f_1(y(t-n_a)) & f_2(y(t-n_a)) & \cdots f_m(y(t-n_a)) \end{bmatrix}\in\mathbf{R}^{n_a\times m},$$
$$\boldsymbol{\varphi}_{\mathrm{s}}(t):=[\boldsymbol{\phi}_1^{\mathrm{T}}(t),\boldsymbol{\phi}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{\phi}_m^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{n_1},$$
$$\boldsymbol{\phi}_j(t):=[f_j(y(t-1)),f_j(y(t-2)),\cdots,f_j(y(t-n_a))]^{\mathrm{T}}\in\mathbf{R}^{n_a},\quad j=1,2,\cdots,m,$$
$$\boldsymbol{\varphi}_u(t):=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_b},$$
$$\boldsymbol{\varphi}_v(t):=[v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_d}.$$

本节针对 3 种有色噪声干扰的输出非线性方程误差系统:输出非线性方程误差滑动平均(ON-EEMA)系统、输出非线性方程误差自回归(ON-EEAR)系统和输出非线性方程误差自回归滑动平均(ON-EEARMA)系统,讨论基于过参数化辨识模型的递推增广最小二乘算法、递推广义最小二乘算法和递推广义增广最小二乘算法.

2.1 基于过参数化模型的递推增广最小二乘方法

考虑下列输出非线性受控自回归滑动平均系统(Output Nonlinear Controlled AutoRegressive Moving Average system,ON-CARMA 系统),即输出非线性方程误差滑动平均系统(Output Nonlinear Equation-Error Moving Average system,ON-EEMA 系统):

$$y(t)=A(z)f(y(t))+B(z)u(t)+D(z)v(t),\tag{1}$$
$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}.\tag{2}$$

系统(1)是输入 $u(t-i)$ 的线性函数,是输出 $y(t-i)$ 的非线性函数,而且是一个方程误差系统,故是一个输出非线性方程误差系统,又噪声 $w(t):=D(z)v(t)$ 是一个滑动平均过程,故系统(1)是一个输出非线性方程误差滑动平均系统.

将多项式 $A(z)$,$B(z)$ 和 $D(z)$ 代入式(1),利用式(2)可得

$$y(t)=A(z)f(y(t))+B(z)u(t)+D(z)v(t)=\sum_{i=1}^{n_a}a_iz^{-i}f(y(t))+\sum_{i=1}^{n_b}b_iz^{-i}u(t)+\sum_{i=1}^{n_d}d_iz^{-i}v(t)+v(t)=\sum_{i=1}^{n_a}a_if(y(t-i))+\sum_{i=1}^{n_b}b_iu(t-i)+\sum_{i=1}^{n_d}d_iv(t-i)+v(t)\tag{3}$$
$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t).\tag{4}$$

系统(4)包含 4 个参数向量,$\boldsymbol{a}$,$\boldsymbol{b}$,$\boldsymbol{d}$ 是线性部分的参数向量,$\boldsymbol{\gamma}$ 是非线性部分的参数向量.辨识的目标是对于给定的输入输出数据 $\{u(t),y(t):t=1,2,\cdots\}$ 和非线性基函数 $f_j(*)$,研究辨识方法来估计这些参数向量,如果基函数是未知的,还要估计基函数 $f_j(*)$.

输出非线性系统(4)是一个典型双线性参数模型,它包含了参数向量 $\boldsymbol{a}$ 与 $\boldsymbol{\gamma}$ 之积,见乘积项 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}$,这也是辨识的困难.注意到式(4)可以等价

写为

$$y(t)=[\beta\boldsymbol{a}^{\mathrm{T}}]\boldsymbol{F}(t)\frac{\boldsymbol{\gamma}}{\beta}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t),$$

对于任意不为零的常数β,它与系统(4)有相同的输入输出关系,因此参数$\boldsymbol{a}$和$\boldsymbol{\gamma}$是不可辨识的.为了保证参数的可辨识性,必须对$\boldsymbol{a}$或$\boldsymbol{\gamma}$进行归一化.不同的辨识方法可能需要不同的归一化假设,典型的归一化假设有:

1) $a_1=1$或$\gamma_1=1$,其实可以规范化任一非零系数a_i或γ_i为1;

2) $A(1)=1$,即$a_1+a_2+\cdots+a_{n_a}=1$;

3) $a_1^2+a_2^2+\cdots+a_{n_a}^2=1, a_1>0$,或$\gamma_1^2+\gamma_2^2+\cdots+\gamma_m^2=1, \gamma_1>0$.

对于研究基于过参数化模型的辨识方法,这里假设参数向量$\boldsymbol{\gamma}$的第一个系数设置为1,即$\gamma_1=1$.读者可以采用假设$a_1=1$,推导相应的辨识方法.

假设$\gamma_1=1$,式(3)可以写为

$$\begin{aligned}y(t)=&\sum_{i=1}^{n_a}a_i\sum_{j=1}^{m}\gamma_jf_j(y(t-i))+\\&\sum_{i=1}^{n_b}b_iu(t-i)+\sum_{i=1}^{n_d}d_iv(t-i)+v(t)=\\&\sum_{i=1}^{n_a}[\gamma_1a_if_1(y(t-i))+\gamma_2a_if_2(y(t-i))+\\&\cdots+\gamma_ma_if_m(y(t-i))]+\\&\sum_{i=1}^{n_b}b_iu(t-i)+\sum_{i=1}^{n_d}d_iv(t-i)+v(t)=\\&\sum_{i=1}^{n_a}a_if_1(y(t-i))+\sum_{i=1}^{n_a}\gamma_2a_if_2(y(t-i))+\\&\cdots+\sum_{i=1}^{n_a}\gamma_ma_if_m(y(t-i))+\sum_{i=1}^{n_b}b_iu(t-i)+\\&\sum_{i=1}^{n_d}d_iv(t-i)+v(t).\end{aligned}\tag{5}$$

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

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\\boldsymbol{d}\end{bmatrix}\in\mathbf{R}^{n_0},$$

$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\\boldsymbol{\varphi}_v(t)\end{bmatrix}\in\mathbf{R}^{n_0},\quad n_0:=mn_a+n_b+n_d.$$

因此,系统(5)可以表示为伪线性回归形式:

$$\begin{aligned}y(t)=&\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\phi}_1(t)+\gamma_2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\phi}_2(t)+\cdots+\gamma_m\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\phi}_m(t)+\\&\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t)=\\&\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t).\end{aligned}\tag{6}$$

令$\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\\hat{\boldsymbol{d}}(t)\end{bmatrix}\in\mathbf{R}^{n_0}$是参数向量$\boldsymbol{\theta}=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\\boldsymbol{d}\end{bmatrix}\in\mathbf{R}^{n_0}$在时刻$t$的估计.定义信息矩阵$\boldsymbol{H}_t$、输出向量$\boldsymbol{Y}_t$、协方差阵$\boldsymbol{P}(t)$如下:

$$\begin{aligned}&\boldsymbol{H}_t:=[\boldsymbol{\varphi}(1),\boldsymbol{\varphi}(2),\cdots,\boldsymbol{\varphi}(t)]^{\mathrm{T}}\in\mathbf{R}^{t\times n_0},\\&\boldsymbol{Y}_t:=[y(1),y(2),\cdots,y(t)]^{\mathrm{T}}\in\mathbf{R}^t,\\&\boldsymbol{P}^{-1}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),\\&\boldsymbol{P}(0)=p_0\boldsymbol{I}_{n_0}.\end{aligned}\tag{7}$$

定义二次准则函数(quadratic criterion function):

$$J_1(\boldsymbol{\theta}):=J_1(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})=\sum_{j=1}^{t}[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2=\|\boldsymbol{Y}_t-\boldsymbol{H}_t\boldsymbol{\theta}\|^2.$$

令$J_1(\boldsymbol{\theta})$关于参数向量$\boldsymbol{\theta}$的偏导数为零,得到

$$\left.\frac{\partial J_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)}=-2\sum_{j=1}^{t}\boldsymbol{\varphi}(j)[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\hat{\boldsymbol{\theta}}(t)]=\mathbf{0},$$

或

$$-2\boldsymbol{H}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{\theta}}(t)]=\mathbf{0}.$$

求解得到

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)=&[\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t]^{-1}\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Y}_t=\boldsymbol{P}(t)\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Y}_t=\\&\boldsymbol{P}(t)[\boldsymbol{H}_{t-1}^{\mathrm{T}},\boldsymbol{\varphi}(t)]\begin{bmatrix}\boldsymbol{Y}_{t-1}\\y(t)\end{bmatrix}=\\&\boldsymbol{P}(t)\boldsymbol{H}_{t-1}^{\mathrm{T}}\boldsymbol{Y}_{t-1}+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)y(t)=\\&\boldsymbol{P}(t)\boldsymbol{P}^{-1}(t-1)\boldsymbol{P}(t-1)\boldsymbol{H}_{t-1}^{\mathrm{T}}\boldsymbol{Y}_{t-1}+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)y(t)=\\&\boldsymbol{P}(t)\boldsymbol{P}^{-1}(t-1)\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)y(t).\end{aligned}\tag{8}$$

为了得到$\hat{\boldsymbol{\theta}}(t)$的递推计算式,应用矩阵求逆引理

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

到式(7),可得

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}.\tag{9}$$

式(7)两边左乘$\boldsymbol{P}(t)$可得

$$\boldsymbol{I}_{n_0}=\boldsymbol{P}(t)\boldsymbol{P}^{-1}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),$$

或

$$\boldsymbol{P}(t)\boldsymbol{P}^{-1}(t-1)=\boldsymbol{I}_{n_0}-\boldsymbol{P}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t).$$

将上式代入式(8)得到递推计算关系:

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)=&[\boldsymbol{I}_{n_0}-\boldsymbol{P}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)y(t)=\\&\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)]=\\&\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t-1)-\\&\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t-1)].\end{aligned}\tag{10}$$

式(9)—(10)构成了计算参数估计向量$\hat{\boldsymbol{\theta}}(t)$的递推算法.然而,式(9)—(10)右边包含了未知噪声向量$\boldsymbol{\varphi}_v(t)$,使得递推计算无法进行,解决的办法是用噪声$v(t)$的估计$\hat{v}(t)$构成估计的噪声向量

$\hat{\boldsymbol{\varphi}}_v(t):=[\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_d}$. (11)

定义 $\boldsymbol{\varphi}(t)$ 的估计 $\hat{\boldsymbol{\varphi}}(t):=\begin{bmatrix}\boldsymbol{\varphi}_s(t)\\ \hat{\boldsymbol{\varphi}}_v(t)\end{bmatrix}\in\mathbf{R}^{n_0}$. 根据式(6)，可知 $v(t)$ 的估计可用下式计算：

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)=y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t). \tag{12}$$

用 $\hat{\boldsymbol{\varphi}}(t)$ 和 $\hat{\boldsymbol{\varphi}}_v(t)$ 分别代替式(10)和式(9)右边的 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{\varphi}_v(t)$，得到

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[y(t)-\boldsymbol{\varphi}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_s(t-1)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t-1)], \tag{13}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)}{1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}. \tag{14}$$

引入增益向量 $\boldsymbol{L}(t):=\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)\in\mathbf{R}^{n_0}$，使用式(14)可得

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)-\frac{\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}{1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)\left[1-\frac{\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}{1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}\right]=\frac{\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}{1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)}. \tag{15}$$

因此式(14)可以表示为

$$\boldsymbol{P}(t)=[\boldsymbol{I}_{n_0}-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1). \tag{16}$$

从以上各式，可以总结出 ON-EEMA 系统的基于过参数化辨识模型的递推增广最小二乘算法(Over-parameterization based Recursive Extended Least Squares algorithm，O-RELS 算法)：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{17}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \tag{18}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_{n_0}-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_{n_0}, \tag{19}$$

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

$$\boldsymbol{\varphi}_s(t)=[\boldsymbol{\phi}_1^{\mathrm{T}}(t),\boldsymbol{\phi}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{\phi}_m^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{21}$$

$$\boldsymbol{\phi}_j(t)=[f_j(y(t-1)),f_j(y(t-2)),\cdots,f_j(y(t-n_a))]^{\mathrm{T}}, \tag{22}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{23}$$

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

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

将估计得到的系统参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 展开为

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t),\widehat{\gamma_2\boldsymbol{a}}^{\mathrm{T}}(t),\cdots,\widehat{\gamma_m\boldsymbol{a}}^{\mathrm{T}}(t),\hat{\boldsymbol{b}}^{\mathrm{T}}(t),\hat{\boldsymbol{d}}^{\mathrm{T}}(t)]^{\mathrm{T}}=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\widehat{\gamma_2a_1}(t),\widehat{\gamma_2a_2}(t),\cdots,\widehat{\gamma_2a_{n_a}}(t),\cdots,\widehat{\gamma_ma_1}(t),\widehat{\gamma_ma_2}(t),\cdots,\widehat{\gamma_ma_{n_a}}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}.$$

从其构成可以看出，$\boldsymbol{a}$，$\boldsymbol{b}$ 和 $\boldsymbol{d}$ 的估计 $\hat{\boldsymbol{a}}(t)$，$\hat{\boldsymbol{b}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$ 可以分别从 $\hat{\boldsymbol{\theta}}(t)$ 的前 n_a 项，第 mn_a+1 到 n_1 项，最后 n_d 项直接读出，剩下的就是如何获得 γ_j 的估计 $\hat{\gamma}_j(t)$. $\hat{\boldsymbol{\theta}}(t)$ 包含了许多参数乘积的估计 $\widehat{\gamma_j\boldsymbol{a}}(t)$ $(j=1,2,\cdots,m)$. 用 $\hat{\theta}_i(t)$ 表示 $\hat{\boldsymbol{\theta}}(t)$ 的第 i 元，根据参数向量 $\boldsymbol{\theta}$ 的定义和 $\hat{\boldsymbol{\theta}}(t)$ 的构成，γ_j 的估计 $\hat{\gamma}_j(t)$ $(j=2,3,\cdots,m)$ 可以近似用 $\hat{\boldsymbol{\theta}}(t)$ 的第 $(j-1)n_a+i$ 项 $\hat{\theta}_{(j-1)n_a+i}(t)$ 除以第 i 项 $\hat{a}_i(t)$ 得到，即

$$\hat{\gamma}_j(t)=\frac{\hat{\theta}_{(j-1)n_a+i}(t)}{\hat{a}_i(t)},\quad j=2,3,\cdots,m.$$

对于 $i=1,2,\cdots,n_a$，参数 γ_j 有 n_a 个估计 $\hat{\gamma}_j(t)$，导致很多冗余估计，记作

$$\hat{\gamma}_{ji}(t):=\frac{\hat{\theta}_{(j-1)n_a+i}(t)}{\hat{a}_i(t)},\quad j=2,3,\cdots,m,\quad i=1,2,\cdots,n_a.$$

一种方法是取它们的平均值作为 γ_j 的估计[10]：

$$\hat{\gamma}_j(t)=\frac{1}{n_a}\sum_{i=1}^{n_a}\hat{\gamma}_{ji}(t)=\frac{1}{n_a}\sum_{i=1}^{n_a}\frac{\hat{\theta}_{(j-1)n_a+i}(t)}{\hat{a}_i(t)},\quad j=2,3,\cdots,m. \tag{26}$$

辨识算法的计算量是评价计算效率的一个重要指标.文献[25]指出，当衡量算法计算效率时，可以粗糙忽略存储量和与程序运行有关的其他间接消耗，直接用浮点运算(floating point operation，flop)数作为衡量算法效率的一种有效方法.一次乘法或加法次数称为一个 flop，除法作乘法对待，减法作加法对待.对于递推辨识算法，可以通过计算每一步的 flop 数来比较算法的计算效率.表 1 列出了 O-RELS 算法每一步的 flop 数.

2.2 基于过参数化模型的递推广义最小二乘方法

考虑下列输出非线性受控自回归自回归系统(Output Nonlinear Controlled AutoRegressive AutoRegressive system，ON-CARAR 系统)，即输出非线性方程误差自回归系统(Output Nonlinear Equation-Error AutoRegressive system，ON-EEAR 系统)：

$$y(t)=A(z)\bar{y}(t)+B(z)u(t)+\frac{1}{C(z)}v(t), \tag{27}$$

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \tag{28}$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \tag{29}$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m, \tag{30}$$

表 1　O-RELS 算法计算量

Table 1　The computational efficiency of the O-RELS algorithm

变量	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)e(t)$	$mn_a+n_b+n_d$	$mn_a+n_b+n_d$
$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)$	$mn_a+n_b+n_d$	$mn_a+n_b+n_d$
$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)$	$(mn_a+n_b+n_d)^2$	$(mn_a+n_b+n_d)^2-(mn_a+n_b+n_d)$
$\boldsymbol{L}(t)=\boldsymbol{\zeta}(t)/(1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t))$	$2(mn_a+n_b+n_d)$	$mn_a+n_b+n_d$
$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)$	$(mn_a+n_b+n_d)^2$	$(mn_a+n_b+n_d)^2$
总数	$2(mn_a+n_b+n_d)^2+4(mn_a+n_b+n_d)$	$2(mn_a+n_b+n_d)^2+2(mn_a+n_b+n_d)$
总 flop 数	$N_1:=4(mn_a+n_b+n_d)^2+6(mn_a+n_b+n_d)$	

有关变量定义同上.对于过参数化辨识方法,归一化 $\gamma_1=1$.

输出非线性方程误差自回归系统的干扰噪声是一个自回归过程:

$$w(t):=\frac{1}{C(z)}v(t)\in\mathbf{R}. \tag{31}$$

定义噪声信息向量

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

将式(31)和(27)表示为

$$w(t)=-c_1w(t-1)-c_2w(t-2)-\cdots-c_{n_c}w(t-n_c)+v(t)=\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+v(t), \tag{32}$$

$$y(t)=A(z)f(y(t))+B(z)u(t)+w(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+w(t)= \tag{33}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{34}$$

其中参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 定义为

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{c}\end{bmatrix}\in\mathbf{R}^{n_0},$$

$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \boldsymbol{\varphi}_w(t)\end{bmatrix}\in\mathbf{R}^{n_0},\quad n_0:=n_1+n_c.$$

令 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{c}}(t)\end{bmatrix}\in\mathbf{R}^{n_0}$ 是参数向量 $\boldsymbol{\theta}=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{c}\end{bmatrix}\in\mathbf{R}^{n_0}$ 在时刻 t 的估计.注意到信息向量 $\boldsymbol{\varphi}(t)$ 包含了未知噪声 $w(t-i)$,为了推导递推算法,用噪声 $w(t)$ 的估计 $\hat{w}(t)$ 构成估计的噪声向量

$$\hat{\boldsymbol{\varphi}}_w(t):=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}}\in\mathbf{R}^{n_c}. \tag{35}$$

根据式(33),可知 $w(t)$ 的估计可用下式计算:

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t). \tag{36}$$

对于辨识模型(34),定义准则函数

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

令 $J_2(\boldsymbol{\theta})$ 关于参数向量 $\boldsymbol{\theta}$ 的偏导数为零,得到

$$\frac{\partial J_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=-2\sum_{j=1}^{t}\boldsymbol{\varphi}(j)[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]=\mathbf{0}.$$

与 O-RELS 算法的推导思路相同,未知量 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{\varphi}_w(t)$ 用其对应的估计 $\hat{\boldsymbol{\varphi}}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_w(t)\end{bmatrix}\in\mathbf{R}^{n_0}$ 和 $\hat{\boldsymbol{\varphi}}_w(t)$ 代替,可以得到 ON-EEAR 系统的基于过参数化辨识模型的递推广义最小二乘算法(Over-parameterization based Recursive Generalized Least Squares algorithm,O-RGLS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{37}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \tag{38}$$

$$\boldsymbol{P}(t)=[\boldsymbol{I}_{n_0}-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1), \tag{39}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{40}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[\boldsymbol{\phi}_1^{\mathrm{T}}(t),\boldsymbol{\phi}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{\phi}_m^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{41}$$

$$\boldsymbol{\phi}_j(t)=[f_j(y(t-1)),f_j(y(t-2)),\cdots,f_j(y(t-n_a))]^{\mathrm{T}}, \tag{42}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{43}$$

$$\hat{\boldsymbol{\varphi}}_w(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}}, \tag{44}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{45}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t)]^{\mathrm{T}}, \tag{46}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t),\widehat{\gamma_2\boldsymbol{a}}^{\mathrm{T}}(t),\cdots,\widehat{\gamma_m\boldsymbol{a}}^{\mathrm{T}}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}}. \tag{47}$$

O-RGLS 算法的计算步骤如下:

1) 算法初始化:令 $t=1$,给定数据长度 L_e 和非线性函数 $f_j(*)$,置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_0}/p_0$,$\boldsymbol{P}(0)=p_0\boldsymbol{I}_{n_0}$,$\hat{w}(t-i)=1/p_0$,$i=1,2,\cdots,n_c$,$p_0=10^6$,$n_0:=mn_a+n_b+n_c$.

2) 收集输入 $u(t)$ 和输出 $y(t)$,用式(42),(43)和(44)分别构造 $\boldsymbol{\phi}_j(t)$,$\boldsymbol{\varphi}_u(t)$ 和 $\hat{\boldsymbol{\varphi}}_w(t)$.

3) 用式(41)构造 $\boldsymbol{\varphi}_{\mathrm{s}}(t)$,用式(40)构造 $\hat{\boldsymbol{\varphi}}(t)$.

4) 用式(38)计算增益向量 $\boldsymbol{L}(t)$,用式(39)计算协方差阵 $\boldsymbol{P}(t)$.

5) 用式(37)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并根据式(46)从 $\hat{\boldsymbol{\theta}}(t)$ 中分离出 $\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)$ 和 $\boldsymbol{c}(t)$.

6) 用式(45)计算 $\hat{w}(t)$.

7) 如果 $t\leqslant L_e$,t 增 1,转步骤 2),继续递推计算;否则,中断循环过程,获得参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.

2.3 基于过参数化模型的递推广义增广最小二乘方法

考虑下列输出非线性受控自回归自回归滑动平均系统(Output Nonlinear Controlled AutoRegressive AutoRegressive Moving Average system, ON-CARARMA 系统),即输出非线性方程误差自回归滑动平均系统(Output Nonlinear Equation-Error AutoRegressive Moving Average system,ON-EEARMA 系统):

$$y(t)=A(z)\bar{y}(t)+B(z)u(t)+\frac{D(z)}{C(z)}v(t), \tag{48}$$

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \tag{49}$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \tag{50}$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m, \tag{51}$$

有关变量定义同上.对于过参数化辨识方法,归一化 $\gamma_1=1$.

输出非线性方程误差自回归滑动平均系统的干扰噪声是一个自回归滑动平均过程:

$$w(t):=\frac{D(z)}{C(z)}v(t)\in\mathbf{R}. \tag{52}$$

定义噪声信息向量

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

将式(52)和(48)表示为

$$w(t)=-c_1w(t-1)-c_2w(t-2)-\cdots-c_{n_c}w(t-n_c)+d_1v(t-1)+d_2v(t-2)+\cdots+d_{n_d}v(t-n_d)+v(t)=\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t), \tag{53}$$

$$y(t)=A(z)f(y(t))+B(z)u(t)+w(t)=\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+w(t)= \tag{54}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{55}$$

其中参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 定义为

$$\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^{n_0},$$

$$\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \boldsymbol{\varphi}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_0},\quad n_0:=n_1+n_c+n_d,$$

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

令 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_0}$ 是参数向量 $\boldsymbol{\theta}=\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{s}}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^{n_0}$ 在时刻 t 的估计.注意到信息向量 $\boldsymbol{\varphi}(t)$ 包含了未知噪声向量 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$,为了推导递推算法,分别用噪声 $w(t)$ 和 $v(t)$ 的估计 $\hat{w}(t)$ 和 $\hat{v}(t)$ 构成估计的噪声向量

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t):=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}. \tag{56}$$

定义 $\boldsymbol{\varphi}(t)$ 的估计 $\hat{\boldsymbol{\varphi}}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_0}$.根据式(54),可知 $w(t)$ 的估计可用下式计算:

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t). \tag{57}$$

根据式(55),可知 $v(t)$ 的估计可用下式计算:

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t). \tag{58}$$

根据辨识模型(55),构造和极小化准则函数

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

令 $J_3(\boldsymbol{\theta})$ 关于参数向量 $\boldsymbol{\theta}$ 的偏导数为零,得到

$$\frac{\partial J_3(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)}=-2\hat{\boldsymbol{\varphi}}(j)\sum_{j=1}^{t}[y(j)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\hat{\boldsymbol{\theta}}(t)]^2=\mathbf{0}.$$

与 O-RELS 算法和 O-RGLS 算法的推导类似,未知量 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 分别用其对应的估计 $\hat{\boldsymbol{\varphi}}(t):=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{s}}(t)\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{n_0}$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 代替,可以得到 ON-EEARMA 系统的基于过参数化辨识模型的递推广义增广最小二乘算法(Over-parameterization based Recursive Generalized Extended Least Squares algorithm, O-RGELS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{59}$$

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

$$\boldsymbol{P}(t)=[\boldsymbol{I}-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1), \tag{61}$$

$$\hat{\boldsymbol{\varphi}}(t)=[\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{62}$$

$$\boldsymbol{\varphi}_{\mathrm{s}}(t)=[\boldsymbol{\phi}_1^{\mathrm{T}}(t),\boldsymbol{\phi}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{\phi}_m^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{63}$$

$$\boldsymbol{\phi}_j(t)=[f_j(y(t-1)),f_j(y(t-2)),\cdots f_j(y(t-n_a))]^{\mathrm{T}}, \tag{64}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{65}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{66}$$

$$\hat{w}(t)=y(t)-\boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{s}}(t), \tag{67}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{68}$$

$$\hat{\boldsymbol{\theta}}(t)=[\hat{\boldsymbol{\theta}}_{\mathrm{s}}^{\mathrm{T}}(t),\hat{\boldsymbol{\theta}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{69}$$

$$\hat{\boldsymbol{\theta}}_{\rm s}(t)=[\hat{\boldsymbol{a}}^{\rm T}(t),\ \widehat{\gamma_2\boldsymbol{a}}^{\rm T}(t),\cdots,\ \widehat{\gamma_m\boldsymbol{a}}^{\rm T}(t),\hat{b}_1(t),\hat{b}_2(t),\cdots,\hat{b}_{n_b}(t)]^{\rm T}, \tag{70}$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\rm T}. \tag{71}$$

O-RGELS 算法的计算步骤如下:

1) 初始化:令 $t=1$,给定非线性函数 $f_j(*)$,置初值 $\boldsymbol{P}(0)=p_0\boldsymbol{I}_{n_0}$,$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_0}/p_0$,$\hat{w}(t-i)=1/p_0$,$\hat{v}(t-i)=1/p_0$,$i=1,2,\cdots,\max[n_c,n_d]$,$p_0=10^6$,$n_0:=mn_a+n_b+n_c+n_d$.

2) 收集系统输入输出数据 $u(t)$ 和 $y(t)$,分别用式(64)—(66)构造 $\boldsymbol{\phi}_j(t)$,$\boldsymbol{\varphi}_u(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\rm n}(t)$.

3) 用式(63)构造 $\boldsymbol{\varphi}_{\rm s}(t)$,用式(62)构造 $\hat{\boldsymbol{\varphi}}(t)$.

4) 用式(60)计算增益向量 $\boldsymbol{L}(t)$,用式(61)计算协方差阵 $\boldsymbol{P}(t)$.

5) 用式(59)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并根据式(69)从 $\hat{\boldsymbol{\theta}}(t)$ 中分离出 $\hat{\boldsymbol{\theta}}_{\rm s}(t)$ 和 $\hat{\boldsymbol{\theta}}_{\rm n}(t)$.

6) 分别用式(67)—(68)计算 $\hat{w}(t-i)$ 和 $\hat{v}(t-i)$.

7) t 增 1,转到步骤 2),继续递推计算.

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

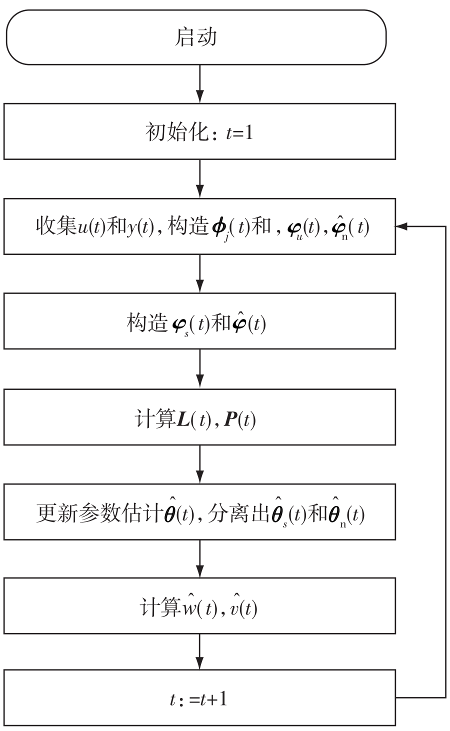

图 1 O-RGELS 算法计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的流程

Fig. 1 The flowchart of computing the O-RGELS parameter estimate $\hat{\boldsymbol{\theta}}(t)$

3 基于模型分解的递推最小二乘方法

过参数化辨识方法由于存在参数乘积,使得算法的计算量变大.以 O-RGELS 算法为例,系统的实际参数数目为 $(m+n_a+n_b+n_c+n_d)$,而协方差维数为 $(mn_a+n_b+n_c+n_d)\times(mn_a+n_b+n_c+n_d)$,这增加了算法的计算量.

对于高维数非线性系统,无论是辨识算法中的矩阵乘积运算还是求逆运算,尤其是对于涉及协方差阵 $\boldsymbol{P}(t)$ 的最小二乘算法,其计算量都是很大的.为减小计算量,将分解思想引入到输出非线性系统的辨识中,研究基于分解的辨识算法.

设非线性函数 $f(y(t))$ 是已知非线性基函数 $\boldsymbol{f}=(f_1,f_2,\cdots,f_m)$,系数为 $(\gamma_1,\gamma_2,\cdots,\gamma_m)$ 的线性组合:

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}. \tag{72}$$

$\boldsymbol{f}(y(t)):=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}$ 是基函数构成的行向量,$\boldsymbol{\gamma}:=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\rm T}\in\mathbf{R}^m$ 是输出非线性部分的参数向量.

定义非线性信息矩阵 $\boldsymbol{F}(t)$,系统信息向量 $\boldsymbol{\varphi}(t)$,$\boldsymbol{\varphi}_u(t)$ 和 $\boldsymbol{\varphi}_{\rm n}(t)$ 如下,

$$\boldsymbol{F}(t):=\begin{bmatrix} f_1(y(t-1)) & f_2(y(t-1)) & \cdots & f_m(y(t-1)) \\ f_1(y(t-2)) & f_2(y(t-2)) & \cdots & f_m(y(t-2)) \\ \vdots & \vdots & & \vdots \\ f_1(y(t-n_a)) & f_2(y(t-n_a)) & \cdots & f_m(y(t-n_a)) \end{bmatrix}\in\mathbf{R}^{n_a\times m},$$

$$\boldsymbol{\varphi}_u(t):=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\rm T}\in\mathbf{R}^{n_b},$$

$$\boldsymbol{\varphi}_{\rm n}(t):=\begin{bmatrix}\boldsymbol{\varphi}_w(t)\\ \boldsymbol{\varphi}_v(t)\end{bmatrix}\in\mathbf{R}^{n_c+n_d},$$

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

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

3.1 基于模型分解的递推增广最小二乘方法

考虑下列输出非线性方程误差滑动平均系统(ON-EEMA 系统):

$$y(t)=A(z)f(y(t))+B(z)u(t)+D(z)v(t), \tag{73}$$

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \tag{74}$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \tag{75}$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\rm T}\in\mathbf{R}^m, \tag{76}$$

其对应的双线性参数辨识模型为

$$y(t)=\sum_{i=1}^{n_a}a_i\boldsymbol{f}(y(t-i))\boldsymbol{\gamma}+\sum_{i=1}^{n_b}b_iu(t-i)+\sum_{i=1}^{n_d}d_iv(t-i)+v(t)=$$

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t). \tag{77}$$

双线性参数系统(77)包含4个参数向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{d}$ 和 $\boldsymbol{\gamma}$.前面借助于过参数化辨识模型，讨论了辨识过参数($\boldsymbol{\gamma}\otimes\boldsymbol{a},\boldsymbol{b},\boldsymbol{d}$)的递推增广最小二乘方法.由于过参数化导致许多冗余参数，要估计的参数数目增加，加大了算法的计算量.为了处理参数的乘积问题，这里利用递阶辨识原理，对系统进行形式上的分解，将4个参数向量($\boldsymbol{a},\boldsymbol{b},\boldsymbol{d},\boldsymbol{\gamma}$)分为2个参数集进行辨识，在推导一个参数集的辨识算法时，另一个集的参数作为常数对待，最后协调2个算法间的未知关联项，得到基于分解的辨识方法，或称为两阶段辨识方法.基于分解的两阶段辨识方法、三阶段辨识方法、多阶段辨识方法也称为递阶辨识方法.例如，对 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{d}$ 和 $\boldsymbol{\gamma}$ 进行不同的组合，2个参数集有($\boldsymbol{a},\boldsymbol{b}$)与($\boldsymbol{\gamma},\boldsymbol{d}$)，($\boldsymbol{a},\boldsymbol{d}$)与($\boldsymbol{\gamma},\boldsymbol{b}$)，($\boldsymbol{a},\boldsymbol{b},\boldsymbol{d}$)与 $\boldsymbol{\gamma}$，$\boldsymbol{a}$ 与($\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{d}$)等.当然也可以分为3个参数集或4个参数集进行辨识，得到三阶段辨识方法或四阶段辨识方法.递阶辨识方法可以大大减小计算量，提高计算效率.

在基于分解的辨识方法中，这里将参数分为2个集($\boldsymbol{a},\boldsymbol{b}$)和($\boldsymbol{\gamma},\boldsymbol{d}$)，令 $\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}$ 和 $\boldsymbol{\vartheta}:=\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{d}\end{bmatrix}$.归一化假设 $\|\boldsymbol{\gamma}\|^2=\gamma_1^2+\gamma_2^2+\cdots+\gamma_m^2=1,\gamma_1>0$.当然也可以假设 $\|\boldsymbol{a}\|=1,a_1>0$.

对于辨识模型(77)，定义2个准则函数：

$$J_4(\boldsymbol{\theta}):=\sum_{j=1}^{t}\{y(j)-\boldsymbol{\varphi}_v^{\mathrm{T}}(j)\boldsymbol{d}-[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}(j),\boldsymbol{\varphi}_u^{\mathrm{T}}(j)]\boldsymbol{\theta}\}^2,$$

$$J_5(\boldsymbol{\vartheta}):=\sum_{j=1}^{t}\{y(j)-\boldsymbol{\varphi}_u^{\mathrm{T}}(j)\boldsymbol{b}-[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(j),\boldsymbol{\varphi}_v^{\mathrm{T}}(j)]\boldsymbol{\vartheta}\}^2.$$

实际上2个准则函数是相等的，都等于 $J_1(*)$，即 $J_4(\boldsymbol{\theta})=J_5(\boldsymbol{\vartheta})=J_1(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})$.

定义输出向量 $\boldsymbol{Y}_t$，信息矩阵 $\boldsymbol{\Phi}_t,\boldsymbol{\Psi}_t,\boldsymbol{\Omega}_t$ 和 $\boldsymbol{\Xi}_t$ 如下：

$$\boldsymbol{Y}_t:=[y(1),y(2),\cdots,y(t)]^{\mathrm{T}}\in\mathbf{R}^t,$$

$$\boldsymbol{\Phi}_t:=[\boldsymbol{\varphi}_u(1),\boldsymbol{\varphi}_u(2),\cdots,\boldsymbol{\varphi}_u(t)]^{\mathrm{T}}\in\mathbf{R}^{t\times n_b},$$

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

$$\boldsymbol{\Omega}_t:=\begin{bmatrix}\boldsymbol{F}(1)\boldsymbol{\gamma} & \boldsymbol{F}(2)\boldsymbol{\gamma}\cdots\boldsymbol{F}(t)\boldsymbol{\gamma}\\ \boldsymbol{\varphi}_u(1) & \boldsymbol{\varphi}_u(2)\ \cdots\ \boldsymbol{\varphi}_u(t)\end{bmatrix}^{\mathrm{T}}\in\mathbf{R}^{t\times(n_a+n_b)},$$

$$\boldsymbol{\Xi}_t:=\begin{bmatrix}\boldsymbol{F}^{\mathrm{T}}(1)\boldsymbol{a} & \boldsymbol{F}^{\mathrm{T}}(2)\boldsymbol{a}\cdots\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{a}\\ \boldsymbol{\varphi}_v(1) & \boldsymbol{\varphi}_v(2)\ \cdots\ \boldsymbol{\varphi}_v(t)\end{bmatrix}^{\mathrm{T}}\in\mathbf{R}^{t\times(m+n_d)}.$$

因此，$J_4(\boldsymbol{\theta})$ 和 $J_5(\boldsymbol{\vartheta})$ 可以写为

$$J_4(\boldsymbol{\theta})=\|\boldsymbol{Y}_t-\boldsymbol{\Psi}_t\boldsymbol{d}-\boldsymbol{\Omega}_t\boldsymbol{\theta}\|^2,$$

$$J_5(\boldsymbol{\vartheta})=\|\boldsymbol{Y}_t-\boldsymbol{\Phi}_t\boldsymbol{b}-\boldsymbol{\Xi}_t\boldsymbol{\vartheta}\|^2.$$

最小二乘辨识原理就是利用系统的观测数据$\{u(t),y(t)\}$，极小化准则函数，得到参数 $\boldsymbol{\theta}$ 和 $\boldsymbol{\vartheta}$ 的估计.令准则函数 $J_4(\boldsymbol{\theta})$ 和 $J_5(\boldsymbol{\vartheta})$ 分别对参数向量 $\boldsymbol{\theta}$ 和 $\boldsymbol{\vartheta}$ 的偏导数为零：

$$\left.\frac{\partial J_4(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)}=-2\boldsymbol{\Omega}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{\Psi}_t\boldsymbol{d}-\boldsymbol{\Omega}_t\hat{\boldsymbol{\theta}}(t)]=\mathbf{0},$$

$$\left.\frac{\partial J_5(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}(t)}=-2\boldsymbol{\Xi}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{\Phi}_t\boldsymbol{b}-\boldsymbol{\Xi}_t\hat{\boldsymbol{\vartheta}}(t)]=\mathbf{0},$$

或

$$\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t\hat{\boldsymbol{\theta}}(t)=\boldsymbol{\Omega}_t^{\mathrm{T}}(\boldsymbol{Y}_t-\boldsymbol{\Psi}_t\boldsymbol{d}),$$

$$\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t\hat{\boldsymbol{\vartheta}}(t)=\boldsymbol{\Xi}_t^{\mathrm{T}}(\boldsymbol{Y}_t-\boldsymbol{\Phi}_t\boldsymbol{b}).$$

这2个方程叫做正则方程.在持续激励条件下，数据长度 t 足够大时，信息矩阵 $\boldsymbol{\Omega}_t$ 和 $\boldsymbol{\Xi}_t$ 是满秩的，矩阵 $\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t$ 和 $\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t$ 的逆存在，上两式两边分别左乘 $(\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t)^{-1}$ 和 $(\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t)^{-1}$，可以求得最小二乘估计：

$$\hat{\boldsymbol{\theta}}(t)=[\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t]^{-1}\boldsymbol{\Omega}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{\Psi}_t\boldsymbol{d}], \tag{78}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t]^{-1}\boldsymbol{\Xi}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{\Phi}_t\boldsymbol{b}]. \tag{79}$$

注意到式(78)和(79)需要计算矩阵$[\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t]$和$[\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t]$的逆，尤其当参数向量 $\hat{\boldsymbol{\theta}}(t)$ 和 $\hat{\boldsymbol{\vartheta}}(t)$ 的维数较大时，求逆运算的计算量会大大增加.为了克服这一缺点，下面推导 $\hat{\boldsymbol{\theta}}(t)$ 和 $\hat{\boldsymbol{\vartheta}}(t)$ 的递推实现方式.

定义信息向量 $\boldsymbol{\varphi}_1(t):=[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{(n_a+n_b)}$，定义协方差阵

$$\boldsymbol{P}_1^{-1}(t):=\boldsymbol{\Omega}_t^{\mathrm{T}}\boldsymbol{\Omega}_t=\sum_{j=1}^{t}\boldsymbol{\varphi}_1(j)\boldsymbol{\varphi}_1^{\mathrm{T}}(j)=\sum_{j=1}^{t-1}\boldsymbol{\varphi}_1(j)\boldsymbol{\varphi}_1^{\mathrm{T}}(j)+\boldsymbol{\varphi}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)=\boldsymbol{P}_1^{-1}(t-1)+\boldsymbol{\varphi}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\in\mathbf{R}^{(n_a+n_b)\times(n_a+n_b)}. \tag{80}$$

因此，式(78)可写为

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)&=\boldsymbol{P}_1(t)\boldsymbol{\Omega}_t^{\mathrm{T}}[\boldsymbol{Y}_t-\boldsymbol{\Psi}_t\boldsymbol{d}]=\\ &\boldsymbol{P}_1(t)[\boldsymbol{\Omega}_{t-1}^{\mathrm{T}},\boldsymbol{\varphi}_1(t)]\begin{bmatrix}\boldsymbol{Y}_{t-1}-\boldsymbol{\Psi}_{t-1}\boldsymbol{d}\\ y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}\end{bmatrix}=\\ &\boldsymbol{P}_1(t)\boldsymbol{\Omega}_{t-1}^{\mathrm{T}}[\boldsymbol{Y}_{t-1}-\boldsymbol{\Psi}_{t-1}\boldsymbol{d}]+\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)[y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}]=\\ &\boldsymbol{P}_1(t)\boldsymbol{P}_1^{-1}(t-1)\boldsymbol{P}_1(t-1)\boldsymbol{\Omega}_{t-1}^{\mathrm{T}}[\boldsymbol{Y}_{t-1}-\boldsymbol{\Psi}_{t-1}\boldsymbol{d}]+\\ &\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)[y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}]=\\ &\boldsymbol{P}_1(t)\boldsymbol{P}_1^{-1}(t-1)\hat{\boldsymbol{\theta}}(t-1)+\\ &\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)[y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}].\end{aligned} \tag{81}$$

注意到式(80)中的协方差阵是以逆 $\boldsymbol{P}_1^{-1}(t)$ 的形式定义的，为了得到 $\boldsymbol{P}_1(t)$ 的表达式，应用矩阵求逆引理

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

于式(80)得到

$$\boldsymbol{P}_1(t)=\boldsymbol{P}_1(t-1)-\frac{\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)}{1+\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}. \tag{82}$$

式(80)两边左乘 $\boldsymbol{P}_1(t)$ 可得

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

经移项可得

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

将上式代入式(81)得到参数 $\boldsymbol{\theta}$ 的递推计算式

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)=&[\boldsymbol{I}-\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)]\hat{\boldsymbol{\theta}}(t-1)+\\&\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)[y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}]=\\&\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)[y(t)-\\&\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}-\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)].\end{aligned}\tag{83}$$

引入增益向量 $\boldsymbol{L}_1(t):=\boldsymbol{P}_1(t)\boldsymbol{\varphi}_1(t)\in\mathbf{R}^{n_a+n_b}$,并将式(82)代入得

$$\begin{aligned}\boldsymbol{L}_1(t)=&\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)-\frac{\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}{1+\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}=\\&\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)\left[1-\frac{\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}{1+\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}\right]=\\&\frac{\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}{1+\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\boldsymbol{\varphi}_1(t)}.\end{aligned}\tag{84}$$

利用式(84)的增益向量 $\boldsymbol{L}_1(t)$,式(83)和(82)分别可以写为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}-\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\tag{85}$$

$$\begin{aligned}\boldsymbol{P}_1(t)=&\boldsymbol{P}_1(t-1)-\boldsymbol{L}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)=\\&[\boldsymbol{I}-\boldsymbol{L}_1(t)\boldsymbol{\varphi}_1^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1).\end{aligned}\tag{86}$$

由于式(84)—(86)右边包含了未知参数向量 $\boldsymbol{d}$、信息向量 $\boldsymbol{\varphi}_1(t)$ 和 $\boldsymbol{\varphi}_v(t)$,故无法计算出参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.解决方法是根据递阶辨识原理,未知 $\boldsymbol{d}$ 用其在时刻 $t-1$ 的估计 $\hat{\boldsymbol{d}}(t-1)$ 代替,未知 $\boldsymbol{\varphi}_1(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}_1(t)$ 代替,未知 $\boldsymbol{\varphi}_v(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}_v(t)$ 代替,得到

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)=&\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\\&\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t-1)-\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\end{aligned}\tag{87}$$

$$\boldsymbol{L}_1(t)=\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)[1+\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)]^{-1},\tag{88}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}-\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1),\quad \boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b},\tag{89}$$

其中

$$\hat{\boldsymbol{\varphi}}_1(t):=[\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}},$$

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

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t),$$

$\hat{\boldsymbol{\gamma}}(t)$ 为 $\boldsymbol{\gamma}$ 在时刻 t 的估计,$\hat{v}(t)$ 是 $v(t)$ 的估计.

定义信息向量 $\boldsymbol{\varphi}_2(t):=[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t),\boldsymbol{\varphi}_v^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbf{R}^{(m+n_d)}$,引入协方差阵和增益向量如下:

$$\boldsymbol{P}_2^{-1}(t):=\boldsymbol{\Xi}_t^{\mathrm{T}}\boldsymbol{\Xi}_t\in\mathbf{R}^{(m+n_d)\times(m+n_d)},$$

$$\boldsymbol{L}_2(t):=\boldsymbol{P}_2(t)\boldsymbol{\varphi}_2(t)\in\mathbf{R}^{m+n_d}.$$

进行类似的推导,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的递推关系式.由此可以总结得到 ON-EEMA 系统的基于模型分解递推增广最小二乘算法(model Decomposition based Recursive Extended Least Squares algorithm, D-RELS 算法):

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)=&\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t-1)-\\&\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\end{aligned}\tag{90}$$

$$\boldsymbol{L}_1(t)=\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)[1+\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)]^{-1},\tag{91}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1),\tag{92}$$

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(t)=&\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_2(t)[y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\\&\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\end{aligned}\tag{93}$$

$$\boldsymbol{L}_2(t)=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)[1+\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)]^{-1},\tag{94}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{m+n_d}-\boldsymbol{L}_2(t)\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)]\boldsymbol{P}_2(t-1),\tag{95}$$

$$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{96}$$

$$\hat{\boldsymbol{\varphi}}_2(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t),\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{97}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\tag{98}$$

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

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t),\tag{100}$$

$$\boldsymbol{F}(t)=\begin{bmatrix}f_1(y(t-1)) & f_2(y(t-1)) & \cdots f_m(y(t-1))\\ f_1(y(t-2)) & f_2(y(t-2)) & \cdots f_m(y(t-2))\\ \vdots & \vdots & \vdots\\ f_1(y(t-n_a)) & f_2(y(t-n_a)) & \cdots f_m(y(t-n_a))\end{bmatrix},\tag{101}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\gamma}}(t)\\ \hat{\boldsymbol{d}}(t)\end{bmatrix},\tag{102}$$

$$\hat{\boldsymbol{\Theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\vartheta}}(t)\end{bmatrix}.\tag{103}$$

D-RELS 算法的计算步骤如下:

1) 初始化:令 $t=1$,给定非线性函数 $f_j(*)$,置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{d}}(0)=1_{n_d}/p_0$,$\|\hat{\boldsymbol{\gamma}}(0)\|=1$,$\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{m+n_d}$,$p_0=10^6$.

2) 采集输入输出 $u(t)$ 和 $y(t)$,通过用式(101)构造 $\boldsymbol{F}(t)$,通过用式(98)构造 $\boldsymbol{\varphi}_u(t)$,用式(99)构造 $\hat{\boldsymbol{\varphi}}_v(t)$,用式(96)构造 $\hat{\boldsymbol{\varphi}}_1(t)$.

3) 用式(91)计算增益向量 $\boldsymbol{L}_1(t)$,用式(92)计算协方差阵 $\boldsymbol{P}_1(t)$.

4) 用式(90)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并从 $\hat{\boldsymbol{\theta}}(t)$ 中分离出 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{b}}(t)$.

5) 用式(97)构造 $\hat{\boldsymbol{\varphi}}_2(t)$,用式(94)和(95)分别计算 $\boldsymbol{L}_2(t)$ 和 $\boldsymbol{P}_2(t)$.

6) 通过式(93)刷新参数估计 $\hat{\boldsymbol{\vartheta}}(t)$,并从 $\hat{\boldsymbol{\vartheta}}(t)$

中分离出 $\hat{\boldsymbol{\gamma}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$.同时将 $\hat{\boldsymbol{\gamma}}(t)$ 进行归一化为单位向量,且第 1 元为正,即

$$\hat{\boldsymbol{\gamma}}(t)=\operatorname{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|},$$

其中 $\operatorname{sgn}[\hat{\gamma}_1(t)]$ 表示参数向量 $\hat{\boldsymbol{\gamma}}(t)$ 的第 1 个元素的符号.

7) 用式(100)计算 $\hat{v}(t)$.

8) t 增 1,转步骤 2),继续递推计算.

D-RELS 算法计算参数估计向量 $\hat{\boldsymbol{\Theta}}(t)$ 的流程如图 2 所示.

与 2.1 节研究的 O-RELS 算法相比,本节的 D-RELS 算法具有更小的计算量,因为 O-RELS 算法中的协方差阵 $\boldsymbol{P}(t)$ 的维数是 $(mn_a+n_b+n_d)\times(mn_a+n_b+n_d)$,而 D-RELS 算法将原系统分解为 2 个子系统,其协方差阵分别为 $\boldsymbol{P}_1(t)$ 和 $\boldsymbol{P}_2(t)$,各自的维数也减小到 $(n_a+n_b)\times(n_a+n_b)$ 和 $(m+n_d)\times(m+n_d)$. D-RELS 算法的 flop 数如表 2 所示. O-RELS 算法和 D-RELS 算法的计算量分别为

$$N_1:=4(mn_a+n_b+n_d)^2+6(mn_a+n_b+n_d),$$

$$N_2:=4(n_a+n_b)^2+4(m+n_d)^2+4mn_a+5n_a+10n_b+10n_d+7m.$$

为了比较哪个算法有更高的计算效率,计算一下这 2 个算法的计算量之差,当 $n_a\geqslant 2$ 和 $n_b\geqslant 2$ 时,$n_an_b>n_a+n_b$,$N_1>4(m+n_a+n_b+n_d)^2+6(m+n_a+n_b+n_d)$,有

$$N_1-N_2>4(n_a+n_b+n_d+m)^2+6(n_a+n_b+n_d+m)-4(n_a+n_b)^2-4(m+n_d)^2-4mn_a-5n\_a-10n_b-10n_d-7m=10mn_a+(8n_d-5)n\_a+(8n_d-4)n_b+(8m-4)n_d>0.$$

容易得出,D-RELS 算法比 O-RELS 算法具有更小的计算量.

3.2 基于模型分解的递推广义最小二乘方法

考虑下列输出非线性方程误差自回归(ON-EE-AR)系统:

$$y(t)=A(z)f(y(t))+B(z)u(t)+\frac{1}{C(z)}v(t), \qquad (104)$$

$$\bar{y}(t)=f(y(t))=\gamma_1f_1(y(t))+\gamma_2f_2(y(t))+\cdots+\gamma_mf_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \qquad (105)$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \qquad (106)$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m,\quad \|\boldsymbol{\gamma}\|=1,\quad \gamma_1>0. \qquad (107)$$

系统的干扰噪声是一个自回归过程:

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

或

$$w(t)=[1-C(z)]w(t)+v(t). \qquad (108)$$

因此,式(108)和(104)分别可以写为

$$w(t)=\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+v(t), \qquad (109)$$

$$y(t)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+w(t) \qquad (110)$$

$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+v(t). \qquad (111)$$

由式(110)可得

$$w(t)=y(t)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}.$$

表 2 D-RELS 算法计算量

Table 2 The computational efficiency of the D-RELS algorithm

变量	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)e_1(t)$	n_a+n_d	n_a+n_d
$e_1(t):=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)-\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)$	$n_a+n_b+n_d$	$n_a+n_b+n_d$
$\boldsymbol{\zeta}_1(t):=\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)$	$(n_a+n_d)^2$	$(n_a+n_d)^2-(n_a+n_d)$
$\boldsymbol{L}_1(t)=\boldsymbol{\zeta}_1(t)/[1+\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\boldsymbol{\zeta}_1(t)]$	$2(n_a+n_d)$	n_a+n_d
$\boldsymbol{P}_1(t)=\boldsymbol{P}_1(t-1)-\boldsymbol{L}_1(t)\boldsymbol{\zeta}_1^{\mathrm{T}}(t)$	$(n_a+n_d)^2$	$(n_a+n_d)^2$
$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{c}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{\mathrm{T}}$	n_am	n_am-n_a
$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_1(t)e_2(t)$	n_b+m	n_b+m
$e_2(t):=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t)-\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)$	n_b+m+n_d	n_b+m+n_d
$\boldsymbol{\zeta}_2(t):=\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)$	$(n_b+m)^2$	$(n_b+m)^2-(n_b+m)$
$\boldsymbol{L}_2(t)=\boldsymbol{\zeta}_2(t)/(1+\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\boldsymbol{\zeta}_2(t))$	$2(n_b+m)$	n_b+m
$\boldsymbol{P}_2(t)=\boldsymbol{P}_2(t-1)-\boldsymbol{L}_2(t)\boldsymbol{\zeta}_2^{\mathrm{T}}(t)$	$(n_b+m)^2$	$(n_b+m)^2$
$\boldsymbol{\zeta}_3(t):=\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)$	n_am	n_am-m
$e_3(t)=y(t)-\boldsymbol{\zeta}_3(t)\hat{\boldsymbol{c}}(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t)$	n_b+m+n_d	n_b+m+n_d
总数	$2(n_a+n_d)^2+2(n_b+m)^2+2n_am+4n_a+6n_b+5m+6n_d$	$2(n_a+n_d)^2+2(n_b+m)^2+2n_am+n_a+4n_b+2m+4n_d$
总 flop 数	$N_2:=4(n_a+n_d)^2+4(n_b+m)^2+4n_am+5n_a+10n_b+7m+10n_d$	

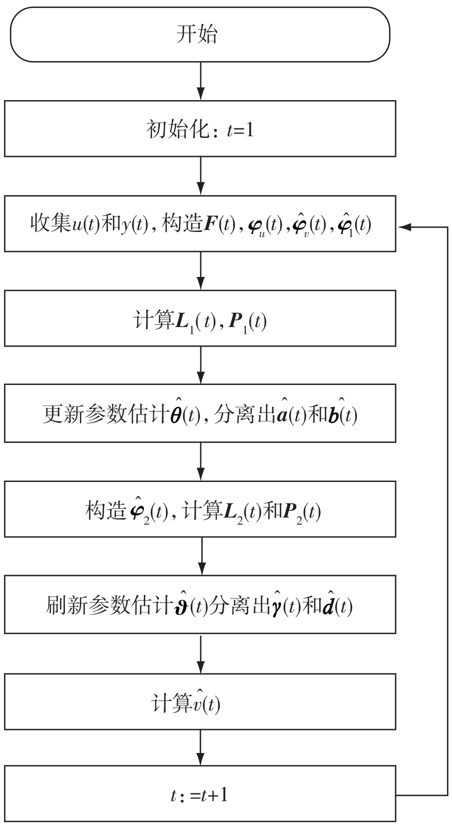

图 2　D-RELS 算法计算参数估计 $\hat{\boldsymbol{\Theta}}(t)$ 的流程

Fig. 2　The folowchart of computing the D-RELS parameter estimate $\hat{\boldsymbol{\Theta}}(t)$

设 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix}$ 和 $\hat{\boldsymbol{\vartheta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\gamma}}(t)\\ \hat{\boldsymbol{c}}(t)\end{bmatrix}$ 分别表示 $\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}$ 和 $\boldsymbol{\vartheta}:=\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{c}\end{bmatrix}$ 在时刻 t 的估计.$w(t)$ 的估计可由下式计算:

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t).$$

对于辨识模型(111),定义二次准则函数:

$$J_6(\boldsymbol{\theta}):=\|y(t)-\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}-[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]\boldsymbol{\theta}\|^2,$$

$$J_7(\boldsymbol{\vartheta}):=\|y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}-[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t),\boldsymbol{\varphi}_w^{\mathrm{T}}(t)]\boldsymbol{\vartheta}\|^2.$$

仿照 D-RELS 算法的推导,未知变量用其估计代替,极小化 $J_6(\boldsymbol{\theta})$ 和 $J_7(\boldsymbol{\vartheta})$,可以得到 ON-EEAR 系统的基于模型分解递推广义最小二乘算法(model Decomposition based Recursive Generalized Least Squares algorithm, D-RGLS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)-\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\tag{112}$$

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

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

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_2(t)[y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\tag{115}$$

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

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

$$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{118}$$

$$\hat{\boldsymbol{\varphi}}_2(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t),\hat{\boldsymbol{\varphi}}_w^{\mathrm{T}}(t)]^{\mathrm{T}},\tag{119}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},\tag{120}$$

$$\hat{\boldsymbol{\varphi}}_w(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c)]^{\mathrm{T}},\tag{121}$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t),\tag{122}$$

$$\boldsymbol{F}(t)=\begin{bmatrix}f_1(y(t-1)) & f_2(y(t-1)) & \cdots f_m(y(t-1))\\ f_1(y(t-2)) & f_2(y(t-2)) & \cdots f_m(y(t-2))\\ \vdots & \vdots & \vdots\\ f_1(y(t-n_a)) & f_2(y(t-n_a)) & \cdots f_m(y(t-n_a))\end{bmatrix},\tag{123}$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\gamma}}(t)\\ \hat{\boldsymbol{c}}(t)\end{bmatrix},\tag{124}$$

$$\hat{\boldsymbol{\Theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\vartheta}}(t)\end{bmatrix}.\tag{125}$$

D-RGLS 算法的计算步骤如下:

1) 初始化:令 $t=1$,给定数据长度和 L_e 和非线性函数 $f_j(*)$,置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\|\hat{\boldsymbol{\gamma}}(0)\|=1$,$\hat{\boldsymbol{c}}(0)=\mathbf{1}_{n_c}/p_0$,$\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{m+n_c}$,$p_0=10^6$.

2) 采集输入输出信息 $u(t)$ 和 $y(t)$,用式(123)构造 $\boldsymbol{F}(t)$,用式(120)构造 $\boldsymbol{\varphi}_u(t)$,用式(121)构造 $\hat{\boldsymbol{\varphi}}_w(t)$,用式(118)构造 $\hat{\boldsymbol{\varphi}}_1(t)$.

3) 用式(113)计算增益向量 $\boldsymbol{L}_1(t)$,用式(114)计算协方差阵 $\boldsymbol{P}_1(t)$.

4) 用式(112)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并从 $\hat{\boldsymbol{\theta}}(t)$ 中分离出 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{b}}(t)$.

5) 用式(119)构造 $\hat{\boldsymbol{\varphi}}_2(t)$,用式(116)和(117)分别计算 $\boldsymbol{L}_2(t)$ 和 $\boldsymbol{P}_2(t)$.

6) 用式(115)刷新参数估计 $\hat{\boldsymbol{\vartheta}}(t)$,并从 $\hat{\boldsymbol{\vartheta}}(t)$ 中分离出 $\hat{\boldsymbol{\gamma}}(t)$ 和 $\hat{\boldsymbol{c}}(t)$.同时将 $\hat{\boldsymbol{\gamma}}(t)$ 进行归一化:

$$\hat{\boldsymbol{\gamma}}(t)=\operatorname{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|},$$

其中 $\operatorname{sgn}[\hat{\gamma}_1(t)]$ 表示参数向量 $\hat{\boldsymbol{\gamma}}(t)$ 的第 1 个元素的符号.

7) 用式(122)计算 $\hat{w}(t)$.

(8) 如果 $t\leqslant L_e$,t 增 1,转步骤 2),继续递推计算;否则,中断循环过程,获得参数估计 $\hat{\boldsymbol{\Theta}}(t)$.

3.3 模型分解的递推广义增广最小二乘方法

考虑下列输出非线性方程误差自回归滑动平均(ON-EEARMA)系统:

$$y(t)=A(z)f(y(t))+B(z)u(t)+\frac{D(z)}{C(z)}v(t), \tag{126}$$

$$\bar{y}(t)=f(y(t))=\gamma_1 f_1(y(t))+\gamma_2 f_2(y(t))+\cdots+\gamma_m f_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \tag{127}$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \tag{128}$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m,\quad \|\boldsymbol{\gamma}\|=1,\quad \gamma_1>0. \tag{129}$$

系统的干扰噪声是一个自回归滑动平均过程:

$$w(t):=\frac{D(z)}{C(z)}v(t). \tag{130}$$

定义参数向量 $\boldsymbol{\theta}_{\mathrm{n}}:=\begin{bmatrix}\boldsymbol{c}\\ \boldsymbol{d}\end{bmatrix}\in\mathbf{R}^{n_c+n_d}$.式(130)可以写为

$$\begin{aligned}w(t)&=[1-C(z)]w(t)+D(z)v(t)=\\ &\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t)=\\ &\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\end{aligned} \tag{131}$$

利用式(130)和(131),由式(126)可得系统辨识模型:

$$y(t)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+w(t) \tag{132}$$

$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t). \tag{133}$$

令 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b}$ 和 $\hat{\boldsymbol{\vartheta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\gamma}}(t)\\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)\end{bmatrix}\in\mathbf{R}^{m+n_c+n_d}$ 分别是 $\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}\in\mathbf{R}^{n_a+n_b}$ 和 $\boldsymbol{\vartheta}:=\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\theta}_{\mathrm{n}}\end{bmatrix}\in\mathbf{R}^{m+n_c+n_d}$ 在时刻 t 的估计.定义信息向量

$$\boldsymbol{\varphi}_1(t):=[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}},$$

$$\boldsymbol{\varphi}_2(t):=[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t),\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}.$$

定义 2 个虚拟子系统输出 $y_1(t)$ 和 $y_2(t)$:

$$y_1(t):=y(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}},$$

$$y_2(t):=y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}.$$

利用以上两式,可将辨识模型(133)分解为如下 2 个子模型:

$$\mathrm{S}_1: y_1(t)=\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t),$$

$$\mathrm{S}_2: y_2(t)=\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\vartheta}+v(t).$$

2 个子系统维数小、变量较少,如图 3 所示.

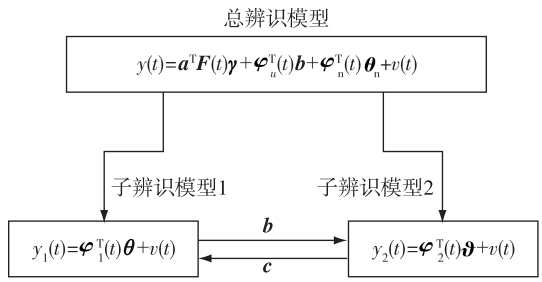

图 3 辨识模型分解为子辨识模型的递阶结构

Fig. 3 The hierarchical structure of the identification models

子模型 S_1 中的 $y_1(t)$ 包含 S_2 的未知参数向量 $\boldsymbol{\theta}_{\mathrm{n}}$,子模型 S_2 中的 $y_2(t)$ 包含模型 S_1 的未知参数向量 $\boldsymbol{b}$.把关联项 $\boldsymbol{\theta}_{\mathrm{n}}$ 和 $\boldsymbol{b}$ 看作是已知的,可以得到下列 2 个递推关系:

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t)&=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y_1(t)-\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)]=\\ &\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}-\boldsymbol{\varphi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(t)&=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_2(t)[y_2(t)-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)]=\\ &\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_2(t)[y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}-\boldsymbol{\varphi}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)].\end{aligned}$$

因为以上两式右边包含了未知信息向量 $\boldsymbol{\varphi}_1(t)$,$\boldsymbol{\varphi}_2(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 以及未知参数向量 $\boldsymbol{\theta}_{\mathrm{n}}$ 和 $\boldsymbol{b}$,解决方法是:未知信息向量 $\boldsymbol{\varphi}_1(t)$,$\boldsymbol{\varphi}_2(t)$ 和 $\boldsymbol{\varphi}_{\mathrm{n}}(t)$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_1(t)$,$\hat{\boldsymbol{\varphi}}_2(t)$ 和 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$ 代替,未知参数向量 $\boldsymbol{\theta}_{\mathrm{n}}$ 和 $\boldsymbol{b}$ 分别用其估计 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)$ 和 $\hat{\boldsymbol{b}}(t)$ 代替,能够得到 ON-EEARMA 系统的基于模型分解递推广义增广最小二乘算法(model Decomposition based Recursive Generalized Extended Least Squares algorithm, D-RGELS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[y(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t-1)-\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{134}$$

$$\boldsymbol{L}_1(t)=\frac{\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)}{1+\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)}, \tag{135}$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}_1^{\mathrm{T}}(t)]\boldsymbol{P}_1(t-1), \tag{136}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{L}_2(t)[y(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{137}$$

$$\boldsymbol{L}_2(t)=\frac{\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)}{1+\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)}, \tag{138}$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_{m+n_c+n_d}-\boldsymbol{L}_2(t)\hat{\boldsymbol{\varphi}}_2^{\mathrm{T}}(t)]\boldsymbol{P}_2(t-1), \tag{139}$$

$$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t),\boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{140}$$

$$\hat{\boldsymbol{\varphi}}_2(t)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t),\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{141}$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{142}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\mathrm{T}}, \tag{143}$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t), \tag{144}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t), \tag{145}$$

$$\boldsymbol{F}(t)=\begin{bmatrix} f_1(y(t-1)) & f_2(y(t-1)) & \cdots f_m(y(t-1)) \\ f_1(y(t-2)) & f_2(y(t-2)) & \cdots f_m(y(t-2)) \\ \vdots & \vdots & \vdots \\ f_1(y(t-n_a)) & f_2(y(t-n_a)) & \cdots f_m(y(t-n_a)) \end{bmatrix}, \quad (146)$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix} \hat{\boldsymbol{a}}(t) \\ \hat{\boldsymbol{b}}(t) \end{bmatrix}, \quad \hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix} \hat{\boldsymbol{\gamma}}(t) \\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t) \end{bmatrix}, \quad (147)$$

$$\hat{\boldsymbol{\Theta}}(t)=\begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \hat{\boldsymbol{\vartheta}}(t) \end{bmatrix}. \quad (148)$$

D-RGELS 算法的计算步骤如下:

1)初始化:令 $t=1$,给定非线性函数 $f_j(*)$,置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(0)=\mathbf{1}_{n_c+n_d}/p_0$,$\|\hat{\boldsymbol{\gamma}}(0)\|=1$,$\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{m+n_c+n_d}$,$p_0=10^6$.

2)利用收集的输入输出数据 $u(t)$ 和 $y(t)$,用式(146)构造 $\boldsymbol{F}(t)$,用式(142)构造 $\boldsymbol{\varphi}_u(t)$,用式(143)构造 $\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$,用式(140)构造 $\hat{\boldsymbol{\varphi}}_1(t)$.

3)用式(135)计算增益向量 $\boldsymbol{L}_1(t)$,用式(136)计算协方差阵 $\boldsymbol{P}_1(t)$.

4)用式(134)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并从 $\hat{\boldsymbol{\theta}}(t)$ 中分离出 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{b}}(t)$.

5)用式(141)构造 $\hat{\boldsymbol{\varphi}}_2(t)$,用式(138)和(139)分别计算 $\boldsymbol{L}_2(t)$ 和 $\boldsymbol{P}_2(t)$.

6)通过式(137)刷新参数估计 $\hat{\boldsymbol{\vartheta}}(t)$,并从 $\hat{\boldsymbol{\vartheta}}(t)$ 中分离出 $\hat{\boldsymbol{\gamma}}(t)$,$\hat{\boldsymbol{c}}(t)$ 和 $\hat{\boldsymbol{d}}(t)$.同时将 $\hat{\boldsymbol{\gamma}}(t)$ 进行归一化:

$$\hat{\boldsymbol{\gamma}}(t):=\operatorname{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|}, \quad \hat{\boldsymbol{\gamma}}(t):=\begin{bmatrix} \hat{\boldsymbol{\gamma}}(t) \\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t) \end{bmatrix}.$$

7)分别用式(144)和(145)计算 $\hat{w}(t)$ 和 $\hat{v}(t)$.

8) t 增 1,转步骤 2),继续递推计算.

D-RGELS 算法计算参数估计 $\hat{\boldsymbol{\Theta}}(t)$ 的流程如图 4 所示.

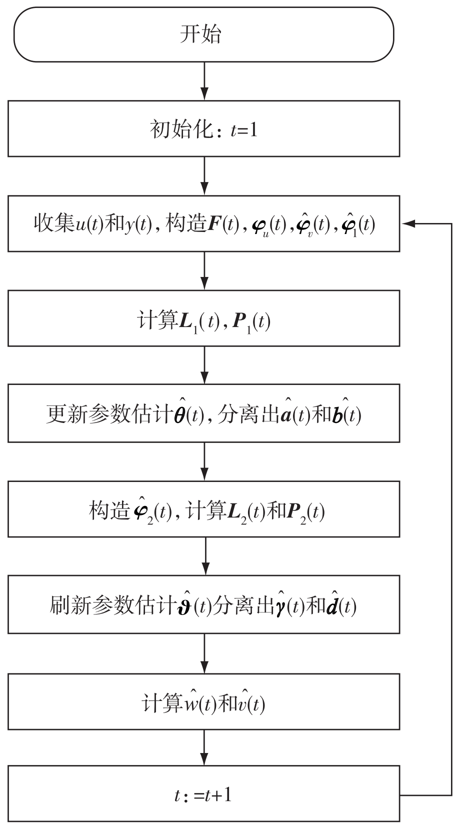

图 4　D-RGELS 算法估计 $\hat{\boldsymbol{\Theta}}(t)$ 的流程

Fig. 4　The flowchart of computing the D-RGELS parameter estimate $\hat{\boldsymbol{\Theta}}(t)$

4　基于数据滤波的递推最小二乘方法

上一节研究了基于分解的输出非线性方程误差类系统的递推最小二乘辨识算法,提高了算法的计算效率.本节从噪声模型入手,引入滤波思想,对系统进行滤波,滤波后的系统就会由有色噪声干扰变成白噪声干扰的系统,这样可以进一步减少算法的计算量,提高参数估计精度.

滤波是信号处理中的一个基本术语.滤波技术广泛应用于各种领域,包括故障检测、参数估计、地面目标追踪等.用一个线性低通滤波器(高通滤波器)对信号进行滤波,可以剔除信号中的高频干扰(低频干扰).在系统辨识中,滤波的作用是不同的,对系统的输入输出数据进行滤波,不会改变系统的输入输出关系,但可改变系统模型中的干扰噪声结构.如果滤波器选择得当,进行滤波后,系统辨识表达式会由有色噪声干扰变成由白噪声干扰,从而提高参数估计精度[26-28].

本文提出的滤波辨识算法的基本原理是,利用噪声模型的传递函数(滤波器)对系统进行滤波后,将系统表达式分解为 2 个辨识模型:滤波后系统模型和噪声模型.然后分别辨识 2 个模型的参数.由于系统的干扰噪声模型往往是未知的,辨识的技巧是用估计的噪声模型进行滤波,采用递推或迭代方案,交替估计系统模型参数和噪声模型参数,实现辨识算法.

4.1　滑动平均滤波递推最小二乘辨识方法

考虑下列输出非线性方程误差滑动平均系统(ON-EEMA 系统):

$$y(t)=A(z)f(y(t))+B(z)u(t)+D(z)v(t), \quad (149)$$

$$f(y(t))=\gamma_1 f_1(y(t))+\gamma_2 f_2(y(t))+\cdots+\gamma_m f_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma}, \tag{150}$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m}, \tag{151}$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m,\quad \|\boldsymbol{\gamma}\|=1,\quad \gamma_1>0. \tag{152}$$

采用前文的定义,其对应的双线性参数辨识模型为

$$y(t)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+D(z)v(t) \tag{153}$$

$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+w(t) \tag{154}$$

$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t). \tag{155}$$

定义滤波输入 $u_{\mathrm{f}}(t)$,滤波输出 $y_{\mathrm{f}}(t)$,滤波信息向量 $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ 和滤波信息矩阵 $\boldsymbol{G}(t)$ 如下:

$$u_{\mathrm{f}}(t):=\frac{1}{D(z)}u(t),\quad y_{\mathrm{f}}(t):=\frac{1}{D(z)}y(t),$$

$$\boldsymbol{\varphi}_{\mathrm{f}}(t):=[u_{\mathrm{f}}(t-1),u_{\mathrm{f}}(t-2),\cdots,u_{\mathrm{f}}(t-n_b)]^{\mathrm{T}}=\frac{1}{D(z)}\boldsymbol{\varphi}_u(t)\in\mathbf{R}^{n_b},$$

$$\boldsymbol{G}(t):=\begin{bmatrix} g_1(t-1) & g_2(t-1) & \cdots & g_m(t-1)\\ g_1(t-2) & g_2(t-2) & \cdots & g_m(t-2)\\ \vdots & \vdots & & \vdots\\ g_1(t-n_a) & g_2(t-n_a) & \cdots & g_m(t-n_a)\end{bmatrix}\in\mathbf{R}^{n_a\times m},$$

$$g_j(t):=\frac{1}{D(z)}f_j(y(t)),\quad j=1,2,\cdots,m.$$

式(149)两边同时乘$\frac{1}{D(z)}$得到

$$\frac{1}{D(z)}y(t)=A(z)\frac{1}{D(z)}f(y(t))+B(z)\frac{1}{D(z)}u(t)+v(t),$$

即

$$y_{\mathrm{f}}(t)=A(z)\frac{1}{D(z)}f(y(t))+B(z)u_{\mathrm{f}}(t)+v(t)=$$
$$A(z)\frac{1}{D(z)}(\gamma_1 f_1(y(t))+\gamma_2 f_2(y(t))+\cdots+\gamma_m f_m(y(t)))+B(z)u_{\mathrm{f}}(t)+v(t)=$$
$$A(z)(\gamma_1 g_1(t)+\gamma_2 g_2(t)+\cdots+\gamma_m g_m(t))+B(z)u_{\mathrm{f}}(t)+v(t)=$$
$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{b}+v(t). \tag{156}$$

由式(153)和(156)可以看出,式(153)中时刻 t 的输出 $y(t)$ 包含了有色噪声项 $D(z)v(t)$,而式(156)中时刻 t 的输出 $y_{\mathrm{f}}(t)$ 只包含了白噪声干扰 $v(t)$.也就是说,经过滤波,模型(154)化为了一个白噪声干扰的输出非线性辨识模型(滤波后系统模型),其不显含噪声模型 $D(z)$ 的参数.下面推导噪声参数辨识模型.定义中间变量 $w(t)$,可以得到噪声模型辨识表达式:

$$w(t):=D(z)v(t)=\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t). \tag{157}$$

辨识模型(156)和(157)包含了系统所有待辨识参数向量 $\boldsymbol{a},\boldsymbol{\gamma},\boldsymbol{b}$ 和 $\boldsymbol{d}$.在辨识模型(156)和(157)中,由于滤波器 $D(z)$ 是未知的,所以变量 $y_{\mathrm{f}}(t)$,$\boldsymbol{G}(t)$,$\boldsymbol{\varphi}_{\mathrm{f}}(t)$ 都是未知的,噪声 $w(t)$ 和噪声信息向量 $\boldsymbol{\varphi}_v(t)$ 也是未知的.因此在推导基于滤波的辨识算法中,解决的办法是借助于辅助模型辨识思想,未知变量用辅助模型的输出代替,或者用其对应的估计代替.由式(154)有

$$w(t)=y(t)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}. \tag{158}$$

令 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b}$,$\hat{\boldsymbol{\gamma}}(t)\in\mathbf{R}^m$ 和 $\hat{\boldsymbol{d}}(t)\in\mathbf{R}^{n_d}$ 分别是 $\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}$,$\boldsymbol{\gamma}$ 和 $\boldsymbol{d}$ 在时刻 t 的估计.将式(158)中的参数向量 $\boldsymbol{a},\boldsymbol{\gamma}$ 和 $\boldsymbol{b}$ 分别用它们在时刻$t-1$ 的估计 $\hat{\boldsymbol{a}}(t-1)$,$\hat{\boldsymbol{\gamma}}(t-1)$ 和 $\hat{\boldsymbol{b}}(t-1)$ 代替,则 $w(t)$ 的估计 $\hat{w}(t)$ 可由下式计算:

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t-1)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1).$$

读者可能注意到:在计算 $w(t)$ 的估计 $\hat{w}(t)$ 时,式(122),式(144)与上式的区别在于采用不同时刻的参数估计,这样做的目的是为了实现递推计算.

令 $\hat{v}(t)$ 是 $v(t)$ 的估计.用其构造 $\boldsymbol{\varphi}_v(t)$ 的估计:

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

由式(155)可得

$$v(t)=y(t)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}-\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}.$$

上式中未知量 $\boldsymbol{a},\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{\varphi}_v(t),\boldsymbol{d}$ 分别用它们在时刻 t 的估计 $\hat{\boldsymbol{a}}(t)$,$\hat{\boldsymbol{\gamma}}(t)$,$\hat{\boldsymbol{b}}(t)$,$\hat{\boldsymbol{\varphi}}_v(t)$,$\hat{\boldsymbol{d}}(t)$ 代替,那么 $v(t)$ 的估计 $\hat{v}(t)$ 可由下式计算:

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_v^{\mathrm{T}}(t)\hat{\boldsymbol{d}}(t).$$

用噪声模型的参数估计

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

构造 $D(z)$ 的估计

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

由于 $D(z)$ 是未知的,故用其估计 $\hat{D}(t,z)$ 对 $u(t)$ 和 $y(t)$ 进行滤波,得到 $u_{\mathrm{f}}(t)$ 和 $y_{\mathrm{f}}(t)$ 的估计如下,

$$\hat{u}_{\mathrm{f}}(t)=\frac{1}{\hat{D}(t,z)}u(t),\quad \hat{y}_{\mathrm{f}}(t)=\frac{1}{\hat{D}(t,z)}y(t),$$

或

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

它们可由以下两式递推计算:

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

$$\hat{y}_{\rm f}(t)=[1-\hat{D}(t,z)]\hat{y}_{\rm f}(t)+y(t)=$$
$$y(t)-\hat{d}_1(t)\hat{y}_{\rm f}(t-1)-\hat{d}_2(t)\hat{y}_{\rm f}(t-2)-\cdots-$$
$$\hat{d}_{n_d}(t)\hat{y}_{\rm f}(t-n_d).$$

定义信息向量:

$$\boldsymbol{\varphi}_1(t):=\begin{bmatrix}\boldsymbol{G}(t)\boldsymbol{\gamma}\\ \boldsymbol{\varphi}_{\rm f}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}_2(t):=\boldsymbol{G}^{\rm T}(t)\boldsymbol{a}\in\mathbf{R}^m.$$

根据滤波模型(156)和噪声模型(157),定义准则函数:

$$J_8(\boldsymbol{\theta}):=\sum_{j=1}^{t}[y_{\rm f}(j)-\boldsymbol{\varphi}_1^{\rm T}(j)\boldsymbol{\theta}]^2,$$

$$J_9(\boldsymbol{\gamma}):=\sum_{j=1}^{t}[y_{\rm f}(j)-\boldsymbol{\varphi}_{\rm f}^{\rm T}(j)\hat{\boldsymbol{b}}(t)-\boldsymbol{\varphi}_2^{\rm T}(t)\boldsymbol{\gamma}]^2,$$

$$J_{10}(\boldsymbol{d}):=\sum_{j=1}^{t}[w(j)-\boldsymbol{\varphi}_v^{\rm T}(j)\boldsymbol{d}]^2.$$

根据最小二乘原理,令 $J_8(\boldsymbol{\theta})$ 对 $\boldsymbol{\theta}$ 的偏导数,$J_9(\boldsymbol{\gamma})$ 对 $\boldsymbol{\gamma}$ 的偏导数,$J_{10}(\boldsymbol{d})$ 对 $\boldsymbol{d}$ 的偏导数都为零,未知变量用其估计代替,能够推导出 ON-EEMA 系统的滑动平均滤波递推最小二乘算法(Filtering based Recursive Least Squares algorithm, F-RLS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[\hat{y}_{\rm f}(t)-\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad(159)$$

$$\boldsymbol{L}_1(t)=\frac{\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)}{1+\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)},\quad(160)$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)]\boldsymbol{P}_1(t-1),\quad(161)$$

$$\hat{\boldsymbol{\gamma}}(t)=\hat{\boldsymbol{\gamma}}(t-1)+\boldsymbol{L}_2(t)[\hat{y}_{\rm f}(t)-\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\hat{\boldsymbol{b}}(t)-$$
$$\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)\hat{\boldsymbol{\gamma}}(t-1)],\quad(162)$$

$$\boldsymbol{L}_2(t)=\frac{\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)}{1+\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)},\quad(163)$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_m-\boldsymbol{L}_2(t)\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)]\boldsymbol{P}_2(t-1),\quad(164)$$

$$\hat{\boldsymbol{d}}(t)=\hat{\boldsymbol{d}}(t-1)+\boldsymbol{L}_3(t)[\hat{w}(t)-\hat{\boldsymbol{\varphi}}_v^{\rm T}(t)\hat{\boldsymbol{d}}(t-1)],\quad(165)$$

$$\boldsymbol{L}_3(t)=\frac{\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_v(t)}{1+\hat{\boldsymbol{\varphi}}_v^{\rm T}(t)\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_v(t)},\quad(166)$$

$$\boldsymbol{P}_3(t)=[\boldsymbol{I}_{n_d}-\boldsymbol{L}_3(t)\hat{\boldsymbol{\varphi}}_v^{\rm T}(t)]\boldsymbol{P}_3(t-1),\quad(167)$$

$$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{\gamma}}^{\rm T}(t-1)\hat{\boldsymbol{G}}^{\rm T}(t),\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)]^{\rm T},\quad(168)$$

$$\hat{\boldsymbol{\varphi}}_2(t)=\hat{\boldsymbol{G}}^{\rm T}(t)\hat{\boldsymbol{a}}(t),\quad(169)$$

$$\hat{\boldsymbol{G}}(t)=\begin{bmatrix}\hat{g}_1(t-1)&\hat{g}_2(t-1)&\cdots&\hat{g}_m(t-1)\\ \hat{g}_1(t-2)&\hat{g}_2(t-2)&\cdots&\hat{g}_m(t-2)\\ \vdots&\vdots&&\vdots\\ \hat{g}_1(t-n_a)&\hat{g}_2(t-n_a)&\cdots&\hat{g}_m(t-n_a)\end{bmatrix},\quad(170)$$

$$\hat{g}_j(t)=f_j(y(t))-\hat{d}_1(t)\hat{g}_j(t-1)-\hat{d}_2(t)\hat{g}_j(t-2)-\cdots-$$
$$\hat{d}_{n_d}(t)\hat{g}_j(t-n_d),\quad j=1,2,\cdots,m,\quad(171)$$

$$\hat{\boldsymbol{\varphi}}_{\rm f}(t)=[\hat{u}_{\rm f}(t-1),\hat{u}_{\rm f}(t-2),\cdots,\hat{u}_{\rm f}(t-n_b)]^{\rm T},\quad(172)$$

$$\hat{u}_{\rm f}(t)=u(t)-\sum_{i=1}^{n_d}\hat{d}_i(t)\hat{u}_{\rm f}(t-i),\quad(173)$$

$$\hat{y}_{\rm f}(t)=y(t)-\sum_{i=1}^{n_d}\hat{d}_i(t)\hat{y}_{\rm f}(t-i),\quad(174)$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\rm T},\quad(175)$$

$$\hat{\boldsymbol{\varphi}}_v(t)=[\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\rm T},\quad(176)$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\rm T}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t-1)-\boldsymbol{\varphi}_u^{\rm T}(t)\hat{\boldsymbol{b}}(t-1),\quad(177)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\rm T}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\rm T}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_v^{\rm T}(t)\hat{\boldsymbol{d}}(t),\quad(178)$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix},\quad\hat{\boldsymbol{\Theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{d}}(t)\\ \hat{\boldsymbol{\gamma}}(t)\end{bmatrix}.\quad(179)$$

F-RLS 算法的计算步骤如下:

1)初始化:令 $t=1$,给定数据长度和 L_e 和非线性函数 $f_j(*)$,置初值 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\gamma}}(0)$ 为随机向量,且 $\|\hat{\boldsymbol{\gamma}}(0)\|=1$,$\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$,$\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_m$,$\boldsymbol{P}_3(0)=p_0\boldsymbol{I}_{n_d}$,$p_0=10^6$.

2)收集输入输出数据 $u(t)$ 和 $y(t)$,用式(175)构造系统信息向量 $\boldsymbol{\varphi}_u(t)$.

3)用式(177)计算 $\hat{w}(t)$,用式(176)构造 $\hat{\boldsymbol{\varphi}}_v(t)$,分别用式(166)和(167)计算 $\boldsymbol{L}_3(t)$ 和 $\boldsymbol{P}_3(t)$.

4)通过式(165)刷新参数估计 $\hat{\boldsymbol{d}}(t)$.

5)分别用式(173)和(174)计算 $\hat{u}_{\rm f}(t)$ 和 $\hat{y}_{\rm f}(t)$,然后通过式(172)构造滤波信息向量 $\hat{\boldsymbol{\varphi}}_{\rm f}(t)$.

6)用式(171)计算 $\hat{g}_j(t)$,用式(170)计算 $\hat{\boldsymbol{G}}(t)$.

7)用式(168)计算 $\hat{\boldsymbol{\varphi}}_1(t)$,分别用式(160)和(161)构造 $\boldsymbol{L}_1(t)$ 和 $\boldsymbol{P}_1(t)$.

8)用式(159)刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$,并从 $\hat{\boldsymbol{\theta}}(t)$ 中读出 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{b}}(t)$.

9)用式(169)计算 $\hat{\boldsymbol{\varphi}}_2(t)$,分别用式(163)和(164)构造 $\boldsymbol{L}_2(t)$ 和 $\boldsymbol{P}_2(t)$.

10)通过式(162)刷新参数估计 $\hat{\boldsymbol{\gamma}}(t)$,同时将 $\hat{\boldsymbol{\gamma}}(t)$ 进行归一化:

$$\hat{\boldsymbol{\gamma}}(t)={\rm sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|},$$

用式(178)计算 $\hat{v}(t)$.

11)如果 $t\leqslant L_e$,t 增 1,转步骤 2),继续递推计算;否则,中断循环过程,获得参数估计 $\hat{\boldsymbol{\Theta}}(t)$.

ON-OEMA 系统的 O-RELS 算法(17)—(25)的

协方差阵 $\boldsymbol{P}(t)=\left[\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\right]^{-1}$ 的维数是 $(mn_a+n_b+n_d)\times(mn_a+n_b+n_d)$，D-RELS 算法(90)—(103)的 2 个协方差阵维数分别为 $(n_a+n_b)\times(n_a+n_b)$ 和 $(m+n_d)\times(m+n_d)$，F-RLS 算法(159)—(179)的 3 个协方差阵维数分别为 $(n_a+n_b)\times(n_a+n_b)$，$m\times m$ 和 $n_d\times n_d$，而递推最小二乘算法的计算量是 $O(n^2)$(n 为协方差阵的维数)，由于 $(n_1+n_2)^2\geqslant n_1^2+n_2^2$，$(n_1+n_2+n_3)^2\geqslant n_1^2+n_2^2+n_3^2(n_i\geqslant 1)$，所以 O-RELS 算法、D-RELS 算法和 F-RLS 算法的计算量依次减小.

4.2 自回归滑动平均滤波递推最小二乘辨识方法

考虑下列输出非线性方程误差自回归滑动平均(ON-EEARMA)系统：

$$y(t)=A(z)f(y(t))+B(z)u(t)+\frac{D(z)}{C(z)}v(t),\qquad(180)$$

$$f(y(t))=\gamma_1 f_1(y(t))+\gamma_2 f_2(y(t))+\cdots+\gamma_m f_m(y(t))=\boldsymbol{f}(y(t))\boldsymbol{\gamma},\qquad(181)$$

$$\boldsymbol{f}(y(t))=[f_1(y(t)),f_2(y(t)),\cdots,f_m(y(t))]\in\mathbf{R}^{1\times m},\qquad(182)$$

$$\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_m]^{\mathrm{T}}\in\mathbf{R}^m,\quad\|\boldsymbol{\gamma}\|=1,\quad\gamma_1>0.\qquad(183)$$

系统的干扰噪声是一个自回归滑动平均过程：

$$w(t):=\frac{D(z)}{C(z)}v(t)\in\mathbf{R}.\qquad(184)$$

令 $\hat{\boldsymbol{\theta}}(t):=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{b}}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b}$，$\hat{\boldsymbol{\gamma}}(t)\in\mathbf{R}^m$ 和 $\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t):=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\\hat{\boldsymbol{d}}(t)\end{bmatrix}\in\mathbf{R}^{n_c+n_d}$ 分别是参数向量 $\boldsymbol{\theta}:=\begin{bmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{bmatrix}\in\mathbf{R}^{n_a+n_b}$，$\boldsymbol{\gamma}\in\mathbf{R}^m$ 和 $\boldsymbol{\theta}_{\mathrm{n}}:=\begin{bmatrix}\boldsymbol{c}\\\boldsymbol{d}\end{bmatrix}\in\mathbf{R}^{n_c+n_d}$ 在时刻 t 的估计.

式(184)可以写为

$$\begin{aligned}w(t)&=[1-C(z)]w(t)+D(z)v(t)=\\&\boldsymbol{\varphi}_w^{\mathrm{T}}(t)\boldsymbol{c}+\boldsymbol{\varphi}_v^{\mathrm{T}}(t)\boldsymbol{d}+v(t)=\\&\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\end{aligned}\qquad(185)$$

利用式(184)和(185)，由式(180)可得系统辨识模型：

$$y(t)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+w(t)\qquad(186)$$

$$=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}+\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}+v(t).\qquad(187)$$

由式(186)有

$$w(t)=y(t)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}.$$

将上式中的参数向量 $\boldsymbol{a}$，$\boldsymbol{\gamma}$ 和 $\boldsymbol{b}$ 分别用它们在时刻 $t-1$ 的估计 $\hat{\boldsymbol{a}}(t-1)$，$\hat{\boldsymbol{\gamma}}(t-1)$ 和 $\hat{\boldsymbol{b}}(t-1)$ 代替，则 $w(t)$ 的估计 $\hat{w}(t)$ 可由下式计算：

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t-1)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1).$$

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

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

由式(187)可得

$$v(t)=y(t)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\gamma}-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\boldsymbol{b}-\boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}}.$$

上式中未知量 $\boldsymbol{a}$，$\boldsymbol{\gamma}$，$\boldsymbol{b}$，$\boldsymbol{\varphi}_{\mathrm{n}}(t)$，$\boldsymbol{\theta}_{\mathrm{n}}$ 分别用它们在时刻 t 的估计 $\hat{\boldsymbol{a}}(t)$，$\hat{\boldsymbol{\gamma}}(t)$，$\hat{\boldsymbol{b}}(t)$，$\hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t)$，$\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)$ 代替，那么 $v(t)$ 的估计 $\hat{v}(t)$ 可由下式计算：

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_{\mathrm{n}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t).$$

定义滤波输入 $u_{\mathrm{f}}(t)$，滤波输出 $y_{\mathrm{f}}(t)$，滤波信息向量 $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ 和滤波信息矩阵 $\boldsymbol{G}(t)$ 如下：

$$u_{\mathrm{f}}(t):=\frac{C(z)}{D(z)}u(t),\quad y_{\mathrm{f}}(t):=\frac{C(z)}{D(z)}y(t),$$

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

$$\boldsymbol{G}(t):=\begin{bmatrix}g_1(t-1)&g_2(t-1)&\cdots&g_m(t-1)\\g_1(t-2)&g_2(t-2)&\cdots&g_m(t-2)\\\vdots&\vdots&&\vdots\\g_1(t-n_a)&g_2(t-n_a)&\cdots&g_m(t-n_a)\end{bmatrix}\in\mathbf{R}^{n_a\times m},$$

$$\begin{aligned}g_j(t)&:=\frac{C(z)}{D(z)}f_j(y(t))=\\&[1-D(z)]g_j(y(t))+C(z)f_j(y(t))=\\&-d_1g_j(y(t-1))-d_2g_j(y(t-2))-\cdots-d_{n_d}g_j(y(t-n_d))+\\&f_j(y(t))+c_1f_j(y(t-1))+c_2f_j(y(t-2))+\cdots+\\&c_{n_c}f_j(y(t-n_c)),\quad j=1,2,\cdots,m.\end{aligned}$$

式(180)两边左乘 $\frac{C(z)}{D(z)}$ 得到

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

或

$$\begin{aligned}y_{\mathrm{f}}(t)&=A(z)\frac{C(z)}{D(z)}f(y(t))+B(z)u_{\mathrm{f}}(t)+v(t)=\\&A(z)[\gamma_1g_1(t)+\gamma_2g_2(t)+\cdots+\gamma_mg_m(t)]+\\&B(z)u_{\mathrm{f}}(t)+v(t)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{b}+v(t).\end{aligned}\qquad(188)$$

用噪声模型的参数向量：

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

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

构造 $C(z)$ 和 $D(z)$ 的估计：

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

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

由于 $C(z)$ 和 $D(z)$ 是未知的，所以用 $\hat{C}(t,z)$ 和 $\hat{D}(t,z)$ 分别对 $u(t)$ 和 $y(t)$ 滤波，得到其估计 $\hat{u}_{\mathrm{f}}(t)$

和$\hat{y}_{\rm f}(t)$：

$$\hat{u}_{\rm f}(t)=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}u(t),\quad \hat{y}_{\rm f}(t)=\frac{\hat{C}(t,z)}{\hat{D}(t,z)}y(t).$$

或

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

$\hat{u}_{\rm f}(t)$和$\hat{y}_{\rm f}(t)$可由下式计算

$$\begin{aligned}\hat{u}_{\rm f}(t)=&[1-\hat{D}(t,z)]\hat{u}_{\rm f}(t)+\hat{C}(t,z)u(t)=\\&-\hat{d}_1(t)\hat{u}_{\rm f}(t-1)-\hat{d}_2(t)\hat{u}_{\rm f}(t-2)-\cdots-\\&\hat{d}_{n_d}(t)\hat{u}_{\rm f}(t-n_d)+\hat{c}_1(t)u(t-1)+\\&\hat{c}_2(t)u(t-2)+\cdots+\hat{c}_{n_c}(t)u(t-n_c)+u(t),\end{aligned}$$

$$\begin{aligned}\hat{y}_{\rm f}(t)=&[1-\hat{D}(t,z)]\hat{y}_{\rm f}(t)+\hat{C}(t,z)y(t)=\\&-\hat{d}_1(t)\hat{y}_{\rm f}(t-1)-\hat{d}_2(t)\hat{y}_{\rm f}(t-2)-\cdots-\\&\hat{d}_{n_d}(t)\hat{y}_{\rm f}(t-n_d)+\hat{c}_1(t)y(t-1)+\\&\hat{c}_2(t)y(t-2)+\cdots+\hat{c}_{n_c}(t)y(t-n_c)+y(t).\end{aligned}$$

定义信息向量：

$$\boldsymbol{\varphi}_1(t):=\begin{bmatrix}\boldsymbol{G}(t)\boldsymbol{\gamma}\\ \boldsymbol{\varphi}_{\rm f}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}_2(t):=\boldsymbol{G}^{\rm T}(t)\boldsymbol{a}\in\mathbf{R}^m.$$

对于滤波辨识模型(188)和噪声模型(185),定义3个准则函数：

$$J_{11}(\boldsymbol{\theta}):=\sum_{j=1}^{t}[\hat{y}_{\rm f}(j)-\boldsymbol{\varphi}_1^{\rm T}(j)\boldsymbol{\theta}]^2,$$

$$J_{12}(\boldsymbol{\gamma}):=\sum_{j=1}^{t}[\hat{y}_{\rm f}(j)-\boldsymbol{\varphi}_{\rm f}^{\rm T}(j)\boldsymbol{b}-\boldsymbol{\varphi}_2^{\rm T}(j)\boldsymbol{\gamma}]^2,$$

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

极小化$J_{11}(\boldsymbol{\theta})$,$J_{12}(\boldsymbol{\gamma})$和$J_{13}(\boldsymbol{\theta}_{\rm n})$,未知变量用其估计代替,能够推导出ON-EEARMA系统的自回归滑动平均滤波递推最小二乘算法(Filtering based Recursive Least Squares algorithm,F-RLS算法)：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}_1(t)[\hat{y}_{\rm f}(t)-\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad(189)$$

$$\boldsymbol{L}_1(t)=\frac{\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)}{1+\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)\boldsymbol{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)},\quad(190)$$

$$\boldsymbol{P}_1(t)=[\boldsymbol{I}_{n_a+n_b}-\boldsymbol{L}_1(t)\hat{\boldsymbol{\varphi}}_1^{\rm T}(t)]\boldsymbol{P}_1(t-1),\quad(191)$$

$$\begin{aligned}\hat{\boldsymbol{\gamma}}(t)=&\hat{\boldsymbol{\gamma}}(t-1)+\boldsymbol{L}_2(t)[\hat{y}_{\rm f}(t)-\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\hat{\boldsymbol{b}}(t)-\\&\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)\hat{\boldsymbol{\gamma}}(t-1)],\end{aligned}\quad(192)$$

$$\boldsymbol{L}_2(t)=\frac{\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)}{1+\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)\boldsymbol{P}_2(t-1)\hat{\boldsymbol{\varphi}}_2(t)},\quad(193)$$

$$\boldsymbol{P}_2(t)=[\boldsymbol{I}_m-\boldsymbol{L}_2(t)\hat{\boldsymbol{\varphi}}_2^{\rm T}(t)]\boldsymbol{P}_2(t-1),\quad(194)$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t)=\hat{\boldsymbol{\theta}}_{\rm n}(t-1)+\boldsymbol{L}_3(t)[\hat{w}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t-1)],\quad(195)$$

$$\boldsymbol{L}_3(t)=\frac{\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_{\rm n}(t)}{1+\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\boldsymbol{P}_3(t-1)\hat{\boldsymbol{\varphi}}_{\rm n}(t)},\quad(196)$$

$$\boldsymbol{P}_3(t)=[\boldsymbol{I}_{n_c+n_d}-\boldsymbol{L}_3(t)\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)]\boldsymbol{P}_3(t-1),\quad(197)$$

$$\hat{\boldsymbol{\varphi}}_1(t)=[\hat{\boldsymbol{\gamma}}^{\rm T}(t-1)\hat{\boldsymbol{G}}^{\rm T}(t),\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)]^{\rm T},\quad(198)$$

$$\hat{\boldsymbol{\varphi}}_2(t)=\hat{\boldsymbol{G}}^{\rm T}(t)\hat{\boldsymbol{a}}(t),\quad(199)$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\rm n}(t)=&[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c),\\&\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)]^{\rm T},\end{aligned}\quad(200)$$

$$\hat{\boldsymbol{\varphi}}_{\rm f}(t)=[\hat{u}_{\rm f}(t-1),\hat{u}_{\rm f}(t-2),\cdots,\hat{u}_{\rm f}(t-n_b)]^{\rm T},\quad(201)$$

$$\hat{\boldsymbol{G}}(t)=\begin{bmatrix}\hat{g}_1(t-1)&\hat{g}_2(t-1)&\cdots&\hat{g}_m(t-1)\\\hat{g}_1(t-2)&\hat{g}_2(t-2)&\cdots&\hat{g}_m(t-2)\\\vdots&\vdots&&\vdots\\\hat{g}_1(t-n_a)&\hat{g}_2(t-n_a)&\cdots&\hat{g}_m(t-n_a)\end{bmatrix},\quad(202)$$

$$\begin{aligned}\hat{g}_j(t)=&f_j(y(t))+\hat{c}_1(t)f_j(y(t-1))+\hat{c}_2(t)f_j(y(t-2))+\\&\cdots+\hat{c}_{n_c}(t)f_j(y(t-n_c))-\hat{d}_1(t)\hat{g}_j(t-1)-\\&\hat{d}_2(t)\hat{g}_j(t-2)-\cdots-\hat{d}_{n_d}(t)\hat{g}_j(t-n_d),\end{aligned}\quad(203)$$

$$\hat{u}_{\rm f}(t)=-\sum_{i=1}^{n_d}\hat{d}_i(t)\hat{u}_{\rm f}(t-i)+\sum_{i=1}^{n_c}\hat{c}_i(t)u(t-i)+u(t),\quad(204)$$

$$\hat{y}_{\rm f}(t)=-\sum_{i=1}^{n_d}\hat{d}_i(t)\hat{y}_{\rm f}(t-i)+\sum_{i=1}^{n_c}\hat{c}_i(t)y(t-i)+y(t),\quad(205)$$

$$\boldsymbol{\varphi}_u(t)=[u(t-1),u(t-2),\cdots,u(t-n_b)]^{\rm T},\quad(206)$$

$$\hat{w}(t)=y(t)-\hat{\boldsymbol{a}}^{\rm T}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t-1)-\boldsymbol{\varphi}_u^{\rm T}(t)\hat{\boldsymbol{b}}(t-1),\quad(207)$$

$$\hat{v}(t)=y(t)-\hat{\boldsymbol{a}}^{\rm T}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\gamma}}(t)-\boldsymbol{\varphi}_u^{\rm T}(t)\hat{\boldsymbol{b}}(t)-\hat{\boldsymbol{\varphi}}_{\rm n}^{\rm T}(t)\hat{\boldsymbol{\theta}}_{\rm n}(t),\quad(208)$$

$$\hat{\boldsymbol{\theta}}(t)=\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\\hat{\boldsymbol{b}}(t)\end{bmatrix},\quad\hat{\boldsymbol{\theta}}_{\rm n}(t)=\begin{bmatrix}\hat{\boldsymbol{c}}(t)\\\hat{\boldsymbol{d}}(t)\end{bmatrix},\quad(209)$$

$$\hat{\boldsymbol{\Theta}}(t)=[\hat{\boldsymbol{\theta}}^{\rm T}(t),\hat{\boldsymbol{\gamma}}^{\rm T}(t),\hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t)]^{\rm T}.\quad(210)$$

F-RLS算法(189)—(202)的计算步骤如下：

1)初始化：令$t=1$,给定非线性函数$f_j(*)$,置初值$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_a+n_b}/p_0$,$\hat{\boldsymbol{\gamma}}(0)$为随机向量,且$\|\hat{\boldsymbol{\gamma}}(0)\|=1$,$\hat{\boldsymbol{\theta}}_{\rm n}(0)=\mathbf{1}_{n_c+n_d}/p_0$,$\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_a+n_b}$,$\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_m$,$\boldsymbol{P}_3(0)=p_0\boldsymbol{I}_{n_c+n_d}$,$p_0=10^6$.

2)收集输入输出数据$u(t)$和$y(t)$,用式(207)计算$\hat{w}(t)$,分别用式(206)和(200)构造向量$\boldsymbol{\varphi}_u(t)$和$\hat{\boldsymbol{\varphi}}_{\rm n}(t)$.

3)分别用式(196)和(197)计算$\boldsymbol{L}_3(t)$和$\boldsymbol{P}_3(t)$.

4)通过式(195)刷新参数估计$\hat{\boldsymbol{\theta}}_{\rm n}(t)$.

5)分别用式(204)和(205)计算$\hat{u}_{\rm f}(t)$和$\hat{y}_{\rm f}(t)$,然后通过式(201)构造滤波信息向量$\hat{\boldsymbol{\varphi}}_{\rm f}(t)$.

6)用式(203)计算$\hat{g}_j(t)$,用式(202)计算$\hat{\boldsymbol{G}}(t)$,

用式（198）计算 $\hat{\boldsymbol{\varphi}}_1(t)$.

7）分别用式（190）和（191）计算 $\boldsymbol{L}_1(t)$ 和 $\boldsymbol{P}_1(t)$.

8）用式（189）刷新参数估计 $\hat{\boldsymbol{\theta}}(t)$，并从 $\hat{\boldsymbol{\theta}}(t)$ 中读出 $\hat{\boldsymbol{a}}(t)$ 和 $\hat{\boldsymbol{b}}(t)$.

9）用式（199）计算 $\hat{\boldsymbol{\varphi}}_2(t)$，分别用式（193）和（194）计算 $\boldsymbol{L}_2(t)$ 和 $\boldsymbol{P}_2(t)$.

10）通过式（192）刷新参数估计 $\hat{\boldsymbol{\gamma}}(t)$，同时将 $\hat{\boldsymbol{\gamma}}(t)$ 进行归一化：

$$\hat{\boldsymbol{\gamma}}(t)=\mathrm{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|},$$

用式（208）计算 $\hat{v}(t)$.

11）t 增 1，转步骤 2），继续递推计算.

F-RLS 算法（189）—（210）的计算参数估计 $\hat{\boldsymbol{\Theta}}(t)$ 的流程如图 5 所示.

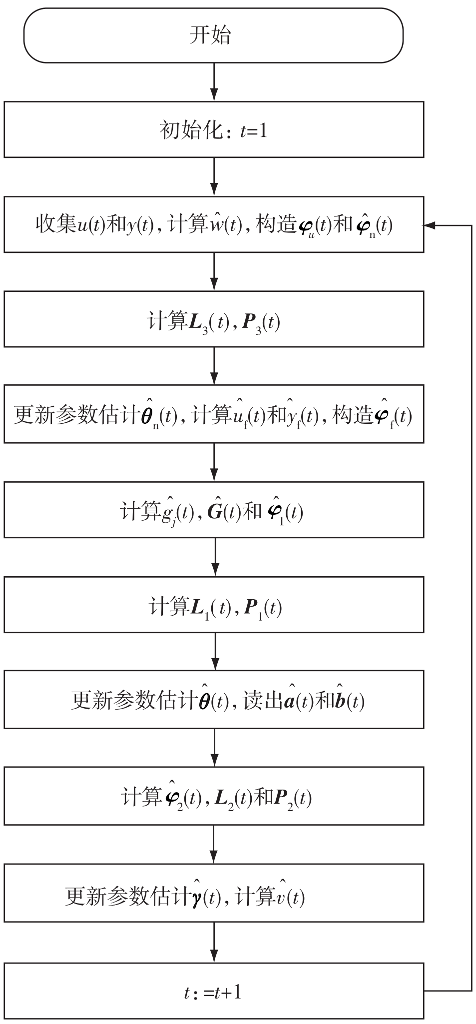

图 5 F-RLS 算法估计 $\hat{\boldsymbol{\Theta}}(t)$ 的流程

Fig. 5 The flowchart of computing the F-RLS parameter estimate $\hat{\boldsymbol{\Theta}}(t)$

5 结语

针对输出非线性方程误差类系统，提出了基于过参数化模型的递推最小二乘类辨识方法，提出了基于模型分解的递推最小二乘类辨识方法，提出了基于数据滤波的递推最小二乘辨识方法；进一步可研究基于过参数化模型的随机梯度类辨识方法、多新息随机梯度类辨识方法、梯度迭代辨识方法、最小二乘迭代辨识方法，研究基于模型分解的随机梯度辨识方法、多新息随机梯度类辨识方法、梯度迭代辨识方法、最小二乘迭代辨识方法，研究基于数据滤波的随机梯度类辨识方法、多新息随机梯度类辨识方法、梯度迭代辨识方法、最小二乘迭代辨识方法.文中的辨识思想和辨识方法可以推广到输出非线性输出误差类系统、输入非线性方程误差类系统、输入非线性输出误差类系统、输入非线性反馈系统、输出非线性反馈系统等.

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

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

[13] 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

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

[15] 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

[16] 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

[17] 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

[18] Ding F,Deng K P,Liu X M.Decomposition based Newton iterative identification method for a Hammerstein nonlinear FIR system with ARMA noise[J].Circuits,Systems and Signal Processing,2014,33(9):2881-2893

[19] 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

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

[21] Chen H B, Ding F. Hierarchical least squares identification for Hammerstein nonlinear controlled autoregressive systems[J].Circuit,System and Signal Processing,2015,34(1):61-75

[22] Chen H B,Xiao Y S,Ding F.Hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle[J].Applied Mathematics and Computation, 2014, 247: 1202-1210

[23] Chen H B,Ding F,Xiao Y S.Decomposition based parameter estimation algorithm for input nonlinear systems using the key term separation technique[J].Nonlinear Dynamics,2015,79(3):2027-2035

[24] Hu Y B,Liu B L,Zhou Q.A multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems[J].Applied Mathematics and Computation,2014,247:218-224

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

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

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

[28] 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

Recursive least squares identification methods for output nonlinear equation-error type systems

DING Feng[1,2,3] CHEN Qijia[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 With the development of control technology,the scales of the control systems become larger and larger, so does the computational load of the identification algorithms.For nonlinear systems with complex structures,especially for the nonlinear systems that contain the products of the unknown parameters of the nonlinear part and linear part,the sizes of the involved matrices in the over-parameterization model based least squares methods greatly increase,this makes the computational amount of the identification algorithms increase dramatically.Therefore, it is necessary to explore new parameter estimation methods with less computation.For output nonlinear equation-error type systems,this paper discusses the over-parameterization model based recursive least squares type identification algorithms; in order to reduce computational loads and improve the identification accuracy,this paper uses the decomposition technique and the filtering technique and presents the model decomposition based recursive least squares identification methods and the filtering based recursive least squares 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;least squares;model decomposition;data filtering;auxiliary model identification idea,multi-innovation identification theory;hierarchical identification principle;coupling identification concept;input nonlinear system;output nonlinear system