



## 辅助模型辨识方法(4):基本思想与梯度辨识

### 摘要

辅助模型辨识思想是针对系统存在一些不可测变量的辨识问题提出的,它是研究含有未知变量的线性系统、非线性系统辨识的重要手段.本文阐述了辅助模型辨识思想,针对白噪声干扰的输入非线性有限脉冲响应系统,研究辅助模型梯度辨识方法、辅助模型多新息梯度辨识方法、变递推间隔辅助模型多新息梯度辨识方法,以及派生的变递推间隔辅助模型梯度辨识方法、等递推间隔辅助模型梯度辨识方法和等递推间隔辅助模型多新息梯度辨识方法.

### 关键词

参数估计;递推辨识;梯度搜索;关键项分离;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;滤波辨识理念;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 2016-07-28

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

### 0 引言

辅助模型辨识方法源于作者的硕士学位论文《多变量系统辨识》(1990年)<sup>[1]</sup>,其首篇辅助模型辨识研究论文《传递函数阵子模型参数递推估计——辅助模型方法》发表在1991年《控制与决策》第6期上<sup>[2]</sup>.后经深入研究,多变量输出误差系统的辅助模型递推最小二乘辨识算法及其收敛性分析先后发表在1992年《清华大学学报(自然科学版)》<sup>[3]</sup>和1997年《控制理论与应用》<sup>[4]</sup>上.笔者于1994年在清华大学获得博士学位(优秀博士学位论文)后留校任教,于2002年6月赴加拿大University of Alberta(阿尔伯塔大学)做博士后,一直潜心从事系统辨识研究,将辅助模型方法加以发展,用于研究双率采样数据输出误差系统的辨识,提出了双率系统的辅助模型递推最小二乘算法和辅助模型随机梯度算法,并研究了其收敛性,研究成果陆续发表在控制领域国际著名期刊《Automatica》<sup>[5]</sup>和《IEEE Transactions on Automatic Control》<sup>[6]</sup>上.

20余年来,笔者及其硕士生和博士生在辅助模型辨识方面进行了不懈的研究,发表了大量论文.随着研究的深入和科学思想的升华,提炼和形成辅助模型辨识思想已是必然.辅助模型辨识思想(auxiliary model identification idea)是针对存在不可测变量的系统辨识的困难而提出的,它是利用系统的可测信息(包括计算得到的信息),建立一个辅助模型,用辅助模型的输出代替系统的不可测变量的一种辨识方法.辅助模型辨识思想的本质是用辅助模型输出代替辨识算法中的未知变量,可简单解释为“代替思想”.如今,辅助模型辨识思想已经推广用于线性输出误差系统<sup>[7-14]</sup>,双率、多率和非均匀采样数据系统<sup>[15-17]</sup>,稀少量测数据系统<sup>[18-19]</sup>,损失数据系统<sup>[20]</sup>,以及线性参数系统<sup>[21-24]</sup>和非线性系统的辨识<sup>[25-26]</sup>.

最近的辅助模型辨识包括自回归输出误差系统的辅助模型递推辨识方法<sup>[27]</sup>、输入非线性输出误差系统的辅助模型递推辨识方法<sup>[28]</sup>、输入非线性输出误差自回归系统辅助模型递推辨识方法<sup>[29]</sup>.本文阐述辅助模型辨识思想及其在输入非线性脉冲响应系统中的应用.辅助模型辨识思想是针对系统存在一些不可测变量的辨识问题提出的,它是研究含有未知变量的线性系统、非线性系统辨识的重要手段.尽管系统包含了大量的测量信息,但是仍然还存在一些关键不可测变量,对于这类系统的辨识,可以用系统的测量信息(包括计算得到

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

的信息)建立辅助模型,用辅助模型的输出代替这些不可测变量,从而使得辨识问题得到解决.本文针对白噪声干扰的输入非线性有限脉冲响应系统,研究辅助模型梯度辨识方法、辅助模型多新息梯度辨识方法、变递推间隔辅助模型多新息梯度辨识方法,以及派生的变递推间隔辅助模型梯度辨识方法、等递推间隔辅助模型梯度辨识方法和等递推间隔辅助模型多新息梯度辨识方法.这些方法可以联合迭代方法<sup>[30-32]</sup>来研究其他有色噪声干扰系统的辨识问题,如双率和多率采样数据系统<sup>[33-34]</sup>、双线性系统<sup>[35]</sup>、非线性系统<sup>[36-41]</sup>.

## 1 辅助模型辨识思想

实际工业系统中,经常有一些变量的值不可能得到(如系统的状态、内部变量或系统的某些输出值),对于这类存在一些不可测变量的系统,如何估计其参数,就是辅助模型辨识方法要回答的问题.

如果系统涉及内部变量  $x(t)$ ,它是系统输入  $u(t)$ 与(或)其他变量,以及参数向量  $\theta$ 的函数,其关系可以表示为

$$x(t) = f[\theta, u(t), u(t-1), \dots, u(t-n)], \quad (1)$$

当包含自回归项时,可以表示为

$$x(t) = f[\theta, x(t-1), x(t-2), \dots, x(t-n), u(t), u(t-1), \dots, u(t-n)], \quad (2)$$

系统输出  $y(t)$ 是内部变量  $x(t)$ 与系统输入  $u(t)$ 或其他变量,以及参数向量  $\rho$ 和可能的干扰  $v(t)$ 的函数,其关系可以表示为

$$y(t) = g[\rho, x(t), x(t-1), \dots, x(t-n), u(t), u(t-1), \dots, u(t-n)] + v(t), \quad (3)$$

当包含自回归项时,可以表示为

$$y(t) = g[\rho, y(t-1), y(t-2), \dots, y(t-n), x(t), x(t-1), \dots, x(t-n), u(t), u(t-1), \dots, u(t-n)] + v(t), \quad (4)$$

式中  $v(t)$ 是零均值随机干扰噪声.

式(1)和(3),式(2)和(3),或式(2)和(4)构成了辅助模型方法的基本辨识模型.典型的离散时间线性状态空间系统是该辨识模型的特例.

为方便起见,记

$$x(k:k-n) = (x(k), x(k-1), \dots, x(k-n)), \quad (5)$$

$$u(k:k-n) = (u(k), u(k-1), \dots, u(k-n)), \quad (6)$$

$$y(k:k-n) = (y(k), y(k-1), \dots, y(k-n)). \quad (7)$$

使用这种记号,辨识模型(1)和(3)可以写为

$$x(t) = f[\theta, u(t:t-n)], \quad (8)$$

$$y(t) = g[\rho, x(t:t-n), u(t:t-n)] + v(t). \quad (9)$$

辨识模型(2)和(3)可以写为

$$x(t) = f[\theta, x(t-1:t-n), u(t:t-n)], \quad (10)$$

$$y(t) = g[\rho, x(t:t-n), u(t:t-n)] + v(t). \quad (11)$$

更一般的辨识模型(2)和(4)可以写为

$$x(t) = f[\theta, x(t-1:t-n), u(t:t-n)], \quad (12)$$

$$y(t) = g[\rho, y(t-1:t-n), x(t:t-n), u(t:t-n)] + v(t). \quad (13)$$

因为模型(8)—(9)和模型(10)—(11)是模型(12)—(13)的特殊情形,故讨论辨识模型(12)—(13)更具有一般性.

对于辨识模型(12)—(13):

1) 可以在约束条件(12)下,通过极小化梯度准则函数

$$J_1(\theta, \rho) := \{y(t) - g[\rho, y(t-1:t-n), x(t:t-n), u(t:t-n)]\}^2,$$

$$x(t) = f[\theta, x(t-1:t-n), u(t:t-n)],$$

研究辅助模型随机梯度辨识方法.

2) 可以通过极小化递增数据长度的约束最小二乘准则函数

$$J_2(\theta, \rho) := \sum_{j=1}^t \{y(j) - g[\rho, y(j-1:j-n),$$

$$x(j:j-n), u(j:j-n)]\}^2,$$

$$x(j) = f[\theta, x(j-1:j-n), u(j:j-n)],$$

研究辅助模型递推梯度辨识方法、辅助模型最小二乘辨识方法、辅助模型递推最小二乘辨识方法、辅助模型牛顿递推辨识方法.

3) 可以通过极小化有限数据长度  $L$ 的约束最小二乘准则函数

$$J_3(\theta, \rho) := \sum_{t=1}^L \{y(t) - g[\rho, y(t-1:t-n),$$

$$x(t:t-n), u(t:t-n)]\}^2,$$

$$x(t) = f[\theta, x(t-1:t-n), u(t:t-n)],$$

研究辅助模型梯度迭代辨识方法、辅助模型最小二乘迭代辨识方法、辅助模型牛顿迭代辨识方法.

4) 可以通过极小化滑动数据窗的约束准则函数

$$J_4(\theta, \rho) := \sum_{j=t-p+1}^t \{y(j) - g[\rho, y(j-1:j-n),$$

$$x(j:j-n), u(j:j-n)]\}^2,$$

$$x(j) = f[\theta, x(j-1:j-n), u(j:j-n)],$$

研究辅助模型多新息随机梯度辨识方法、辅助模型梯度迭代辨识方法、辅助模型最小二乘迭代辨识方法、辅助模型牛顿递推辨识方法、辅助模型牛顿迭代辨识方法等.

由于准则函数  $J_1(\boldsymbol{\theta}, \boldsymbol{\rho}), J_2(\boldsymbol{\theta}, \boldsymbol{\rho}), J_3(\boldsymbol{\theta}, \boldsymbol{\rho}), J_4(\boldsymbol{\theta}, \boldsymbol{\rho})$  包含了未知变量  $x(t)$  或  $x(j)$ , 使得辨识优化问题极为困难. 一种可行的解决方案是用系统的可测信息(包括计算得到的信息)建立一个辅助模型, 用辅助模型的输出  $x_a(t)$  代替这些未知变量  $x(t)$ , 从而使辨识问题得到解决. 这种辨识中的“代替思想”称为辅助模型辨识思想(auxiliary model identification idea). 基于辅助模型辨识思想的辨识方法称为辅助模型辨识方法或基于辅助模型的辨识方法(auxiliary model based identification method).

准则函数  $J_1(\boldsymbol{\theta}, \boldsymbol{\rho}), J_2(\boldsymbol{\theta}, \boldsymbol{\rho}), J_3(\boldsymbol{\theta}, \boldsymbol{\rho}), J_4(\boldsymbol{\theta}, \boldsymbol{\rho})$  通常不是一个二次优化问题, 而是带约束条件的非线性优化问题, 求解起来特别困难. 下面使用梯度搜索, 以白噪声干扰的输入非线性有限脉冲响应系统为例, 基于辅助模型辨识思想, 研究一类线性参数系统的辅助模型梯度辨识方法.

## 2 辅助模型梯度辨识方法

梯度搜索可以用于线性回归系统(即线性参数系统)和非线性参数系统的辨识方法研究. 本节以输入非线性有限脉冲响应系统为例, 讨论一类线性参数系统的辅助模型的梯度辨识方法, 包括辅助模型投影算法、辅助模型广义投影算法、辅助模型随机梯度算法等.

### 2.1 系统描述与辨识模型

考虑由输入非线性有限脉冲响应模型(Input Nonlinear Finite Impulse Response model, IN-FIR 模型)描述的非线性系统(图 1), 它是由一个已知基静态非线性环节串联一个线性 FIR 子系统构成的, 其线性环节是一个 FIR 模型:

$$y(t) = B(z)x(t) + v(t), \quad (14)$$

$x(t) \in \mathbf{R}$  和  $y(t) \in \mathbf{R}$  分别为线性动态 FIR 子系统的输入和输出,  $v(t) \in \mathbf{R}$  是零均值白噪声,  $B(z)$  是移位算子  $z^{-1}[z^{-1}y(t) = y(t-1)]$  的多项式, 即

$$B(z) := b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b}.$$

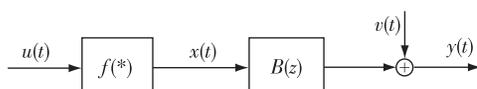


图 1 输入非线性有限脉冲响应(IN-FIR)系统

Fig. 1 An input nonlinear finite impulse response system

非线性环节的输出  $x(t)$  是系数为  $c_j$  的一组已知基函数  $f_j(u(t))$  的线性组合, 即

$$x(t) := f(u(t)) = \sum_{j=1}^m c_j f_j(u(t)), \quad (15)$$

$u(t)$  为系统输入, 也是非线性环节的输入. 非线性基函数  $f_j(*)$  可以是多项式、三角函数、指数函数、分段线性函数等. 式(14)–(15)描述的系统也称为输入非线性 FIR 系统. 把图 1 中  $f(*)$  和  $B(z)$  调换位置, 就得到输出非线性 FIR 系统.

在输入非线性 FIR 系统(14)–(15)中, 中间变量或内部变量(即非线性环节的输出, 也是线性环节的输入)  $x(t) \in \mathbf{R}$  是不可测的. 辨识目标是假设基函数  $f_j(*)$  已知, 利用观测数据  $\{u(t), y(t)\}$ , 提出辨识方法估计线性动态子系统的参数  $b_i$  和非线性环节的参数  $c_j$ .

式(14)可以等价写为

$$y(t) = B(z)x(t) + v(t) = [B(z)/\alpha][\alpha x(t)] + v(t).$$

由此可见: 对于任意非零常数  $\alpha$ ,  $(B(z)/\alpha, \alpha x(t))$  和  $(B(z), x(t))$ , 系统有相同的输入输出关系, 因此多项式  $B(z)$  的参数  $b_i$  和非线性函数  $x(t) = f(u(t))$  的系数  $c_j$  不是唯一的, 即  $b_i$  和  $c_j$  不是参数可辨识的<sup>[42]</sup>. 为了保证参数可辨识, 必须对  $x(t)$  或  $B(z)$  进行归一化. 不同的归一化假设可能导致不同的辨识方法. 典型的归一化假设有

- 1)  $b_0 = 1$  或  $c_1 = 1$ , 即多项式  $B(z)$  或线性函数  $x(t) = f(u(t))$  的第 1 个非零系数为 1, 这相当于取  $\alpha = b_0$ , 且令  $B(z) := B(z)/\alpha$  或  $\alpha = 1/c_1$ , 且令  $f(*) := \alpha f(*)$ . 其实可以规范化任一非零系数  $b_i$  或  $c_i$  为 1.
- 2)  $B(1) = 1$ , 即线性动态子系统的增益为 1.
- 3)  $b_0^2 + b_1^2 + \dots + b_{n_b}^2 = 1, b_0 > 0$ , 或  $c_1^2 + c_2^2 + \dots + c_m^2 = 1, c_1 > 0$ .

本文采用第 1) 个假设  $b_0 = 1$ . 利用移位算子  $z^{-1}$  的性质, 将多项式  $B(z)$  的表达式代入式(14)整理可得

$$y(t) = (b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b})x(t) + v(t) = b_0 x(t) + b_1 x(t-1) + b_2 x(t-2) + \dots + b_{n_b} x(t-n_b) + v(t) = x(t) + b_1 x(t-1) + b_2 x(t-2) + \dots + b_{n_b} x(t-n_b) + v(t). \quad (16)$$

定义参数向量  $\boldsymbol{\theta}$  和  $\boldsymbol{\rho}$  及信息向量  $\boldsymbol{\phi}(t)$  和  $\boldsymbol{\psi}(t)$  如下:

$$\boldsymbol{\theta} := \mathbf{c} = [c_1, c_2, \dots, c_m]^T \in \mathbf{R}^m, \quad (17)$$

$$\boldsymbol{\rho} := \mathbf{b} = [b_1, b_2, \dots, b_{n_b}]^T \in \mathbf{R}^{n_b}, \quad (18)$$

$$\boldsymbol{\phi}(t) := \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m, \quad (19)$$

$$\boldsymbol{\psi}(t) := [x(t-1), x(t-2), \dots, x(t-n_b)]^T \in \mathbf{R}^{n_b}. \quad (20)$$

按照式(5)—(7)的定义,式(20)可以简单写为  
 $\boldsymbol{\psi}(t) := x(t-1:t-n_b) \in \mathbf{R}^{n_b}$ .

于是, IN-FIR 系统(15)和(14)可写成线性回归形式:

$$x(t) = \boldsymbol{\phi}^T(t) \boldsymbol{\theta}, \quad (21)$$

$$y(t) = x(t) + \boldsymbol{\psi}^T(t) \boldsymbol{\rho} + v(t) =$$

$$\begin{aligned} & \boldsymbol{\phi}^T(t) \boldsymbol{\theta} + \boldsymbol{\psi}^T(t) \boldsymbol{\rho} + v(t) = [\boldsymbol{\phi}^T(t), \boldsymbol{\psi}^T(t)] \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{bmatrix} + v(t) = \\ & \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta} + v(t), \end{aligned} \quad (22)$$

其中系统的参数向量为

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{bmatrix} \in \mathbf{R}^n, \quad n := n_b + m, \quad (23)$$

系统的信息向量为

$$\boldsymbol{\varphi}(t) := \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x(t-1:t-n_b) \end{bmatrix} \in \mathbf{R}^n, \quad (24)$$

$$\boldsymbol{\phi}(t) := \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m, \quad (25)$$

$$\boldsymbol{\psi}(t) := [x(t-1), x(t-2), \dots, x(t-n_b)]^T \in \mathbf{R}^{n_b}. \quad (26)$$

式(21)—(22)称为 IN-FIR 系统的辨识模型 (identification model), 其中  $y(t)$  和  $u(t)$  是可测的输出和输入,  $x(t)$  是中间变量 (未知内部变量),  $\boldsymbol{\vartheta}$  是系统的参数向量,  $\boldsymbol{\theta}$  是静态非线性环节的参数向量,  $\boldsymbol{\rho}$  是动态 FIR 子系统的参数向量,  $v(t)$  是零均值随机白噪声.

## 2.2 IN-FIR 系统的辅助模型建立

设  $\hat{X}(t)$  为  $X$  在时刻  $t$  的估计, 这意味着  $\hat{\boldsymbol{\theta}}(t)$  是  $\boldsymbol{\theta}$  在时刻  $t$  的估计,  $\hat{\boldsymbol{\rho}}(t)$  是  $\boldsymbol{\rho}$  在时刻  $t$  的估计,  $\hat{\boldsymbol{\vartheta}}(t) := \begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \hat{\boldsymbol{\rho}}(t) \end{bmatrix} \in \mathbf{R}^n$  是  $\boldsymbol{\vartheta} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{bmatrix}$  在时刻  $t$  的估计,  $\hat{v}(t)$  是  $v(t)$  的估计.

辨识模型(21)—(22)中的信息向量  $\boldsymbol{\varphi}(t)$  是未知的, 因为其包含了已知基函数构成的非线性基函数向量  $\mathbf{f}(u(t))$ , 还包含了未知内部变量  $x(t-1:t-n_b)$ , 即  $x(t-i)$ , 这是 IN-FIR 系统辨识的困难. 解决这一困难的途径是利用辅助模型辨识思想, 用系统的可测信息 (包括计算获得的信息) 建立一个辅助模型, 用辅助模型的输出代替这些未知变量, 从而获得 IN-FIR 系统的辅助模型辨识方法.

具体到 IN-FIR 系统, 就是用  $u(t)$  作为输入, 建立一个静态辅助模型, 如图 2 所示, 其中  $f_a(u(t))$  是与非线性函数  $f(u(t))$  解析结构式相同、系数不同的静态辅助模型. 根据式(15)中  $x(t)$  表达式的结构, 假

设辅助模型的参数向量为  $\boldsymbol{\theta}_a(t) = [\theta_{a1}(t), \theta_{a2}(t), \dots, \theta_{am}(t)]^T \in \mathbf{R}^m$ , 用系统输入  $u(t)$  和辅助模型参数  $\boldsymbol{\theta}_a(t)$  构造估算  $x(t)$  的非线性辅助模型:

$$x_a(t) = f_a(u(t)) = \sum_{j=1}^m \theta_{aj}(t) f_j(u(t)) = \mathbf{f}(u(t)) \boldsymbol{\theta}_a(t).$$

这里采用递推方案实现提出的辅助模型辨识算法 (当然也可以用迭代方案). 因此, 一种方案是用参数向量  $\boldsymbol{\theta}$  的估计  $\hat{\boldsymbol{\theta}}(t) := [\hat{c}_1(t), \hat{c}_2(t), \dots, \hat{c}_m(t)]^T \in \mathbf{R}^m$  作为辅助模型的参数向量  $\boldsymbol{\theta}_a(t)$ , 即  $\boldsymbol{\theta}_a(t) := \hat{\boldsymbol{\theta}}(t)$ , 这样上式的辅助模型可以表示为

$$x_a(t) = f_a(u(t)) = \sum_{j=1}^m \hat{c}_j(t) f_j(u(t)) = \mathbf{f}(u(t)) \hat{\boldsymbol{\theta}}(t). \quad (27)$$

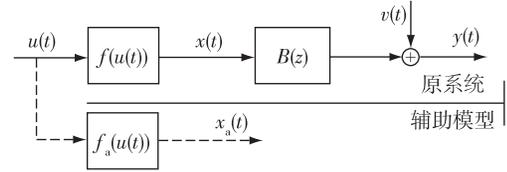


图 2 带辅助模型的 IN-FIR 系统

Fig. 2 The IN-FIR system with an auxiliary model

辅助模型的输出  $x_a(t)$  可作为  $x(t)$  的估计  $\hat{x}(t)$ , 即  $\hat{x}(t) := x_a(t)$ . 用辅助模型的输出  $x_a(t-i)$  构造信息向量  $\hat{\boldsymbol{\psi}}(t)$  的估计

$$\hat{\boldsymbol{\psi}}(t) := [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b) \in \mathbf{R}^{n_b}. \quad (28)$$

根据式(24)中  $\boldsymbol{\varphi}(t)$  的定义和上述辅助模型, 用非线性基函数向量  $\mathbf{f}(u(t))$  和  $\hat{\boldsymbol{\psi}}(t)$  的定义  $\hat{\boldsymbol{\varphi}}(t)$  的估计

$$\hat{\boldsymbol{\varphi}}(t) := \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix} \in \mathbf{R}^n. \quad (29)$$

由于辨识模型(21)—(22)中信息向量  $\boldsymbol{\varphi}(t)$  包含了未知变量  $x(t-i)$ , 而在对应的辨识算法中, 我们使用辅助模型的输出  $x_a(t-i)$  和估计的信息向量  $\hat{\boldsymbol{\varphi}}(t)$ , 从而使得辨识问题得到解决. 这就是辅助模型辨识思想的奇妙之处. IN-FIR 系统的辅助模型总结如下:

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (30)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix} \in \mathbf{R}^n, \quad (31)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m, \quad (32)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T =$$

$$x_a(t-1:t-n_b) \in \mathbf{R}^{n_b}. \quad (33)$$

下面借助于上述辅助模型,研究辅助模型投影辨识方法、辅助模型广义投影算法、辅助模型随机梯度辨识方法。

### 2.3 辅助模型投影辨识算法

对于辨识模型(21)–(22),定义梯度准则函数

$$J_5(\boldsymbol{\vartheta}) = \frac{1}{2} [y(t) - \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta}]^2.$$

设  $\mu(t) \geq 0$  为步长,  $\|\mathbf{X}\|^2 = \text{tr}[\mathbf{X}\mathbf{X}^T]$ . 采用负梯度搜索(negative gradient search)(参见文献[42]第5章中负梯度搜索算法),可得梯度递推关系:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) - \mu(t) \text{grad}_{\boldsymbol{\vartheta}} [J(\hat{\boldsymbol{\vartheta}}(t-1))] = \hat{\boldsymbol{\vartheta}}(t-1) + \mu(t) \boldsymbol{\varphi}(t) [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)]. \quad (34)$$

将  $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t)$  代入准则函数  $J_5(\boldsymbol{\vartheta})$  中,可得

$$g[\mu(t)] := 2J_5[\hat{\boldsymbol{\vartheta}}(t)] = \{y(t) - \boldsymbol{\varphi}^T(t) [\hat{\boldsymbol{\vartheta}}(t-1) + \mu(t) \boldsymbol{\varphi}(t) [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)]]\}^2 = [1 - \mu(t) \|\boldsymbol{\varphi}(t)\|^2] [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)]^2.$$

极小化  $g[\mu(t)]$ , 令其导数  $g'[\mu(t)] = 0$ , 可求得最佳步长(best step-size):

$$\mu(t) = \frac{1}{\|\boldsymbol{\varphi}(t)\|^2}.$$

将上式代入式(34)即得投影算法:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{\|\boldsymbol{\varphi}(t)\|^2} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)].$$

这个算法不可实现,因为信息向量  $\boldsymbol{\varphi}(t)$  包含了未知内部变量  $x(t-1:t-n_b)$ . 为了解决这一问题,用式(29)中  $\hat{\boldsymbol{\varphi}}(t)$  代替上式中未知向量  $\boldsymbol{\varphi}(t)$ , 联立辅助模型(30)–(33), 可得到辨识 IN-FIR 系统参数向量  $\boldsymbol{\vartheta}$  的辅助模型投影算法(Auxiliary Model based Projection algorithm, AM-Proj 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{\|\hat{\boldsymbol{\varphi}}(t)\|^2} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (35)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (36)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (37)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (38)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (39)$$

$$\hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \hat{\boldsymbol{\rho}}(t) \end{bmatrix}. \quad (40)$$

为避免分母为零,可将式(35)修改为

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{1 + \|\hat{\boldsymbol{\varphi}}(t)\|^2} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)]. \quad (41)$$

**注1** 尽管投影算法是通过一个优化准则函数得到的,其步长也是最佳的,但是对于存在干扰的随机系统,投影辨识算法不是一个好的选择. AM-Proj 算法也是如此。

**注2** 辅助模型随机梯度算法收敛速度慢,辅助模型投影算法收敛速度快但对噪声敏感. 为了取二者的优点,我们提出了辅助模型广义投影辨识算法。

### 2.4 辅助模型广义投影辨识算法

在 AM-Proj 算法(35)–(40)中,如果将式(35)右边分母修改为

$$r(t) = \sum_{j=0}^{q-1} \|\hat{\boldsymbol{\varphi}}(t-j)\|^2 =$$

$$r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 - \|\hat{\boldsymbol{\varphi}}(t-q)\|^2, \quad r(0) = 1, \quad (42)$$

就得到辨识 IN-FIR 系统参数向量  $\boldsymbol{\vartheta}$  的辅助模型广义投影算法(Auxiliary Model based Generalized Projection algorithm, AM-GPJ 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (43)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2 - \|\hat{\boldsymbol{\varphi}}(t-q)\|^2, \quad r(0) = 1, \quad (44)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (45)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (46)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (47)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (48)$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (49)$$

这里  $q$  为数据窗长度。

**注3** 值得指出的是:辅助模型投影算法与投影算法性能类似,辅助模型广义投影算法与广义投影算法性能类似,辅助模型随机梯度算法与随机梯度算法性能类似。

**注4** 尽管投影算法能够跟踪时变参数,但是对噪声敏感,参数估计波动(参数估计方差)大. 为了克服这一缺点,笔者发展了投影算法,提出了广义投影算法,它是根据长度为  $q$  的数据窗数据定义  $r(t)$ , 动态调整算法增益,故可以克服对噪声的敏感性. 此外,可以适当选择  $q$  来减小时变参数估计误差上界。

**注5** 虽然投影算法和广义投影算法能够跟踪

时变参数,但是即使对于时不变参数随机系统,投影算法和广义投影算法的参数估计不断波动,参数估计误差不趋于零.下面的随机梯度算法可以解决这一问题.

## 2.5 辅助模型随机梯度辨识算法

在 AM-GPJ 算法(43)—(49)中,将式(44)修改为

$$r(t) = \sum_{j=0}^{t-1} \|\hat{\boldsymbol{\varphi}}(t-j)\|^2 = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2, \quad r(0) = 1, \quad (50)$$

就得到辨识 IN-FIR 系统参数向量  $\boldsymbol{\vartheta}$  的辅助模型随机梯度算法 (Auxiliary Model based Stochastic Gradient algorithm, AM-SG 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (51)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2, \quad (52)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\varphi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (53)$$

$$\boldsymbol{\varphi}(t) = \boldsymbol{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (54)$$

$$\boldsymbol{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (55)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (56)$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (57)$$

AM-SG 算法(51)—(57)的计算步骤如下:

1) 初始化:令  $t=1$ .置初值  $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0, r(0) = 1, x_a(t-i) = 1/p_0, i=1, 2, \dots, n_b$ , 给定基函数  $f_j(\cdot)$ .  $\mathbf{1}_n$  是元均为 1 的  $n$  维列向量,即  $\mathbf{1}_n = [1, 1, \dots, 1]^T \in \mathbf{R}^n, p_0$  取一个很大的正数,如  $p_0 = 10^6$ .

2) 收集数据  $u(t)$  和  $y(t)$ ,用式(54)构造基函数向量  $\boldsymbol{\phi}(t)$ .

3) 用式(55)构造信息向量  $\boldsymbol{\psi}(t)$ ,用式(53)构造信息向量  $\hat{\boldsymbol{\varphi}}(t)$ .

4) 用式(52)计算  $r(t)$ .

5) 根据式(51)刷新参数估计向量  $\hat{\boldsymbol{\vartheta}}(t)$ .

6) 从式(57)的  $\hat{\boldsymbol{\vartheta}}(t)$  中读出  $\hat{\boldsymbol{\theta}}(t)$ ,用式(56)计算辅助模型的输出  $x_a(t)$ .

7)  $t$  增 1,转到第 2)步.

**注 6** AM-SG 算法中步长  $1/r(t)$  的定义式(52)是随机梯度算法的特征.由于算法利用了辅助模型辨识思想,故称为辅助模型随机梯度辨识算法. AM-Proj 算法、AM-GPJ 算法、AM-SG 算法是辅助模型辨

识思想结合梯度搜索在线性参数系统辨识中的应用.

**注 7** AM-SG 算法没有跟踪时变参数的能力,因为随着数据的增加,  $r(t)$  趋于无穷大,算法的增益向量  $\hat{\boldsymbol{\varphi}}(t)/r(t)$  趋于零.但是对于时不变线性参数系统,AM-SG 算法能够给出一致参数估计,但收敛速度极其慢.

**注 8** 为改进 AM-SG 算法的性能,我们可以写出辅助模型修正随机梯度辨识算法、辅助模型遗忘因子随机梯度辨识算法、辅助模型遗忘因子修正随机梯度辨识算法、辅助模型遗忘因子修正广义投影辨识算法等.

**注 9** 如果将多项式  $B(z)$  修改为

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

那么规范化假设应该做相应修正,如令  $b_1 = 1$ .在这种假设下,按照本文方法推导相应的辨识算法是不难的.

## 3 辅助模型多新息梯度辨识方法

本节以输入非线性有限脉冲响应系统为例,讨论基于辅助模型的多新息梯度辨识方法,包括辅助模型多新息投影算法、辅助模型多新息广义投影算法、辅助模型多新息随机梯度算法等.

### 3.1 系统描述与辨识模型

考虑 IN-FIR 系统(14)的辨识问题,其对应的辨识模型(21)—(26)重写如下:

$$x(t) = \boldsymbol{\phi}^T(t) \boldsymbol{\theta}, \quad (58)$$

$$y(t) = x(t) + \boldsymbol{\psi}^T(t) \boldsymbol{\rho} + v(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta} + v(t), \quad (59)$$

其中系统的参数向量为

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{bmatrix} \in \mathbf{R}^n, \quad n := n_b + m, \quad (60)$$

系统的信息向量为

$$\boldsymbol{\varphi}(t) := \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(u(t)) \\ x(t-1:t-n_b) \end{bmatrix} \in \mathbf{R}^n, \quad (61)$$

$$\boldsymbol{\phi}(t) := \boldsymbol{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m, \quad (62)$$

$$\boldsymbol{\psi}(t) := [x(t-1), x(t-2), \dots, x(t-n_b)]^T \in \mathbf{R}^{n_b}, \quad (63)$$

其中  $y(t) \in \mathbf{R}$  为观测输出,  $\boldsymbol{\vartheta} \in \mathbf{R}^n$  为系统待辨识的参数向量,  $v(t) \in \mathbf{R}$  为零均值随机噪声.

### 3.2 辅助模型多新息投影辨识算法

#### 3.2.1 辅助模型多新息投影辨识算法

比较 AM-SG 算法(51)—(57)与 AM-Proj 算法

(35)–(40)的异同,可以发现随机梯度算法与投影算法的差别在于 $r(t)$ 的选择.在AM-MISG算法(104)–(113)中,直接用 $\|\hat{\Phi}(p,t)\|^2$ 代替式(104)中的 $r(t)$ ,或将式(106)修改为 $r(t)=\|\hat{\Phi}(p,t)\|^2$ ,就得到辨识IN-FIR系统(59)参数向量 $\vartheta$ 的辅助模型多新息投影算法(Auxiliary Model based Multi-Innovation Projection algorithm, AM-MI-Proj算法):

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \frac{\hat{\Phi}(p,t)}{\|\hat{\Phi}(p,t)\|^2} E(p,t), \quad (64)$$

$$E(p,t) = Y(p,t) - \hat{\Phi}^T(p,t) \hat{\vartheta}(t-1), \quad (65)$$

$$Y(p,t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (66)$$

$$\hat{\Phi}(p,t) = [\hat{\varphi}(t), \hat{\varphi}(t-1), \dots, \hat{\varphi}(t-p+1)], \quad (67)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \varphi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (68)$$

$$\varphi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (69)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (70)$$

$$x_a(t) = \Phi^T(t) \hat{\theta}(t), \quad (71)$$

$$\hat{\vartheta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (72)$$

从下面的推导可知,这实际上是一个简化的辅助模型多新息投影算法.

### 3.2.2 辅助模型多新息投影辨识算法的推导

下面利用最速下降法或梯度搜索推导辅助模型多新息投影辨识算法(64)–(72).

在IN-FIR系统的辨识模型(58)–(59)中,考虑了从 $j=t-p+1$ 到 $j=t$ 的数据窗里的 $p$ 组数据,定义堆积输出向量 $Y(p,t)$ 和堆积信息矩阵 $\Phi(p,t)$ 如下:

$$Y(p,t) := \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix} \in \mathbf{R}^p, \quad \Phi^T(p,t) := \begin{bmatrix} \varphi^T(t) \\ \varphi^T(t-1) \\ \vdots \\ \varphi^T(t-p+1) \end{bmatrix} \in \mathbf{R}^{p \times n}. \quad (73)$$

定义准则函数为

$$J_6(\vartheta) := \frac{1}{2} \|Y(p,t) - \Phi^T(p,t) \vartheta\|^2,$$

这是一个有限数据窗准则函数.假设步长为 $\mu_i$ ,极小化 $J_6(\vartheta)$ 的梯度递推关系如下:

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) - \mu_i \text{grad}_{\vartheta} [J_6(\hat{\vartheta}(t-1))] =$$

$$\hat{\vartheta}(t-1) + \mu_i \Phi(p,t) [Y(p,t) - \Phi^T(p,t) \hat{\vartheta}(t-1)] = \hat{\vartheta}(t-1) + \mu_i \Phi(p,t) E(p,t), \quad (74)$$

其中新息向量定义为

$$E(p,t) := Y(p,t) - \Phi^T(p,t) \hat{\vartheta}(t-1) \in \mathbf{R}^p. \quad (75)$$

下面求最佳步长 $\mu_i$ .设 $I_n$ 是一个 $n$ 阶单位阵.将 $\vartheta = \hat{\vartheta}(t)$ 代入准则函数 $J_6(\vartheta)$ 中,可得

$$g(\mu_i) := 2J_6(\hat{\vartheta}(t)) = \|Y(p,t) - \Phi^T(p,t) [\hat{\vartheta}(t-1) + \mu_i \Phi(p,t) E(p,t)]\|^2 = \|[I_n - \mu_i \Phi^T(p,t) \Phi(p,t)] E(p,t)\|^2 = E^T(p,t) [I_n - \mu_i \Phi^T(p,t) \Phi(p,t)]^2 E(p,t) = E^T(p,t) [I_n - 2\mu_i \Phi^T(p,t) \Phi(p,t) + \mu_i^2 \Phi^T(p,t) \Phi(p,t) \Phi^T(p,t) \Phi(p,t)] E(p,t) = \|E(p,t)\|^2 - 2\mu_i \|\Phi(p,t) E(p,t)\|^2 + \mu_i^2 \|\Phi^T(p,t) \Phi(p,t) E(p,t)\|^2.$$

极小化 $g(\mu_i)$ ,令 $g'(\mu_i) = 0$ ,即

$$-2 \|\Phi(p,t) E(p,t)\|^2 + 2\mu_i \|\Phi^T(p,t) \Phi(p,t) E(p,t)\|^2 = 0.$$

由此可求得最佳步长为

$$\mu_i = \frac{\|\Phi(p,t) E(p,t)\|^2}{\|\Phi^T(p,t) \Phi(p,t) E(p,t)\|^2} = \frac{E^T(p,t) \Phi^T(p,t) \Phi(p,t) E(p,t)}{E^T(p,t) \Phi^T(p,t) \Phi(p,t) \Phi^T(p,t) \Phi(p,t) E(p,t)}. \quad (76)$$

式(74), (76), (75)和(73)构成了多新息投影算法:

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \mu_i \Phi(p,t) E(p,t), \quad (77)$$

$$\mu_i = \frac{\|\Phi(p,t) E(p,t)\|^2}{\|\Phi^T(p,t) \Phi(p,t) E(p,t)\|^2}, \quad (78)$$

$$E(p,t) = Y(p,t) - \Phi^T(p,t) \hat{\vartheta}(t-1), \quad (79)$$

$$Y(p,t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (80)$$

$$\Phi(p,t) = [\varphi(t), \varphi(t-1), \dots, \varphi(t-p+1)]. \quad (81)$$

然而,这个算法无法实现,因为算法中信息向量 $\varphi(t)$ 包含了未知中间变量 $x(t)$ .解决的方法是利用辅助模型辨识思想,构造辅助模型(30)–(33),用 $\varphi(t)$ 的估计 $\hat{\varphi}(t)$ 构造 $\Phi(p,t)$ 的估计:

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

用 $\hat{\Phi}(p,t)$ 代替式(77)–(81)中的 $\Phi(p,t)$ ,联立辅助模型(30)–(33),就得到辨识IN-FIR系统(59)参数向量 $\vartheta$ 的辅助模型多新息投影算法(AM-MI-Proj算法):

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \mu_i \hat{\Phi}(p,t) E(p,t), \quad \hat{\vartheta}(0) = \mathbf{1}_n/p_0, \quad (82)$$

$$\mu_i = \frac{\|\hat{\Phi}(p,t) E(p,t)\|^2}{\|\hat{\Phi}^T(p,t) \hat{\Phi}(p,t) E(p,t)\|^2}, \quad (83)$$

$$E(p,t) = Y(p,t) - \hat{\Phi}^T(p,t) \hat{\vartheta}(t-1), \quad (84)$$

$$Y(p,t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (85)$$

$$\hat{\Phi}(p, t) = [\hat{\varphi}(t), \hat{\varphi}(t-1), \dots, \hat{\varphi}(t-p+1)], \quad (86)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \psi(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (87)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (88)$$

$$\psi(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (89)$$

$$x_a(t) = \Phi^T(t) \hat{\theta}(t), \quad (90)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (91)$$

如果式(83)的分母为零, 就令  $\hat{\theta}(t) = \hat{\theta}(t-1)$ , 或在式(83)的分母上加 1. 因为收敛因子  $\mu_i$  的计算式比较复杂, 故有必要对其进行简化.

由于对于任意实向量  $x$  和非负定对称矩阵  $Q$ ,  $\lambda_{\max}[Q]$  是  $Q$  的最大特征值, 利用关系  $x^T Q x \leq \lambda_{\max}[Q] x^T x \leq \|Q\| x^T x$ , 有

$$\mu_i = \frac{\|\hat{\Phi}(p, t) E(p, t)\|^2}{\|\hat{\Phi}^T(p, t) \hat{\Phi}(p, t) E(p, t)\|^2} \geq \frac{E^T(p, t) \hat{\Phi}^T(p, t) \hat{\Phi}(p, t) E(p, t)}{\lambda_{\max}[\hat{\Phi}(p, t) \hat{\Phi}^T(p, t)] E^T(p, t) \hat{\Phi}^T(p, t) \hat{\Phi}(p, t) E(p, t)} = \frac{1}{\lambda_{\max}[\hat{\Phi}(p, t) \hat{\Phi}^T(p, t)] \|\hat{\Phi}(p, t)\|^2},$$

于是, 收敛因子 (convergence factor) 可以保守取为

$$\mu_i = \frac{1}{\lambda_{\max}[\hat{\Phi}(p, t) \hat{\Phi}^T(p, t)]}, \quad (92)$$

因为计算矩阵的迹 (trace) 比计算特征值 (eigenvalue) 简单, 所以收敛因子可以更保守取为

$$\mu_i = \frac{1}{\|\hat{\Phi}(p, t)\|^2}. \quad (93)$$

如果采用式(92)的收敛因子  $\mu_i$ , 则式(82)—(83)可以简化为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\hat{\Phi}(p, t)}{\lambda_{\max}[\hat{\Phi}(p, t) \hat{\Phi}^T(p, t)]} E(p, t).$$

如果在算法(82)—(91)中取式(93)的  $\mu_i$  作为收敛因子, 就得到简化辅助模型多新息投影算法(64)—(72).

此外, 由式(82)和(84)可得

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu_i \hat{\Phi}(p, t) [Y(p, t) - \hat{\Phi}^T(p, t) \hat{\theta}(t-1)] = [I_n - \mu_i \hat{\Phi}(p, t) \hat{\Phi}^T(p, t)] \hat{\theta}(t-1) + \mu_i \hat{\Phi}(p, t) Y(p, t).$$

如果  $\mu_i \hat{\Phi}(p, t) \hat{\Phi}^T(p, t)$  的特征值大于 2,  $\hat{\theta}(t)$  就不可能收敛. 因此, 为保证辅助模型多新息算法的收敛性, 收敛因子  $\mu_i$  的选择必须使得  $[I_n - \mu_i \hat{\Phi}(p, t) \cdot$

$\hat{\Phi}^T(p, t)]$  的特征值在单位圆内. 上述收敛因子的选择满足这一条件.

**注 10** 值得指出的是: 在我们提出了变递推间隔的辨识方法后 (实际上为变递推步长更合理), 显然原来梯度搜索算法中的“步长”术语命名不甚合理. 因此, 在笔者的有关论文和著作中, 多将“步长”称为“收敛因子”.

### 3.3 辅助模型多新息广义投影算法

辅助模型广义投影算法(43)—(49)是将辅助模型投影算法(35)—(40)中的  $r(t) = \|\hat{\varphi}(t)\|^2$  扩展为长度为  $q$  的数据窗里的信息向量的平方和:

$$r(t) = \sum_{j=0}^{q-1} \|\hat{\varphi}(t-j)\|^2 = r(t-1) + \|\hat{\varphi}(t)\|^2 - \|\hat{\varphi}(t-q)\|^2.$$

基于这种思想, 可以将辅助模型多新息投影算法(64)—(72)中的  $r(t) = \|\hat{\Phi}(p, t)\|^2$  扩展为  $r(t) = \|\hat{\Phi}(q, t)\|^2$ , 就得到辨识 IN-FIR 系统(59)参数向量  $\theta$  的辅助模型多新息广义投影算法 (Auxiliary Model based Multi-Innovation Generalized Projection algorithm, AM-MI-GPJ 算法):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\hat{\Phi}(p, t)}{r(t)} E(p, t), \quad \hat{\theta}(0) = \mathbf{1}_n / p_0, \quad (94)$$

$$E(p, t) = Y(p, t) - \hat{\Phi}^T(p, t) \hat{\theta}(t-1), \quad (95)$$

$$r(t) = \|\hat{\Phi}(q, t)\|^2 = r(t-1) + \|\hat{\varphi}(t)\|^2 - \|\hat{\varphi}(t-q)\|^2, \quad q \geq p, \quad (96)$$

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (97)$$

$$\hat{\Phi}(p, t) = [\hat{\varphi}(t), \hat{\varphi}(t-1), \dots, \hat{\varphi}(t-p+1)], \quad (98)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \psi(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (99)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (100)$$

$$\psi(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (101)$$

$$x_a(t) = \Phi^T(t) \hat{\theta}(t), \quad (102)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (103)$$

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

### 3.4 辅助模型多新息随机梯度辨识算法

下面讨论线性参数系统(21)—(22)的辅助模型多新息随机梯度辨识方法. 设整数  $p \geq 1$  表示新息长度. 根据多新息辨识理论<sup>[43]</sup>, 将 AM-SG 算法

(51)–(57)中的标量新息

$$e(t) := y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1) \in \mathbf{R}$$

扩展为新息向量

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

将系统输出  $y(t)$  和信息向量  $\hat{\boldsymbol{\varphi}}(t)$  扩展为堆积输出向量  $\mathbf{Y}(p, t)$  和堆积信息矩阵  $\hat{\boldsymbol{\Phi}}(p, t)$  如下:

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

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

便得到辨识 IN-FIR 系统参数向量  $\boldsymbol{\vartheta}$  的辅助模型多新息随机梯度算法 (Auxiliary Model based Multi-Innovation Stochastic Gradient algorithm, AM-MISG 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p, t)}{r(t)} \mathbf{E}(p, t), \quad (104)$$

$$\mathbf{E}(p, t) = \mathbf{Y}(p, t) - \hat{\boldsymbol{\Phi}}^T(p, t) \hat{\boldsymbol{\vartheta}}(t-1), \quad (105)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(p, t)\|^2, \quad (106)$$

$$\mathbf{Y}(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (107)$$

$$\hat{\boldsymbol{\Phi}}(p, t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \dots, \hat{\boldsymbol{\varphi}}(t-p+1)], \quad (108)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (109)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (110)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (111)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (112)$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (113)$$

当新息长度  $p=1$  时, AM-MISG 算法退化为 AM-SG 算法 (51)–(57), 此即 AM-SG 算法是 AM-MISG 算法的特例. 为提高收敛速度, 式 (106) 中的  $r(t)$  也可修改为

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

AM-MISG 算法 (104)–(113) 的实现步骤如下:

1) 初始化: 令  $t=1$ , 给定数据长度  $L_e$ , 新息长度  $p$  和基函数  $f_j(\cdot)$ ; 置初值  $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0$ ,  $r(0) = 1$ ,  $x_a(t-i) = 1/p_0, i=1, 2, \dots, n_b$ .

2) 采集输入输出数据  $u(t)$  和  $y(t)$ , 用式 (107) 构造堆积输出向量  $\mathbf{Y}(p, t)$ .

3) 用式 (110) 构造基函数向量  $\boldsymbol{\phi}(t)$ , 用式

(111) 构造信息向量  $\hat{\boldsymbol{\psi}}(t)$ , 用式 (109) 构造信息向量  $\hat{\boldsymbol{\varphi}}(t)$ , 用式 (108) 构造堆积信息矩阵  $\hat{\boldsymbol{\Phi}}(p, t)$ .

4) 用式 (105) 计算新息向量  $\mathbf{E}(p, t)$ , 用式 (106) 计算  $r(t)$ .

5) 根据式 (104) 刷新参数估计向量  $\hat{\boldsymbol{\vartheta}}(t)$ .

6) 从式 (113) 的  $\hat{\boldsymbol{\vartheta}}(t)$  中读出  $\hat{\boldsymbol{\theta}}(t)$ , 用式 (112) 计算辅助模型的输出  $x_a(t)$ .

7) 如果  $t < L_e$ ,  $t$  增加 1, 转到第 2) 步; 否则获得参数估计  $\hat{\boldsymbol{\vartheta}}(L_e)$ .

AM-MISG 算法 (104)–(113) 计算 IN-FIR 系统参数估计向量  $\hat{\boldsymbol{\vartheta}}(t)$  的流程如图 3 所示.

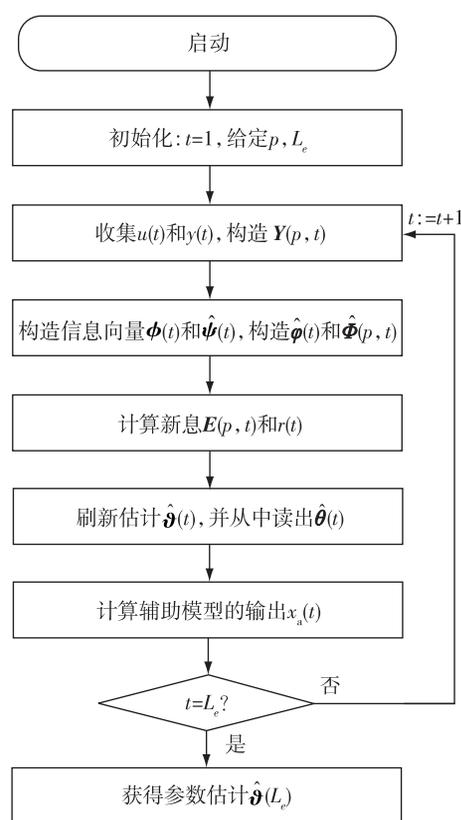


图 3 基于辅助模型的多新息随机梯度算法流程

Fig. 3 The flowchart of the AM-MISG algorithm

**注 11** 与 AM-SG 算法 (51)–(57) 相比, AM-MISG 算法 (104)–(113) 通过新息扩展, 充分地利用了系统的观测数据和辨识新息, 故收敛速度加快, 辨识精度提高, 不过计算量有所增大.

### 3.5 辅助模型多新息遗忘梯度算法

在辅助模型多新息随机梯度算法 (104)–(113) 中引入遗忘因子  $\lambda$ , 将式 (106) 修改为

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

或

$r(t) = \lambda r(t-1) + \|\hat{\varphi}(t)\|^2$ ,  $0 \leq \lambda \leq 1$ ,  $r(0) = 1$ , 就得到辨识 IN-FIR 系统(59) 参数向量  $\boldsymbol{\vartheta}$  的遗忘因子辅助模型多新息随机梯度算法 (Forgetting Factor AM-MISG algorithm, FF-AM-MISG 算法) 或称为辅助模型遗忘因子多新息随机梯度算法 (AM-FF-MISG 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)} E(p,t), \quad \hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0, \quad (114)$$

$$E(p,t) = Y(p,t) - \hat{\boldsymbol{\Phi}}^T(p,t) \hat{\boldsymbol{\vartheta}}(t-1), \quad (115)$$

$$r(t) = \lambda r(t-1) + \|\hat{\varphi}(t)\|^2, \quad 0 \leq \lambda \leq 1, \quad r(0) = 1, \quad (116)$$

$$Y(p,t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (117)$$

$$\hat{\boldsymbol{\Phi}}(p,t) = [\hat{\varphi}(t), \hat{\varphi}(t-1), \dots, \hat{\varphi}(t-p+1)], \quad (118)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (119)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (120)$$

$$\boldsymbol{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (121)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad (122)$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (123)$$

辅助模型多新息遗忘因子算法具有跟踪时变参数的能力.

**注 12** 辅助模型投影算法(35)—(40) 对噪声很敏感, 辅助模型广义投影算法(43)—(49) 和辅助模型多新息投影算法(64)—(72) 可以在一定程度上降低对噪声的敏感性; 辅助模型多新息投影算法(64)—(72) 和辅助模型多新息广义投影算法(94)—(103) 可以进一步降低对噪声的敏感性.

**注 13** 对于时变参数系统, 辅助模型广义投影算法(43)—(49)、辅助模型遗忘因子随机梯度算法、辅助模型多新息广义投影算法(94)—(103)、辅助模型遗忘因子多新息随机梯度算法(114)—(123) 可以通过调节数据窗长度  $q$  或遗忘因子来限制参数估计误差上界.

#### 4 变递推间隔辅助模型多新息梯度辨识方法

本节以输入非线性有限脉冲响应系统为例, 考虑稀少量测数据情形和损失数据情形, 研究变递推间隔的辅助模型多新息梯度辨识方法: 变递推间隔辅助模型多新息投影算法、变递推间隔辅助模型多新息广义投影算法、变递推间隔辅助模型多新息随机梯度算法等.

#### 4.1 系统描述与辨识模型

考虑 IN-FIR 系统(14) 的辨识问题, 其对应的辨识模型(58)—(63), 重写如下:

$$x(t) = \boldsymbol{\phi}^T(t) \boldsymbol{\theta}, \quad (124)$$

$$y(t) = x(t) + \boldsymbol{\psi}^T(t) \boldsymbol{\rho} + v(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta} + v(t), \quad (125)$$

其中系统的参数向量为

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{bmatrix} \in \mathbf{R}^n, \quad n := n_b + m, \quad (126)$$

系统的信息向量为

$$\boldsymbol{\varphi}(t) := \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x(t-1:t-n_b) \end{bmatrix} \in \mathbf{R}^n, \quad (127)$$

$$\boldsymbol{\phi}(t) := \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m, \quad (128)$$

$$\boldsymbol{\psi}(t) := [x(t-1), x(t-2), \dots, x(t-n_b)]^T \in \mathbf{R}^{n_b}, \quad (129)$$

$y(t) \in \mathbf{R}$  为观测输出,  $\boldsymbol{\vartheta} \in \mathbf{R}^n$  为系统待辨识的参数向量,  $v(t) \in \mathbf{R}$  为零均值随机噪声.

#### 4.2 稀少量测数据系统与损失数据系统

在实际中, 由于传感器故障或硬件设备限制, 会造成某些采样点的观测数据丢失, 这样的系统称为损失数据系统或稀少量测数据系统.

也就是说, 对每一个  $t = 0, 1, 2, \dots$ , 观测数据  $y(t)$  不可能总是可得到的系统称为损失数据系统. 为了处理数据丢失情况, 定义一个整数序列 (integer sequence)  $\{t_s : s = 0, 1, 2, \dots\}$  满足

$$0 = t_0 < t_1 < t_2 < t_3 < \dots < t_{s-1} < t_s < \dots, \quad (130)$$

且  $t_s^* := t_s - t_{s-1} \geq 1$ . 假设当  $(s = 0, 1, 2, \dots)$  时,  $y(t)$  都可得到, 即对所有  $s = 0, 1, 2, \dots$ , 观测  $y(t_s)$  都可得到, 也就是  $y(t_0), y(t_1), y(t_2), \dots$  都可得到. 在这种情况下, 数据集  $\{y(t_s) : s = 0, 1, 2, \dots\}$  包含所有可得到的观测输出数据, 而不可得到的输出数据  $\{y(t_s+1), y(t_s+2), \dots, y(t_{s+1}-1) : s = 0, 1, 2, \dots\}$  是损失数据.

损失数据系统有两种类型: 一种是稀少量测数据系统, 一种是损失数据系统. 与可得到的数据量相比, 当丢失的数据占大部分, 就称为稀少量测数据系统, 如图 4 所示 (图中 29 个数据点中只有 9 个数据可得到); 当丢失的数据占小部分, 就称为损失数据系统, 如图 5 所示 (图中 29 个数据点中有 25 个数据可得到, 只有 4 个数据不可得到)<sup>[18, 20]</sup>.

对于图 4 的稀少量测数据情形, 可得到的观测输出为  $y(0), y(1), y(3), y(6), y(10), y(15), y(16), y(21), y(28)$ , 即  $y(t_0), y(t_1), y(t_2), y(t_3), y(t_4), y(t_5), y(t_6), y(t_7), y(t_8)$ , 这里  $t_0 = 0, t_1 = 1, t_2 = 3, t_3 = 6, t_4 = 10, t_5 = 15, t_6 = 16, t_7 = 21, t_8 = 28$ , 这是

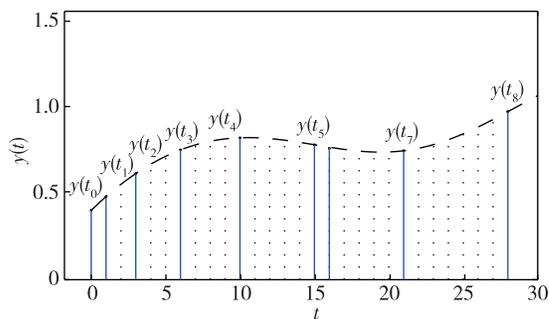


图4 稀少量测输出数据情形( $y(0), y(1), y(3), y(6), y(10), y(15), y(16), y(21), y(28)$ 是可得到数据)

Fig. 4 A scarce measurement pattern

稀少输出量测可得到的一般框架.当然包括对所有  $s, t_s^* = 1$  时所有观测输出可得到情形.

图5是损失输出数据情形(稀少损失数据),+代表损失输出数据或坏数据(异常数据或不可信数据), $y(3), y(8), y(9), y(23)$ 是损失数据, $y(15)$ 是不可信数据.

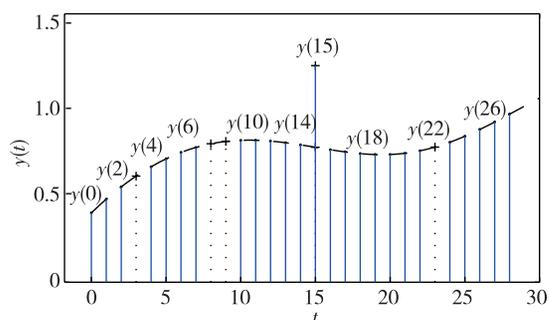


图5 损失输出数据情形(+:损失数据或不可信数据; $y(3), y(8), y(9), y(23)$ 是损失数据; $y(15)$ 是不可信数据)

Fig. 5 A missing output data pattern

比较图4和图5可知,稀少量测数据系统意味着大部分数据丢失了,只有少部分数据可得到;相反,损失数据系统意味着少部分数据丢失了,大部分数据可得到.如果系统的输入数据在每一个时刻都可得到,当  $t_s^* = d$  为大于1的常数时, $y(dt)$ 可得到,就可以得到双率采样数据系统,且输入输出采样数据比为  $d^{[5-6]}$ .

值得指出的是,对于系统的输出数据丢失情形,有的文献用估计的模型来计算损失输出,进而用这个估计输出来辨识系统参数,这种方法是徒劳的,是不可能改进参数估计精度的,应该直接采用跳过这些损失输出数据点的变递推间隔辨识方法.本节讨论了能克服存在损失数据问题的变递推间隔辅助模

型多新息梯度辨识方法.

### 4.3 变递推间隔辅助模型多新息投影算法

#### 4.3.1 稀少量测数据系统与损失数据系统的辨识模型

在实际问题中,可能发生数据丢失的情况,包括输出数据丢失、输入数据丢失、状态数据丢失等<sup>[18-20]</sup>.为了使辨识算法跳过这些丢失的数据点,需要根据系统的可得到的采样数据,定义一个整数序列  $\{t_s : s=0, 1, 2, \dots\}$ .

对于稀少量测数据系统,假设观测数据  $y(t_s)$  可得到,  $p \geq 1$  为新息长度,定义堆积输出向量  $\mathbf{Y}(p, t_s)$  和堆积信息矩阵  $\Phi(p, t_s)$  如下:

$$\mathbf{Y}(p, t_s) := \begin{bmatrix} y(t_s) \\ y(t_{s-1}) \\ y(t_{s-2}) \\ \vdots \\ y(t_{s-p+1}) \end{bmatrix} \in \mathbf{R}^p,$$

$$\Phi^T(p, t_s) := \begin{bmatrix} \varphi^T(t_s) \\ \varphi^T(t_{s-1}) \\ \varphi^T(t_{s-2}) \\ \vdots \\ \varphi^T(t_{s-p+1}) \end{bmatrix} \in \mathbf{R}^{p \times n}. \quad (131)$$

如果再定义堆积噪声向量(stacked noise vector)  $\mathbf{V}(p, t_s) := [v(t_s), v(t_{s-1}), v(t_{s-2}), \dots, v(t_{s-p+1})]^T \in \mathbf{R}^p$ , 则由式(125)可得到一类稀少量测数据系统的辅助模型多新息辨识模型:

$$\mathbf{Y}(p, t_s) = \Phi^T(p, t_s) \boldsymbol{\theta} + \mathbf{V}(p, t_s). \quad (132)$$

对于损失数据系统,假设  $y(t_s), y(t_s-1), \dots, y(t_s-p+1)$  都可得到,定义堆积输出向量  $\mathbf{Y}(p, t_s)$  和堆积信息矩阵  $\Phi(p, t_s)$  如下:

$$\mathbf{Y}(p, t_s) := \begin{bmatrix} y(t_s) \\ y(t_s-1) \\ y(t_s-2) \\ \vdots \\ y(t_s-p+1) \end{bmatrix} \in \mathbf{R}^p,$$

$$\Phi^T(p, t_s) := \begin{bmatrix} \varphi^T(t_s) \\ \varphi^T(t_s-1) \\ \varphi^T(t_s-2) \\ \vdots \\ \varphi^T(t_s-p+1) \end{bmatrix} \in \mathbf{R}^{p \times n}. \quad (133)$$

如果再定义堆积噪声向量

$V(p, t_s) := [v(t_s), v(t_s-1), v(t_s-2), \dots, v(t_s-p+1)]^T \in \mathbf{R}^p$ , 则由式(125)可得到一类损失数据系统的辅助模型多新息辨识模型:

$$Y(p, t_s) = \Phi^T(p, t_s) \vartheta + V(p, t_s). \quad (134)$$

**注 14** 从式(133)与(131)可以看出: 损失数据系统的  $Y(p, t_s)$  和  $\Phi(p, t_s)$  只使用了最近的直到时刻  $t=t_s$  的  $p$  组数据  $\{y(t_s-j), \varphi(t_s-j), j=0, 1, \dots, p-1\}$ ; 稀少量测数据系统的  $Y(p, t_s)$  和  $\Phi(p, t_s)$  使用了从  $t=t_{s-p+1}$  到  $t_s$  的  $t_s-t_{s-p+1}+1$  组数据中的  $p$  组可得到的数据  $\{y(t_{s-j}), \varphi(t_{s-j}), j=0, 1, \dots, p-1\}$ .

**注 15** 对于稀少量测数据系统和损失数据系统, 不管堆积输出向量  $Y(p, t_s)$  和堆积信息矩阵  $\Phi(p, t_s)$  内部是怎么构成的, 它们的辨识模型(132)与(134)的形式是一样的, 它们的辅助模型多新息辨识方法的推导过程是一样的.

在这种定义下, 当系统输入是由计算机产生的数字信号, 是可测的, 当系统输出只有稀少采样数据或系统输出存在数据损失时, 本节的变递推间隔辅助模型多新息辨识方法在一定程度上能处理这类损失数据系统的辨识问题.

#### 4.3.2 稀少量测数据系统的变递推间隔辅助模型多新息投影算法

根据稀少量测数据系统的辨识模型(132), 定义准则函数

$$J_7(\vartheta) := \frac{1}{2} \| Y(p, t_s) - \Phi^T(p, t_s) \vartheta \|^2.$$

这是一个有限数据窗准则函数. 假设步长为  $\mu(t_s)$ ,

$$\hat{\vartheta}(t_s) := \begin{bmatrix} \hat{\theta}(t_s) \\ \hat{\rho}(t_s) \end{bmatrix} \in \mathbf{R}^n \text{ 是 } \vartheta = \begin{bmatrix} \theta \\ \rho \end{bmatrix} \text{ 在时刻 } t=t_s \text{ 的估计.}$$

极小化  $J_7(\vartheta)$ , 可以得到梯度递推关系:

$$\hat{\vartheta}(t_s) = \hat{\vartheta}(t_{s-1}) - \mu(t_s) \text{grad}[J_7(\hat{\vartheta}(t_{s-1}))] = \hat{\vartheta}(t_{s-1}) + \mu(t_s) \Phi(p, t_s) [Y(p, t_s) - \Phi^T(p, t_s) \hat{\vartheta}(t_{s-1})].$$

定义新息向量

$$E(p, t_s) := Y(p, t_s) - \Phi^T(p, t_s) \hat{\vartheta}(t_{s-1}) \in \mathbf{R}^p, \quad (135)$$

则有

$$\hat{\vartheta}(t_s) = \hat{\vartheta}(t_{s-1}) + \mu(t_s) \Phi(p, t_s) E(p, t_s). \quad (136)$$

下面求最佳步长  $\mu(t_s)$ . 将  $\vartheta = \hat{\vartheta}(t_s)$  代入准则函数  $J_7(\vartheta)$  中可得

$$g(\mu(t_s)) := 2J_7(\vartheta) =$$

$$\begin{aligned} & \| Y(p, t_s) - \Phi^T(p, t_s) [\hat{\vartheta}(t_{s-1}) + \mu(t_s) \Phi(p, t_s) E(p, t_s)] \|^2 = \\ & \| [I_n - \mu(t_s) \Phi^T(p, t_s) \Phi(p, t_s)] E(p, t_s) \|^2 = \\ & E^T(p, t_s) [I_n - \mu(t_s) \Phi^T(p, t_s) \Phi(p, t_s)]^2 E(p, t_s) = \end{aligned}$$

$$\begin{aligned} & E^T(p, t_s) [I_n - 2\mu(t_s) \Phi^T(p, t_s) \Phi(p, t_s) + \\ & \mu^2(t_s) \Phi^T(p, t_s) \Phi(p, t_s) \Phi^T(p, t_s) \Phi(p, t_s)] E(p, t_s) = \\ & \| E(p, t_s) \|^2 - 2\mu(t_s) \| \Phi(p, t_s) E(p, t_s) \|^2 + \\ & \mu^2(t_s) \| \Phi^T(p, t_s) \Phi(p, t_s) E(p, t_s) \|^2. \end{aligned}$$

极小化  $g(\mu(t_s))$ , 可求得最佳步长为

$$\mu(t_s) = \frac{\| \Phi(p, t_s) E(p, t_s) \|^2}{\| \Phi^T(p, t_s) \Phi(p, t_s) E(p, t_s) \|^2} = \frac{E^T(p, t_s) \Phi^T(p, t_s) \Phi(p, t_s) E(p, t_s)}{E^T(p, t_s) \Phi^T(p, t_s) \Phi(p, t_s) \Phi^T(p, t_s) \Phi(p, t_s) E(p, t_s)}.$$

为简化计算, 步长可简单地保守取为

$$\mu(t_s) = \frac{1}{\| \Phi(p, t_s) \|^2}. \quad (137)$$

式(136), (137), (135)和(131)构成了多新息投影算法:

$$\hat{\vartheta}(t_s) = \hat{\vartheta}(t_{s-1}) + \mu(t_s) \Phi(p, t_s) E(p, t_s), \quad (138)$$

$$\mu(t_s) = \frac{1}{\| \Phi(p, t_s) \|^2}, \quad (139)$$

$$E(p, t_s) = Y(p, t_s) - \Phi^T(p, t_s) \hat{\vartheta}(t_{s-1}), \quad (140)$$

$$Y(p, t_s) = [y(t_s), y(t_{s-1}), y(t_{s-2}), \dots, y(t_{s-p+1})]^T, \quad (141)$$

$$\Phi(p, t_s) = [\varphi(t_s), \varphi(t_{s-1}), \varphi(t_{s-2}), \dots, \varphi(t_{s-p+1})], \quad (142)$$

$$0 = t_0 < t_1 < t_2 < \dots, \quad 1 \leq t_s^* := t_s - t_{s-1}, \quad (143)$$

然而, 这个算法无法实现, 因为算法中信息向量  $\varphi(t_s)$  包含了未知中间变量  $x(t_s-i)$ . 解决的方法是利用辅助模型辨识思想, 构造辅助模型(30)–(33), 用  $\varphi(t)$  的估计  $\hat{\varphi}(t)$  构造  $\Phi(p, t_s)$  的估计:

$$\hat{\Phi}(p, t_s) := [\hat{\varphi}(t_s), \hat{\varphi}(t_{s-1}), \dots, \hat{\varphi}(t_{s-p+1})] \in \mathbf{R}^{n \times p}.$$

用  $\hat{\Phi}(p, t_s)$  代替式(138)–(143)中的  $\Phi(p, t_s)$ , 联立辅助模型(30)–(33), 就得到辨识稀少量测数据 IN-FIR 系统(125)参数向量  $\vartheta$  的变递推间隔辅助模型多新息投影算法 (interval-Varying AM-MI-Projection algorithm, V-AM-MI-Proj 算法):

$$\hat{\vartheta}(t_s) = \hat{\vartheta}(t_{s-1}) + \mu(t_s) \hat{\Phi}(p, t_s) E(p, t_s),$$

$$s = 1, 2, 3, \dots, \quad \hat{\vartheta}(t_0) = \mathbf{1}_n / p_0, \quad (144)$$

$$\hat{\vartheta}(t) = \hat{\vartheta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (145)$$

$$\mu(t_s) = \frac{1}{\| \hat{\Phi}(p, t_s) \|^2}, \quad (146)$$

$$E(p, t_s) = Y(p, t_s) - \hat{\Phi}^T(p, t_s) \hat{\vartheta}(t_{s-1}), \quad (147)$$

$$Y(p, t_s) = [y(t_s), y(t_{s-1}), y(t_{s-2}), \dots, y(t_{s-p+1})]^T, \quad (148)$$

$$\hat{\Phi}(p, t_s) = [\hat{\varphi}(t_s), \hat{\varphi}(t_{s-1}), \hat{\varphi}(t_{s-2}), \dots, \hat{\varphi}(t_{s-p+1})], \quad (149)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \hat{\phi}(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1 : t-n_b) \end{bmatrix}, \quad (150)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (151)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (152)$$

$$x_a(t) = \Phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (153)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (154)$$

**注 16** V-AM-MI-Proj 算法在可得到的采样数据点  $t=t_s$  刷新参数估计  $\hat{\theta}(t_s)$ , 而在采样区间  $T_s = \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}$  保持估计不变, 式(145)表示在数据丢失的区间内参数估计保持不变. 该算法是依变量  $s$  进行递推计算的,  $t$  的取值是  $\{0, 1, 2, \dots\}$ .

### 4.3.3 损失数据系统的变递推间隔辅助模型多新息投影算法

根据损失数据系统的辨识模型(134), 定义准则函数

$$J_s(\theta) := \frac{1}{2} \| Y(p, t_s) - \Phi^T(p, t_s)\theta \|^2.$$

采用与算法(144)~(154)类似的推导, 便可以得到辨识损失数据 IN-FIR 系统(125)参数向量  $\theta$  的变递推间隔辅助模型多新息投影算法(V-AM-MI-Proj 算法):

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \mu(t_s)\hat{\Phi}(p, t_s)E(p, t_s), \quad s=1, 2, 3, \dots, \quad \hat{\theta}(t_0) = \mathbf{1}_n/p_0, \quad (155)$$

$$\hat{\theta}(t) = \hat{\theta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (156)$$

$$\mu(t_s) = \frac{1}{\|\hat{\Phi}(p, t_s)\|^2}, \quad (157)$$

$$E(p, t_s) = Y(p, t_s) - \hat{\Phi}^T(p, t_s)\hat{\theta}(t_{s-1}), \quad (158)$$

$$Y(p, t_s) = [y(t_s), y(t_s-1), \dots, y(t_s-p+1)]^T, \quad (159)$$

$$\hat{\Phi}(p, t_s) = [\hat{\varphi}(t_s), \hat{\varphi}(t_s-1), \dots, \hat{\varphi}(t_s-p+1)], \quad (160)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (161)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (162)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (163)$$

$$x_a(t) = \Phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0, 1, \dots, t_1-1, \quad (164)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (165)$$

**注 17** 为了防止式(146)和(157)右边第二项分母为零, 解决的方法之一是在分子加上 1, 即

$$\mu(t_s) = \frac{1}{1 + \|\hat{\Phi}(p, t_s)\|^2}, \quad (166)$$

或将式(144)和(155)修改为

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\Phi}(p, t_s)}{1 + \|\hat{\Phi}(p, t_s)\|^2} E(p, t_s), \quad s=1, 2, 3, \dots, \quad \hat{\theta}(t_0) = \mathbf{1}_n/p_0. \quad (167)$$

**注 18** 变递推间隔辅助模型多新息辨识算法可以这样理解: 实际中一些系统的控制周期与参数估计刷新周期不相同. 在这种情形下,  $t_s^* := t_s - t_{s-1}$  可认为是动态长度的控制周期, 但是在一个周期  $t_s^*$  内的每一时刻都可获得一组可用的数据信息, 可利用其中  $p$  组数据(或直到时刻  $t_s$  的  $p$  组数据)所包含的刷新参数估计, 即用这些信息产生的新息向量  $E(p, t_s)$  来对  $\hat{\theta}(t_{s-1})$  进行修正、更新参数估计, 多新息(修正)辨识算法就是因此而得名. 这里  $E(p, t_s)$  的每一元都是新息(第 1 元是当前新息, 其余为历史新息), 与新息的定义吻合, 易于理解. 当  $t_s^* = p = 1$  时, 变递推间隔辅助模型多新息算法就退化为辅助模型多新息投影辨识算法.

### 4.4 变递推间隔辅助模型多新息广义投影算法

#### 4.4.1 稀少量测数据系统的变递推间隔辅助模型多新息广义投影算法

对稀少量测数据系统的 V-AM-MI-Proj 算法(144)~(154)中的收敛因子进行调整, 令  $r(t_s) := 1/\mu(t_s)$ , 用长度  $q$  的数据窗数据定义

$$r(t_s) := \|\hat{\Phi}(q, t_s)\|^2 = r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2 - \|\hat{\varphi}(t_{s-q})\|^2,$$

就得到辨识稀少量测数据 IN-FIR 系统(125)参数向量  $\theta$  的变递推间隔辅助模型多新息广义投影算法(interval-Varying AM-MI-GPJ algorithm, V-AM-MI-GPJ 算法):

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\Phi}(p, t_s)}{r(t_s)} E(p, t_s), \quad s=1, 2, 3, \dots, \quad \hat{\theta}(t_0) = \mathbf{1}_n/p_0, \quad (168)$$

$$\hat{\theta}(t) = \hat{\theta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (169)$$

$$r(t_s) = \|\hat{\Phi}(q, t_s)\|^2 = r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2 - \|\hat{\varphi}(t_{s-q})\|^2, \quad r(t_0) = 1, \quad (170)$$

$$E(p, t_s) = Y(p, t_s) - \hat{\Phi}^T(p, t_s)\hat{\theta}(t_{s-1}), \quad (171)$$

$$Y(p, t_s) = [y(t_s), y(t_{s-1}), y(t_{s-2}), \dots, y(t_{s-p+1})]^T, \quad (172)$$

$$\hat{\Phi}(p, t_s) = [\hat{\varphi}(t_s), \hat{\varphi}(t_{s-1}), \hat{\varphi}(t_{s-2}), \dots, \hat{\varphi}(t_{s-p+1})], \quad (173)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (174)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (175)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (176)$$

$$x_a(t) = \Phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (177)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (178)$$

#### 4.4.2 损失数据系统的变递推间隔辅助模型多新息广义投影算法

对损失数据系统的 V-AM-MI-Proj 算法(155)—(156)中的收敛因子进行调整,令  $r(t_s) := 1/\mu(t_s)$ , 用长度  $q$  的数据窗数据定义

$$r(t_s) := r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2 - \|\hat{\varphi}(t_{s-q})\|^2,$$

就得到辨识损失数据 IN-FIR 系统(125)参数向量  $\theta$  的变递推间隔辅助模型多新息广义投影算法(interval-Varying AM-MI-GPJ algorithm, V-AM-MI-GPJ 算法):

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\Phi}(p, t_s)}{r(t_s)} E(p, t_s), \quad s=1,2,3,\dots,$$

$$\hat{\theta}(t_0) = \mathbf{1}_n/p_0, \quad (179)$$

$$\hat{\theta}(t) = \hat{\theta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (180)$$

$$r(t_s) = r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2 - \|\hat{\varphi}(t_{s-q})\|^2, \quad r(t_0) = 1, \quad (181)$$

$$E(p, t_s) = Y(p, t_s) - \hat{\Phi}^T(p, t_s) \hat{\theta}(t_{s-1}), \quad (182)$$

$$Y(p, t_s) = [y(t_s), y(t_s-1), \dots, y(t_s-p+1)]^T, \quad (183)$$

$$\hat{\Phi}(p, t_s) = [\hat{\varphi}(t_s), \hat{\varphi}(t_s-1), \dots, \hat{\varphi}(t_s-p+1)], \quad (184)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \psi(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (185)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (186)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (187)$$

$$x_a(t) = \Phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (188)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (189)$$

### 4.5 变递推间隔辅助模型多新息随机梯度算法

#### 4.5.1 稀少量测数据系统的变递推间隔辅助模型多新息随机梯度算法

比较辅助模型投影算法、辅助模型广义投影算法、辅助模型多新息算法之间异同特征,不难基于 V-AM-MI-Proj 算法(144)—(154)和 V-AM-MI-GPJ 算法(168)—(178),重新定义收敛因子  $r(t_s) = 1/\mu(t_s)$ ,就得到辨识稀少量测数据 IN-FIR 系统(125)参数向量  $\theta$  的变递推间隔辅助模型多新息随机梯度算法(interval-Varying AM-MISG algorithm, V-AM-

MISG 算法):

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\Phi}(p, t_s)}{r(t_s)} E(p, t_s), \quad s=1,2,3,\dots, \quad (190)$$

$$\hat{\theta}(t_0) = \mathbf{1}_n/p_0, \quad (190)$$

$$\hat{\theta}(t) = \hat{\theta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (191)$$

$$r(t_s) = \|\hat{\Phi}(p, t_s)\|^2 = \sum_{j=0}^{s-1} \|\hat{\varphi}(t_{s-j})\|^2 = r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2, \quad r(t_0) = 1, \quad (192)$$

$$E(p, t_s) = Y(p, t_s) - \hat{\Phi}^T(p, t_s) \hat{\theta}(t_{s-1}), \quad (193)$$

$$Y(p, t_s) = [y(t_s), y(t_{s-1}), y(t_{s-2}), \dots, y(t_{s-p+1})]^T, \quad (194)$$

$$\hat{\Phi}(p, t_s) = [\hat{\varphi}(t_s), \hat{\varphi}(t_{s-1}), \hat{\varphi}(t_{s-2}), \dots, \hat{\varphi}(t_{s-p+1})], \quad (195)$$

$$\hat{\varphi}(t) = \begin{bmatrix} \phi(t) \\ \psi(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (196)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (197)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (198)$$

$$x_a(t) = \Phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (199)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (200)$$

在收敛性分析中,也可将式(192)修改为

$$r(t_s) = r(t_{s-1}) + \|\hat{\Phi}(p, t_s)\|^2, \quad r(t_0) = 1. \quad (201)$$

注 19 在 V-AM-MISG 算法(190)—(200)中引入遗忘因子(forgetting factor)  $\lambda$ ,将式(192)修改为

$$r(t_s) = \lambda r(t_{s-1}) + \|\hat{\varphi}(t_s)\|^2, \quad 0 \leq \lambda \leq 1, \quad (201)$$

就得到遗忘因子 V-AM-MISG 算法(FF-V-AM-MISG 算法),也称为变递推间隔辅助模型多新息遗忘因子随机梯度算法(V-AM-MI-FFSG 算法),简称变递推间隔辅助模型多新息遗忘梯度算法(V-AM-MIFG 算法).

注 20 在 V-AM-MISG 算法(190)—(200)中引入收敛指数(convergence index)  $\varepsilon$ ,将式(190)修改为

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\Phi}(p, t_s)}{r^\varepsilon(t_s)} E(p, t_s), \quad \frac{1}{2} < \varepsilon \leq 1, \quad (202)$$

就得到修正 V-AM-MISG 算法(M-V-AM-MISG 算法),或称为 Epsilon V-AM-MISG 算法( $\varepsilon$ -V-AM-MISG 算法).

注 21 在 V-AM-MISG 算法(190)—(200)中引入收敛指数  $\varepsilon$  和遗忘因子  $\lambda$ ,将式(190)和(192)修改为

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\hat{\boldsymbol{\Phi}}(p, t_s)}{r^\varepsilon(t_s)} \mathbf{E}(p, t_s), \quad \frac{1}{2} < \varepsilon \leq 1,$$

$$\hat{\boldsymbol{\theta}}(t_0) = \mathbf{1}_n / p_0, \quad (203)$$

$$r(t_s) = \lambda r(t_{s-1}) + \|\hat{\boldsymbol{\varphi}}(t_s)\|^2, \quad 0 \leq \lambda \leq 1, \\ r(t_0) = 1, \quad (204)$$

就得到修正 FF-V-AM-MISG 算法 (M-FF-V-AM-MISG 算法), 或称为 Epsilon FF-V-AM-MISG 算法 ( $\varepsilon$ -FF-V-AM-MISG 算法)。

#### 4.5.2 损失数据系统的变递推间隔辅助模型多新息随机梯度算法

比较辅助模型投影算法、辅助模型广义投影算法、辅助模型多新息算法之间异同特征, 不难基于 V-AM-MI-Proj 算法 (155) — (165) 和 V-AM-MI-GPJ 算法 (179) — (189), 重新定义收敛因子  $r(t_s) = 1/\mu(t_s)$ , 就得到辨识损失数据 IN-FIR 系统 (125) 参数向量  $\boldsymbol{\theta}$  的变递推间隔辅助模型多新息随机梯度算法 (interval-Varying AM-MISG algorithm, V-AM-MISG 算法):

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\hat{\boldsymbol{\Phi}}(p, t_s)}{r(t_s)} \mathbf{E}(p, t_s), \quad s = 1, 2, 3, \dots,$$

$$\hat{\boldsymbol{\theta}}(t_0) = \mathbf{1}_n / p_0, \quad (205)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (206)$$

$$r(t_s) = \sum_{j=0}^{s-1} \|\hat{\boldsymbol{\varphi}}(t_{s-j})\|^2 = r(t_{s-1}) + \|\hat{\boldsymbol{\varphi}}(t_s)\|^2, \\ r(t_0) = 1, \quad (207)$$

$$\mathbf{E}(p, t_s) = \mathbf{Y}(p, t_s) - \hat{\boldsymbol{\Phi}}^T(p, t_s) \hat{\boldsymbol{\theta}}(t_{s-1}), \quad (208)$$

$$\mathbf{Y}(p, t_s) = [y(t_s), y(t_s-1), \dots, y(t_s-p+1)]^T, \quad (209)$$

$$\hat{\boldsymbol{\Phi}}(p, t_s) = [\hat{\boldsymbol{\varphi}}(t_s), \hat{\boldsymbol{\varphi}}(t_s-1), \dots, \hat{\boldsymbol{\varphi}}(t_s-p+1)], \quad (210)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (211)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, \\ f_m(u(t))]^T, \quad (212)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = \\ x_a(t-1:t-n_b), \quad (213)$$

$$x_a(i) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0, 1, \dots, t_1-1, \quad (214)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (215)$$

对损失数据系统的 V-AM-MISG 算法也可引入遗忘因子和收敛指数, 得到相应的遗忘因子辨识算法和修正辨识算法。

**注 22** 当新息长度  $p=1$  时, 本节稀少量测数据系统的辨识方法与损失数据系统的对应辨识方法表达式相同。

## 5 几个特殊的辅助模型梯度辨识方法

V-AM-MI-GPJ 算法中包含两个参数, 一个是新息长度  $p$ , 一个是变递推步长  $t_s^*$ . 当它们取不同值时, 就得到不同的辨识方法. 也就是说, 许多梯度辨识方法是 V-AM-MI-GPJ 算法的特例. 如在 V-AM-MI-GPJ 算法中取  $q=p$ , 就得到 V-AM-MI-Proj 算法; 将 V-AM-MI-GPJ 算法中  $r(t_s) = \|\hat{\boldsymbol{\Phi}}(q, t_s)\|^2$  修改为  $r(t_s) = \|\hat{\boldsymbol{\Phi}}(s, t_s)\|^2$ , 就得到 V-AM-MISG 算法. 下面的算法是从 V-AM-MI-Proj 算法、V-AM-MI-GPJ 算法、V-AM-MI-SG 算法派生出来的。

### 5.1 变递推间隔辅助模型梯度辨识方法

#### 5.1.1 变递推间隔辅助模型投影算法

在稀少量测数据系统的 V-AM-MI-Proj 算法 (144) — (154) 中令新息长度  $p=1$ , 得到变递推间隔辅助模型投影算法 (V-AM-Proj 算法):

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + \frac{\hat{\boldsymbol{\varphi}}(t_s)}{1 + \|\hat{\boldsymbol{\varphi}}(t_s)\|^2} [y(t_s) - \hat{\boldsymbol{\varphi}}^T(t_s) \hat{\boldsymbol{\theta}}(t_{s-1})],$$

$$\hat{\boldsymbol{\theta}}(t_0) = \mathbf{1}_n / p_0, \quad (216)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (217)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (218)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, \\ f_m(u(t))]^T, \quad (219)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = \\ x_a(t-1:t-n_b), \quad (220)$$

$$x_a(i) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0, 1, \dots, t_1-1, \quad (221)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (222)$$

当变递推间隔长度  $t_s^* \equiv 1$  时, 有  $t_s = t$ , 则 V-AM-Proj 算法 (216) — (222) 退化为辅助模型投影算法 (AM-Proj 算法) (35) — (40)。

#### 5.1.2 变递推间隔辅助模型广义投影算法

在 V-AM-MI-GPJ 算法 (168) — (178) 中令新息长度  $p=1$ , 则得到变递推间隔辅助模型广义投影算法 (V-AM-GPJ 算法):

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

$$\hat{\boldsymbol{\theta}}(t_0) = \mathbf{1}_n / p_0, \quad (223)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (224)$$

$$r(t_s) = \sum_{j=0}^{q-1} \|\hat{\boldsymbol{\varphi}}(t_{s-j})\|^2 = r(t_{s-1}) + \\ \|\hat{\boldsymbol{\varphi}}(t_s)\|^2 - \|\hat{\boldsymbol{\varphi}}(t_{s-q})\|^2, \quad r(t_0) = 1, \quad (225)$$

$$\hat{\phi}(t) = \begin{bmatrix} \phi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (226)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (227)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (228)$$

$$x_a(t) = \phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (229)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (230)$$

当记忆长度和数据窗长度  $q=1$  时, V-AM-GPJ 算法(223) — (230) 退化为 V-AM-Proj 算法(216) — (222).

当变递推间隔长度  $t_s^* \equiv 1$  时, 有  $t_s = t$ , 则 V-AM-GPJ 算法(223) — (230) 退化为辅助模型广义投影算法(AM-GPJ 算法)(43) — (49).

### 5.1.3 变递推间隔辅助模型随机梯度算法

在 V-AM-MISG 算法(190) — (200) 中令新息长度  $p=1$ , 则得到变递推间隔辅助模型随机梯度算法(V-AM-SG 算法):

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\phi}(t_s)}{r(t_s)} [y(t_s) - \hat{\phi}^T(t_s) \hat{\theta}(t_{s-1})], \quad (231)$$

$$\hat{\theta}(t) = \hat{\theta}(t_{s-1}), \quad t \in T_s := \{t_{s-1}, t_{s-1}+1, \dots, t_s-1\}, \quad (232)$$

$$r(t_s) = \sum_{j=0}^{s-1} \|\hat{\phi}(t_{s-j})\|^2 = r(t_{s-1}) + \|\hat{\phi}(t_s)\|^2, \quad (233)$$

$$r(t_0) = 1,$$

$$\hat{\phi}(t) = \begin{bmatrix} \phi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (234)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (235)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (236)$$

$$x_a(t) = \phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,t_1-1, \quad (237)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (238)$$

**注 23** 当变递推间隔长度  $t_s^* \equiv 1$  时, 有  $t_s = t$ , 则 V-AM-SG 算法(231) — (238) 退化为辅助模型随机算法(AM-SG 算法)(51) — (57).

**注 24** 在 V-AM-SG 算法(231) — (238) 中引入遗忘因子  $\lambda$ , 将式(233)修改为

$$r(t_s) = \lambda r(t_{s-1}) + \|\hat{\phi}(t_s)\|^2, \quad 0 \leq \lambda \leq 1, \quad (239)$$

$$r(t_0) = 1,$$

就得到遗忘因子 V-AM-SG 算法(FF-V-AM-SG 算法), 也称为变递推间隔辅助模型遗忘因子随机梯度

算法(V-AM-FFSG 算法), 简称变递推间隔辅助模型遗忘梯度算法(V-AM-FG 算法).

**注 25** 在 V-AM-SG 算法(231) — (238) 中引入收敛指数(convergence index)  $\varepsilon$ , 将式(231)修改为

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\phi}(t_s)}{r^\varepsilon(t_s)} [y(t_s) - \hat{\phi}^T(t_s) \hat{\theta}(t_{s-1})], \quad (240)$$

$$\frac{1}{2} < \varepsilon \leq 1, \quad \hat{\theta}(t_0) = \mathbf{1}_n/p_0,$$

就得到修正 V-AM-SG 算法(M-V-AM-SG 算法), 或称为 Epsilon V-AM-SG 算法( $\varepsilon$ -V-AM-SG 算法).

**注 26** 在 V-AM-SG 算法(231) — (238) 中引入收敛指数(convergence index)  $\varepsilon$  和遗忘因子  $\lambda$ , 将式(231)和(233)修改为

$$\hat{\theta}(t_s) = \hat{\theta}(t_{s-1}) + \frac{\hat{\phi}(t_s)}{r^\varepsilon(t_s)} [y(t_s) - \hat{\phi}^T(t_s) \hat{\theta}(t_{s-1})], \quad (241)$$

$$\frac{1}{2} < \varepsilon \leq 1, \quad \hat{\theta}(t_0) = \mathbf{1}_n/p_0,$$

$$r(t_s) = \lambda r(t_{s-1}) + \|\hat{\phi}(t_s)\|^2, \quad 0 \leq \lambda \leq 1, \quad (242)$$

$$r(t_0) = 1,$$

就得到修正 FF-V-AM-SG 算法(M-FF-V-AM-SG 算法), 或称为 Epsilon FF-V-AM-SG 算法( $\varepsilon$ -FF-V-AM-SG 算法).

本节的稀少量测数据系统辨识方法也适合损失数据系统.

## 5.2 等递推间隔辅助模型梯度辨识方法

### 5.2.1 等递推间隔辅助模型投影算法

在 V-AM-Proj 算法(216) — (222) 中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s=1, 2, \dots$ , 则得到等递推间隔辅助模型投影算法(interval-Equating AM-Proj algorithm, E-AM-Proj 算法):

$$\hat{\theta}(ds) = \hat{\theta}(ds-d) + \frac{\hat{\phi}(ds)}{1 + \|\hat{\phi}(ds)\|^2} [y(ds) - \hat{\phi}^T(ds) \hat{\theta}(ds-d)], \quad \hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (243)$$

$$\hat{\theta}(t) = \hat{\theta}(ds-d), \quad t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (244)$$

$$\hat{\phi}(t) = \begin{bmatrix} \phi(t) \\ \hat{\psi}(t) \end{bmatrix} = \begin{bmatrix} f^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (245)$$

$$\phi(t) = f^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (246)$$

$$\hat{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (247)$$

$$x_a(t) = \phi^T(t)\hat{\theta}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,d-1, \quad (248)$$

$$\hat{\theta}(t) = [\hat{\theta}^T(t), \hat{\rho}^T(t)]^T. \quad (249)$$

当记忆长度  $d = 1$  时, E-AM-Proj 算法 (243) — (249) 退化为 AM-Proj 算法 (35) — (40).

### 5.2.2 等递推间隔辅助模型广义投影算法

在 V-AM-GPJ 算法 (223) — (230) 中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s = 1, 2, \dots$ , 则得到等递推间隔辅助模型广义投影算法 (interval-Equating AM-GPJ algorithm, E-AM-GPJ 算法):

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

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \quad (250)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(ds-d),$$

$$t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (251)$$

$$r(ds) = \sum_{j=0}^{q-1} \|\hat{\boldsymbol{\varphi}}(ds-jd)\|^2 = r(ds-d) + \|\hat{\boldsymbol{\varphi}}(ds)\|^2 - \|\hat{\boldsymbol{\varphi}}(ds-qd)\|^2, \quad r(0) = 1, \quad (252)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (253)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (254)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (255)$$

$$x_a(t) = \boldsymbol{\Phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0, 1, \dots, d-1, \quad (256)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (257)$$

当记忆长度  $q = 1$  时, E-AM-GPJ 算法 (250) — (257) 退化为 E-AM-Proj 算法 (243) — (249); 当记忆长度  $d = 1$  时, E-AM-GPJ 算法 (250) — (257) 退化为 AM-GPJ 算法 (43) — (49).

### 5.2.3 等递推间隔辅助模型随机梯度算法

在 V-AM-SG 算法 (231) — (238) 中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s = 1, 2, \dots$ , 则得到等递推间隔辅助模型随机梯度算法 (interval-Equating AM-SG algorithm, E-AM-SG 算法):

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

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \quad (258)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(ds-d),$$

$$t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (259)$$

$$r(ds) = \sum_{j=0}^{s-1} \|\hat{\boldsymbol{\varphi}}(ds-jd)\|^2 = r(ds-d) + \|\hat{\boldsymbol{\varphi}}(ds)\|^2, \quad r(0) = 1, \quad (260)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (261)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (262)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (263)$$

$$x_a(t) = \boldsymbol{\Phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0, 1, \dots, d-1, \quad (264)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (265)$$

当然这个算法也可以引入收敛指数和(或)遗忘因子来改善性能. 等递推间隔辨识方法的一个应用是双率采样数据系统的辨识. 如等递推间隔辅助模型随机梯度算法 (E-AM-SG 算法), 即《双率采样数据系统的辅助模型随机梯度算法及真实输出估计》论文发表在《IEEE Transactions on Automatic Control》2005 年第 9 期上<sup>[6]</sup>.

本节的稀少量测数据系统辨识方法也适合损失数据系统.

## 5.3 等递推间隔辅助模型多新息梯度辨识方法

### 5.3.1 等递推间隔辅助模型多新息投影算法

在稀少量测数据系统的 V-AM-MI-Proj 算法 (144) — (154) 中令变递推间隔步长  $t_s^* \equiv 1$ , 有  $t = t_s$ , 则得到辅助模型多新息投影算法 (AM-MI-Proj 算法) (64) — (72).

在稀少量测数据系统的 V-AM-MI-Proj 算法 (144) — (154) 中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s = 1, 2, \dots$ , 则得到稀少量测数据系统的等递推间隔辅助模型多新息投影算法 (interval-Equating AM-MI-Proj algorithm, E-AM-MI-Proj 算法):

$$\hat{\boldsymbol{\theta}}(ds) = \hat{\boldsymbol{\theta}}(ds-d) + \frac{\hat{\boldsymbol{\Phi}}(p, ds)}{\|\hat{\boldsymbol{\Phi}}(p, ds)\|^2} \mathbf{E}(p, ds),$$

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \quad (266)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(ds-d),$$

$$t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (267)$$

$$\mathbf{E}(p, ds) = \mathbf{Y}(p, ds) - \hat{\boldsymbol{\Phi}}^T(p, ds) \hat{\boldsymbol{\theta}}(ds-d), \quad (268)$$

$$\mathbf{Y}(p, ds) = [y(ds), y(ds-d), y(ds-2d), \dots, y(ds-pd+d)]^T, \quad (269)$$

$$\hat{\boldsymbol{\Phi}}(p, ds) = [\hat{\boldsymbol{\varphi}}(ds), \hat{\boldsymbol{\varphi}}(ds-d), \hat{\boldsymbol{\varphi}}(ds-2d), \dots, \hat{\boldsymbol{\varphi}}(ds-pd+d)], \quad (270)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (271)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (272)$$

$$\hat{\boldsymbol{\psi}}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T =$$

$$x_a(t-1:t-n_b), \quad (273)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,d-1, \quad (274)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (275)$$

将式(269)—(270)修改为

$$\mathbf{Y}(p, ds) = [y(ds), y(ds-1), y(ds-2), \dots, y(ds-p+1)]^T, \quad (276)$$

$$\hat{\boldsymbol{\Phi}}(p, ds) = [\hat{\boldsymbol{\phi}}(ds), \hat{\boldsymbol{\phi}}(ds-1), \hat{\boldsymbol{\phi}}(ds-2), \dots, \hat{\boldsymbol{\phi}}(ds-p+1)], \quad (277)$$

就得到损失数据系统的等递推间隔辅助模型多新息投影算法.

### 5.3.2 等递推间隔辅助模型多新息广义投影算法

在稀少量测数据系统的 V-AM-MI-GPJ 算法(168)—(178)中令变递推间隔步长  $t_s^* \equiv 1$ , 有  $t = t_s$ , 则得到辅助模型多新息广义投影算法(AM-MI-GPJ 算法)(94)—(103).

在稀少量测数据系统的 V-AM-MI-GPJ 算法(168)—(178)中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s = 1, 2, \dots$ , 则得到稀少量测数据系统的等递推间隔辅助模型多新息广义投影算法(interval-Equating AM-MI-GPJ algorithm, E-AM-MI-GPJ 算法):

$$\hat{\boldsymbol{\theta}}(ds) = \hat{\boldsymbol{\theta}}(ds-d) + \frac{\hat{\boldsymbol{\Phi}}(p, ds)}{r(ds)} \mathbf{E}(p, ds), \quad (278)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(ds-d), \quad t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (279)$$

$$r(ds) = \|\hat{\boldsymbol{\Phi}}(q, ds)\|^2 = r(ds-d) + \|\hat{\boldsymbol{\phi}}(ds)\|^2 - \|\hat{\boldsymbol{\phi}}(ds-qd)\|^2, \quad r(0) = 1, \quad (280)$$

$$\mathbf{E}(p, ds) = \mathbf{Y}(p, ds) - \hat{\boldsymbol{\Phi}}^T(p, ds) \hat{\boldsymbol{\theta}}(ds-d), \quad (281)$$

$$\mathbf{Y}(p, ds) = [y(ds), y(ds-d), y(ds-2d), \dots, y(ds-pd+d)]^T, \quad (282)$$

$$\hat{\boldsymbol{\Phi}}(p, ds) = [\hat{\boldsymbol{\phi}}(ds), \hat{\boldsymbol{\phi}}(ds-d), \hat{\boldsymbol{\phi}}(ds-2d), \dots, \hat{\boldsymbol{\phi}}(sd-pd+d)], \quad (283)$$

$$\hat{\boldsymbol{\phi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (284)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (285)$$

$$\boldsymbol{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (286)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,d-1, \quad (287)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (288)$$

将式(280)和(282)—(283)修改为

$$r(ds) = r(ds-d) + \|\hat{\boldsymbol{\phi}}(ds)\|^2 - \|\hat{\boldsymbol{\phi}}(ds-qd)\|^2, \quad r(0) = 1, \quad (289)$$

$$\mathbf{Y}(p, ds) = [y(ds), y(ds-1), y(ds-2), \dots, y(ds-p+1)]^T, \quad (290)$$

$$\hat{\boldsymbol{\Phi}}(p, ds) = [\hat{\boldsymbol{\phi}}(ds), \hat{\boldsymbol{\phi}}(ds-1), \hat{\boldsymbol{\phi}}(ds-2), \dots, \hat{\boldsymbol{\phi}}(ds-p+1)], \quad (291)$$

就得到损失数据系统的等递推间隔辅助模型多新息广义投影算法.

当取记忆长度  $q=p$  时, E-AM-MI-GPJ 算法退化为 E-AM-MI-Proj 算法(266)—(275).

### 5.3.3 等递推间隔辅助模型多新息随机梯度算法

在稀少量测数据系统的 V-AM-MISG 算法(190)—(200)中令变递推间隔步长  $t_s^* \equiv 1$ , 有  $t = t_s$ , 则得到辅助模型多新息随机梯度算法(AM-MISG 算法)(104)—(113).

在稀少量测数据系统的 V-AM-MISG 算法(190)—(200)中令变递推间隔步长  $t_s^* \equiv d$  (正整数), 有  $t_s = ds, s = 1, 2, \dots$ , 则得到稀少量测数据系统的等递推间隔辅助模型多新息随机梯度算法(interval-Equating AM-MISG algorithm, E-AM-MISG 算法):

$$\hat{\boldsymbol{\theta}}(ds) = \hat{\boldsymbol{\theta}}(ds-d) + \frac{\hat{\boldsymbol{\Phi}}(p, ds)}{r(ds)} \mathbf{E}(p, ds), \quad \hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \quad (292)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(ds-d), \quad t \in T_s := \{ds-d, ds-d+1, \dots, ds-1\}, \quad (293)$$

$$r(ds) = \|\hat{\boldsymbol{\Phi}}(s, ds)\|^2 = \sum_{j=0}^{s-1} \|\hat{\boldsymbol{\phi}}(ds-jd)\|^2 = r(ds-d) + \|\hat{\boldsymbol{\phi}}(ds)\|^2, \quad r(0) = 1, \quad (294)$$

$$\mathbf{E}(p, ds) = \mathbf{Y}(p, ds) - \hat{\boldsymbol{\Phi}}^T(p, ds) \hat{\boldsymbol{\theta}}(ds-d), \quad (295)$$

$$\mathbf{Y}(p, ds) = [y(ds), y(ds-d), y(ds-2d), \dots, y(ds-pd+d)]^T, \quad (296)$$

$$\hat{\boldsymbol{\Phi}}(p, ds) = [\hat{\boldsymbol{\phi}}(ds), \hat{\boldsymbol{\phi}}(ds-d), \hat{\boldsymbol{\phi}}(ds-2d), \dots, \hat{\boldsymbol{\phi}}(ds-pd+d)], \quad (297)$$

$$\hat{\boldsymbol{\phi}}(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \boldsymbol{\psi}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(u(t)) \\ x_a(t-1:t-n_b) \end{bmatrix}, \quad (298)$$

$$\boldsymbol{\phi}(t) = \mathbf{f}^T(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (299)$$

$$\boldsymbol{\psi}(t) = [x_a(t-1), x_a(t-2), \dots, x_a(t-n_b)]^T = x_a(t-1:t-n_b), \quad (300)$$

$$x_a(t) = \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t), \quad x_a(i) = 1/p_0, \quad i=0,1,\dots,d-1, \quad (301)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\boldsymbol{\rho}}^T(t)]^T. \quad (302)$$

将式(294)和(296)—(297)修改为

$$r(ds) = \sum_{j=0}^{s-1} \|\hat{\varphi}(ds-jd)\|^2 = r(ds-d) + \|\hat{\varphi}(ds)\|^2, \\ r(0) = 1, \quad (303)$$

$$Y(p, ds) = [y(ds), y(ds-1), y(ds-2), \dots, \\ y(ds-p+1)]^T, \quad (304)$$

$$\hat{\Phi}(p, ds) = [\hat{\varphi}(ds), \hat{\varphi}(ds-1), \hat{\varphi}(ds-2), \dots, \\ \hat{\varphi}(ds-p+1)], \quad (305)$$

则得到损失数据系统的等递推间隔辅助模型多新息随机梯度算法。

**注 27** 在 E-AM-MISG 算法 (292) — (302) 中引入遗忘因子  $\lambda$ , 将式 (294) 修改为

$$r(ds) = \lambda r(ds-d) + \|\hat{\varphi}(ds)\|^2, \quad 0 \leq \lambda \leq 1, \\ r(0) = 1, \quad (306)$$

就得到遗忘因子 E-AM-MISG 算法 (FF-E-AM-MISG 算法), 也称为等递推间隔辅助模型多新息遗忘因子随机梯度算法 (E-AM-FFSG 算法), 简称等递推间隔辅助模型多新息遗忘梯度算法 (E-AM-MIFG 算法)。

**注 28** 在 E-AM-MISG 算法 (292) — (302) 中引入收敛指数 (convergence index)  $\varepsilon$ , 将式 (292) 修改为

$$\hat{\vartheta}(ds) = \hat{\vartheta}(ds-d) + \frac{\hat{\Phi}(p, ds)}{r^\varepsilon(ds)} E(p, ds), \quad \frac{1}{2} < \varepsilon \leq 1, \\ \hat{\vartheta}(0) = \mathbf{1}_n / p_0, \quad (307)$$

就得到修正 E-AM-MISG 算法 (M-E-AM-MISG 算法), 或称为 Epsilon E-AM-MISG 算法 ( $\varepsilon$ -E-AM-MISG 算法)。

**注 29** 在 E-AM-MISG 算法 (292) — (302) 中引入收敛指数 (convergence index)  $\varepsilon$  和遗忘因子  $\lambda$ , 将式 (292) 和 (294) 修改为

$$\hat{\vartheta}(ds) = \hat{\vartheta}(ds-d) + \frac{\hat{\Phi}(p, ds)}{r^\varepsilon(ds)} E(p, ds), \quad \frac{1}{2} < \varepsilon \leq 1, \\ \hat{\vartheta}(0) = \mathbf{1}_n / p_0, \quad (308)$$

$$r(ds) = \lambda r(ds-d) + \|\hat{\varphi}(ds)\|^2, \quad 0 \leq \lambda \leq 1, \\ r(0) = 1, \quad (309)$$

就得到修正 FF-E-AM-MISG 算法 (M-FF-E-AM-MISG 算法), 或称为 Epsilon FF-E-AM-MISG 算法 ( $\varepsilon$ -FF-E-AM-MISG 算法)。

## 6 结语

辅助模型辨识思想已经被用于线性输出误差类系统、多输入输出误差系统、多变量输出误差系统、双率采样数据输出误差系统、非均匀采样输出误差类系统、输入非线性系统的辨识方法研究。本文介绍

了辅助模型辨识思想, 以白噪声干扰下的输入非线性有限脉冲响应系统为例, 讨论了基于辅助模型的梯度辨识方法、多新息梯度辨识方法、变递推间隔梯度辨识方法、等递推间隔梯度辨识方法等。这些方法可以推广, 并结合递阶辨识原理、耦合辨识概念、滤波辨识理念来研究有色噪声干扰的线性系统和非线性系统的辅助模型递阶辨识方法、辅助模型耦合辨识方法、辅助模型滤波辨识方法等。

## 参考文献

### References

- [1] 丁锋. 多变量系统辨识[D]. 北京: 清华大学自动化系, 1990  
DING Feng. Identification of multivariable systems [D]. Beijing: Department of Automation, Tsinghua University, 1990
- [2] 丁锋, 谢新民. 传递函数阵子模型参数递推估计: 辅助模型方法[J]. 控制与决策, 1991, 6(6): 447-452  
DING Feng, XIE Xinmin. Recursive parameter estimation of transfer matrix sub-submodels: An auxiliary model method [J]. Control and Decision, 1991, 6(6): 447-452
- [3] 丁锋, 谢新民. 多变量系统的辅助模型辨识算法[J]. 清华大学学报(自然科学版), 1992, 32(4): 100-106  
DING Feng, XIE Xinmin. Auxiliary model identification method for multivariable systems [J]. Journal of Tsinghua University: Science and Technology, 1992, 32(4): 100-106
- [4] 丁锋. 多变量系统的辅助模型辨识方法的收敛性分析[J]. 控制理论与应用, 1997, 14(2): 192-200  
DING Feng. Convergence analysis of auxiliary model identification algorithms for multivariable systems [J]. Control Theory and Applications, 1997, 14(2): 192-200
- [5] Ding F, Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model [J]. Automatica, 2004, 40(10): 1739-1748
- [6] Ding F, Chen T. Parameter estimation of dual-rate stochastic systems by using an output error method [J]. IEEE Transactions on Automatic Control, 2005, 50(9): 1436-1441
- [7] 王冬青. 基于辅助模型的递推增广最小二乘辨识方法[J]. 控制理论与应用, 2009, 26(1): 51-56  
WANG Dongqing. Recursive extended least squares identification method based on auxiliary models [J]. Control Theory and Applications, 2009, 26(1): 51-56
- [8] Liu Y J, Xiao Y S, Zhao X L. Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model [J]. Applied Mathematics and Computation, 2009, 215(4): 1477-1483
- [9] Xiang L L, Xie L B, Ding R F. Hierarchical least squares algorithms for single-input multiple-output systems based on the auxiliary model [J]. Mathematical and Computer Modelling, 2010, 52(5/6): 918-924
- [10] 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
- [11] Han H Q,Song G L,Xiao Y S,et al.Performance analysis of the AM-SG parameter estimation for multivariable systems[J].Applied Mathematics and Computation,2011,217(12):5566-5572
- [12] Zhang Z N,Jia J,Ding R F.Hierarchical least squares based iterative estimation algorithm for multivariable Box-Jenkins-like systems using the auxiliary model[J].Applied Mathematical Computation,2012,218(9):5580-5587
- [13] Xie L,Yang H Z.Interactive parameter estimation for output error moving average systems[J].Transactions of the Institute of Measurement and Control,2013,35(1):34-43
- [14] Zhang W G.Decomposition based least squares iterative estimation for output error moving average systems[J].Engineering Computations,2014,31(4):709-725
- [15] Ding J,Fan C X,Lin J X.Auxiliary model based parameter estimation for dual-rate output error systems with colored noise[J].Applied Mathematical Modelling,2013,37(6):4051-4058
- [16] Liu X G,Lu J.Least squares based iterative identification for a class of multirate systems[J].Automatica,2010,46(3):549-554
- [17] Xie L,Yang H Z.Gradient based iterative identification for nonuniform sampling output error systems[J].Journal of Vibration and Control,2011,17(3):471-478
- [18] Ding F,Liu G,Liu X P.Parameter estimation with scarce measurements[J].Automatica,2011,47(8):1646-1655
- [19] Ding F.State filtering and parameter identification for state space systems with scarce measurements[J].Signal Processing,2014,104:369-380
- [20] Ding F,Ding J.Least squares parameter estimation with irregularly missing data [J].International Journal of Adaptive Control and Signal Processing,2010,24(7):540-553
- [21] Wang C,Tang T.Recursive least squares estimation algorithm applied to a class of linear-in-parameters output error moving average systems [J].Applied Mathematics Letters,2014,29:36-41
- [22] Wang C,Tang T.Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique [J].Nonlinear Dynamics,2014,77(3):769-780
- [23] Wang C,Zhu L.Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory[J].Journal of the Franklin Institute,2015,352(10):4624-4637
- [24] Guo L J,Wang Y J,Wang C.A recursive least squares algorithm for pseudo-linear arma systems using the auxiliary model and the filtering technique[J].Circuits, Systems and Signal Processing,2016,35(7):2655-2667
- [25] Chen J,Zhang Y,Ding R F.Auxiliary model based multi-innovation algorithms for multivariable nonlinear systems [J].Mathematical and Computer Modelling,2010,52(9/10):1428-1434
- [26] Hu H Y,Ding R.Least squares based iterative identification algorithms for input nonlinear controlled autoregressive systems based on the auxiliary model[J].Nonlinear Dynamics,2014,76(1):777-784
- [27] 丁锋.辅助模型辨识方法(1):自回归输出误差系统[J].南京信息工程大学学报(自然科学版),2016,8(1):1-22  
DING Feng.Auxiliary model based identification methods. Part A:Autoregressive output-error systems[J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2016,8(1):1-22
- [28] 丁锋,陈慧波.辅助模型辨识方法(2):输入非线性输出误差系统[J].南京信息工程大学学报(自然科学版),2016,8(2):97-115  
DING Feng,CHEN Huibo.Auxiliary model based identification methods. Part B: Input nonlinear output-error systems[J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2016,8(2):97-115
- [29] 丁锋,毛亚文.辅助模型辨识方法(3):输入非线性输出误差自回归系统[J].南京信息工程大学学报(自然科学版),2016,8(3):193-214  
DING Feng,MAO Yawen.Auxiliary model based identification methods. Part C: Input nonlinear output-error autoregressive systems[J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2016,8(3):193-214
- [30] Xu L.Application of the Newton iteration algorithm to the parameter estimation for dynamical systems [J].Journal of Computational and Applied Mathematics,2015,288:33-43
- [31] Xu L,Chen L,Xiong W L.Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration [J].Nonlinear Dynamics,2015,79(3):2155-2163
- [32] Xu L.The damping iterative parameter identification method for dynamical systems based on the sine signal measurement[J].Signal Processing,2016,120:660-667
- [33] Ding J,Lin J X.Modified subspace identification for periodically non-uniformly sampled systems by using the lifting technique [J].Circuits, Systems and Signal Processing,2014,33(5):1439-1449
- [34] Xie L,Yang H Z,Ding F,et al.Novel model of non-uniformly sampled-data systems based on a time-varying backward shift operator [J].Journal of Process Control,2016,43:38-52
- [35] Meng D D.Recursive least squares and multi-innovation gradient estimation algorithms for bilinear stochastic systems[J].Circuits, Systems and Signal Processing,2016. doi:10.1007/s00034-016-0337-0
- [36] Chen J,Ni Y X.Parameter identification methods for an additive nonlinear system [J].Circuits, Systems and Signal Processing,2014,33(10):3053-3064
- [37] Chen J,Wang X P.Identification of Hammerstein systems with continuous nonlinearity [J].Information Processing Letters,2015,115(11):822-827
- [38] Wang D Q,Zhang W.Improved least squares identification algorithm for multivariable Hammerstein systems[J].Journal of the Franklin Institute,2015,352

- (11):5292-5307
- [39] Wang D Q. Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models[J]. Applied Mathematics Letters, 2016, 57: 13-19
- [40] Li J H. Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration[J]. Applied Mathematics Letters, 2013, 26(1): 91-96
- [41] Xu L. A proportional differential control method for a time-delay system using the Taylor expansion approximation[J]. Applied Mathematics and Computation, 2014, 236: 391-399
- [42] 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
- [43] 丁锋. 系统辨识: 多新息辨识理论与方法[M]. 北京: 科学出版社, 2016  
DING Feng. System identification: Multi-innovation identification theory and methods [M]. Beijing: Science Press, Beijing, 2016

## Auxiliary model based identification methods. Part D: Basic idea and gradient identification

DING Feng<sup>1,2,3</sup>

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

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

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

**Abstract** The auxiliary model identification idea is presented for the systems with unmeasurable variables. It is an important method for studying the system identification with unknown variables. Introducing the auxiliary model identification idea and taking the input nonlinear finite impulse response system with white noise as an example, this paper studies the auxiliary model (AM) based gradient identification algorithms, the AM multi-innovation gradient identification algorithms, the interval-varying AM multi-innovation gradient identification algorithms; the interval-equating AM gradient identification algorithms and the interval-equating AM multi-innovation gradient identification algorithms.

**Key words** parameter estimation; recursive identification; gradient search; key term separation; auxiliary model identification idea, multi-innovation identification theory; hierarchical identification principle; coupling identification concept; filtering identification idea; input nonlinear system; output nonlinear system