



输入非线性方程误差系统的多新息辨识方法

摘要

针对输入非线性方程误差系统,即输入非线性受控自回归系统,研究了基于过参数化模型的多新息辨识方法和基于过参数化模型的递阶多新息辨识方法;研究了基于关键项分离原理的多新息辨识方法;使用辨识模型分解技术,研究了基于关键项分离原理的两阶段多新息辨识方法和三阶段多新息辨识方法.这些方法可以推广到其他输入非线性方程误差系统、输入非线性输出误差类系统、输出非线性方程误差类系统、输出非线性输出类系统、反馈非线性系统等.同时,给出了几个典型辨识算法的计算量、计算步骤和流程图.

关键词

参数估计;递推辨识;梯度搜索;最小二乘;过参数化模型;关键项分离原理;模型分解;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 2015-04-02

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

随着控制科学的发展和控制技术的应用,时代发生了根本性变革,社会文明取得了极大的进步.与时俱进,作为控制科学的基础学科——系统辨识自形成的半个多世纪以来,也得到空前的发展.在此背景下,2011—2012年、2014年笔者等已经在《南京信息工程大学学报》连载了系统辨识论文18篇,并出版了著作《系统辨识新论》^[1]与《系统辨识——辨识方法性能分析》^[2].2014年的连载论文研究了多元系统、多变量系统、状态空间系统的多新息辨识方法等^[3-8].2015年的连载论文主要基于数据滤波技术,研究了输入非线性方程误差自回归系统的多新息辨识方法^[9].

本文分别基于过参数化模型和基于关键项分离原理,研究输入非线性方程误差系统的多新息辨识方法.非线性现象和非线性系统无处不在,现实中几乎所有的系统都存在一定程度的非线性.研究非线性系统的建模和辨识对此类系统的分析、控制、预测以及优化有着重要的意义.由于非线性系统结构特性复杂,表现形式多样,加上应用领域非常广泛,不可能有统一的非线性系统描述模型和统一的数学理论支持.

对于控制而言,人们不希望使用复杂的模型结构和复杂的数学表达式,能满足精度要求的简单对象模型易于设计控制器,且方便实现,稳定性和可靠性好.数学表达式十分复杂的数学模型,即使精确,也不利于实际应用,不会被控制工程师所接受.

一些特殊模块化结构非线性系统受到了控制科学家的广泛关注.块结构非线性系统是由非线性块与线性动态块两部分构成的,或是由非线性部分和线性部分构成的.非线性部分是一个静态非线性特性,线性部分通常是一个动态子系统.这样的块结构非线性比一般非线性系统处理起来相对容易,也是受到关注的原因之一.值得指出的是,如果非线性部分是一个非线性动态子系统,这样非线性系统也是极其复杂的.因此,通常说的块结构非线性系统都是指非线性部分是静态的.

块结构非线性系统包括:1) 输入非线性系统(非线性模块位于线性模块之前:N—L);2) 输出非线性系统(非线性模块位于线性模块之后:L—N);3) 输入输出非线性系统(两个非线性模块夹着一个线性模块:N—L—N);4) 中间非线性系统(两端线性动态子系统夹着一

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

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,这些典型块结构非线性系统示意如图1和图2所示.

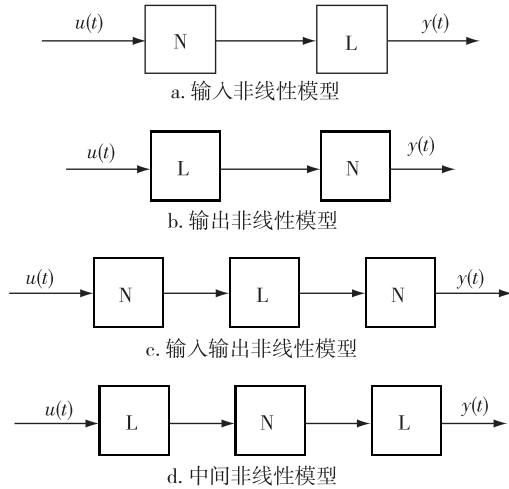


图1 几种非线性系统模型结构

Fig. 1 Several class of nonlinear systems

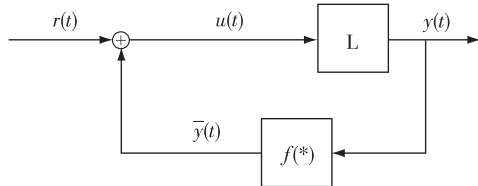


图2 反馈非线性系统模型结构

Fig. 2 A feedback nonlinear system

当输入非线性是一个静态多项式非线性环节,线性环节是一个动态子系统时,这样的非线性系统也称为 Hammerstein 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.

按照干扰噪声性质区分,输入非线性方程误差类系统(IN-EET系统)包含:

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

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

3) 输入非线性方程误差自回归系统(IN-EEAR

系统),又称输入非线性受控自回归自回归系统(IN-CARAR系统),也称输入非线性动态调节系统;

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

输入非线性输出误差类系统(IN-OET系统)包含:

1) 基本输入非线性输出误差系统(IN-OE系统);

2) 输入非线性输出误差滑动平均系统(IN-OEMA系统);

3) 输入非线性输出误差自回归系统(IN-OEAR系统);

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

关键项分离原理是 Vörös 提出的,并被用于研究非线性部分是分段线性、多段线性和间断分段线性化的 Hammerstein 系统的辨识^[10-12].其思想是选择合适的中间变量作为关键项,通过可辨识性要求,将参数归一化,将系统的输出表示为系统参数的线性组合,能避免辨识模型中出现线性块参数与非线性块参数的乘积项.关键项分离原理只解决了分离问题,其分离出来的关键项通常是一个中间变量(即系统内容变量),是未知的,需要借助其他思想方法实现系统未知参数和关键项的联合估计.一种方法是在辨识算法中用关键项的估计代替,也可以采用辅助模型辨识思想——用辅助模型的输出代替.

多新息辨识理论是一种新的辨识理论与方法,已经成功用于线性系统和非线性系统的辨识,在《系统辨识新论》第6章中有专门介绍^[1],多新息辨识方法的一些重要成果都发表在国内外期刊上^[13-34].

本文分别基于过参数化模型,基于关键项分离原理,基于辨识模型分解技术,研究具有已知基的输入非线性方程误差系统的多新息辨识方法、递阶多新息辨识方法、两阶段多新息辨识方法、三阶段多新息辨识方法等.这类非线性系统将 Hammerstein 非线性系统作为特例.当然,还有一些非线性特性,如预负载死区非线性、饱和非线性、滞后继电器非线性、滞后非线性等(图3).由于非线性特性形状各异,不可能存在通用的辨识方法,必须针对不同的非线性特性,采用不同技巧,研究相应的递推辨识方法或迭代辨识方法.这也给辨识科学家提供了许多研究课题^[1].

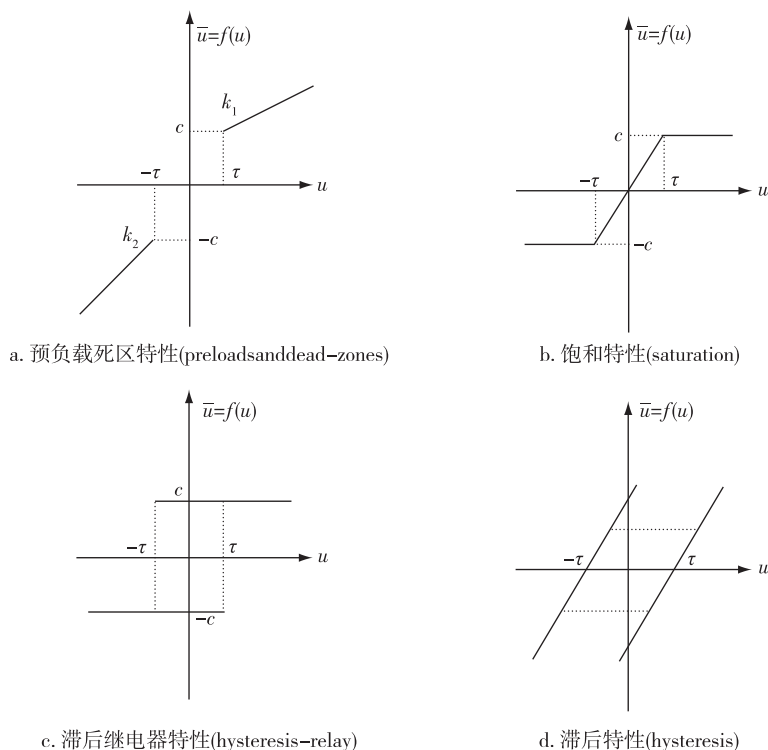


图3 其他非线性特性

Fig. 3 The other input nonlinearities

1 基于过参数化模型的多新息辨识方法

本节基于过参数辨识模型研究输入非线性方程误差系统的多新息辨识方法。

考虑输入非线性方程误差系统的参数辨识问题,它是由一个已知基静态非线性环节串联一个线性动态子系统构成的,其线性环节是一个受控自回归子系统:

$$A(z)y(t) = B(z)\bar{u}(t) + v(t), \quad (1)$$

$\bar{u}(t)$ 和 $y(t)$ 分别为线性动态子系统的输入和输出, $v(t) \in \mathbf{R}$ 是零均值白噪声,中间变量(即非线性环节的输出,也是线性环节的输入) $\bar{u}(t) \in \mathbf{R}$ 是不可测的, $A(z)$ 和 $B(z)$ 是移位算子 $z^{-1} [z^{-1}y(t) = y(t-1)]$ 的多项式,即:

$$A(z) := 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a},$$

$$B(z) := b_1z^{-1} + b_2z^{-2} + \dots + b_{n_b}z^{-n_b}.$$

非线性环节的输出 $\bar{u}(t)$ 是一组系数为 c_j 的已知基函数 $f_j(u(t))$ 的线性组合($u(t)$ 为系统输入,也是非线性环节的输入),即:

$$\bar{u}(t) = \sum_{j=1}^m c_j f_j(u(t)). \quad (2)$$

非线性基函数 $f_j(\cdot)$ 可以是多项式、三角函数、

指数函数、分段线性函数等。式(1)~(2)描述的系统称为输入非线性方程误差系统(Input Nonlinear Equation Error system, IN-EE 系统),即输入非线性受控自回归系统(Input Nonlinear Controlled AutoRegressive system, IN-CAR 系统)。

对于任意非零常数 α , $(\alpha\bar{u}(t), B(z)/\alpha)$ 和 $(\bar{u}(t), B(z))$, 式(1)~(2)是同一个系统。为了确保辨识结果的唯一性,必须对 $\bar{u}(t)$ 或 $B(z)$ 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) $b_1 = 1$ 或 $c_1 = 1$, 其实可以规范化任一非零系数 b_i 或 c_i 为 1;

2) $B(1) = 1$ (或线性动态子系统增益 $B(1)/A(1) = 1$);

3) $b_1^2 + b_2^2 + \dots + b_{n_b}^2 = 1, b_1 > 0$, 或 $c_1^2 + c_2^2 + \dots + c_m^2 = 1, c_1 > 0$ 。

对于基于过参数化模型的辨识方法,这里采用第1)个假设 $b_1 = 1$ 。读者可以采用假设 $c_1 = 1$, 推导相应的辨识方法。

利用移位算子 z^{-1} 的性质,将式(2)代入式(1)整理可得:

$$\begin{aligned}
y(t) = & - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + v(t) = \\
& - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \sum_{j=1}^m c_j f_j(u(t-i)) + v(t) = \\
& - \sum_{i=1}^{n_a} a_i y(t-i) + b_1 c_1 f_1(u(t-1)) + \\
& b_1 c_2 f_2(u(t-1)) + \cdots + b_1 c_m f_m(u(t-1)) + \\
& b_2 c_1 f_1(u(t-2)) + b_2 c_2 f_2(u(t-2)) + \cdots + \\
& b_2 c_m f_m(u(t-2)) + \cdots + b_{n_b} c_1 f_1(u(t-n_b)) + \\
& b_{n_b} c_2 f_2(u(t-n_b)) + \cdots + b_{n_b} c_m f_m(u(t-n_b)) + v(t). \quad (3)
\end{aligned}$$

定义参数向量 θ 和信息向量 $\phi(t)$ 如下:

$$\theta := \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_b} \end{bmatrix} \in \mathbf{R}^n, \quad n := n_a + mn_b, \quad (4)$$

$$\phi(t) := \begin{bmatrix} \phi_0(t) \\ \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_{n_b}(t) \end{bmatrix} \in \mathbf{R}^n, \quad (5)$$

$$\theta_0 := [a_1, a_2, \dots, a_{n_a}]^T \in \mathbf{R}^{n_a}, \quad (6)$$

$$\theta_i := [b_i c_1, b_i c_2, \dots, b_i c_m]^T \in \mathbf{R}^m, \quad i=1, 2, \dots, n_b, \quad (7)$$

$$\phi_0(t) := [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T \in \mathbf{R}^{n_a}, \quad (8)$$

$$\phi_i(t) := [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T \in \mathbf{R}^m. \quad (9)$$

于是, IN-EE 系统(3)可写成线性回归形式:

$$y(t) = \phi^T(t) \theta + v(t), \quad (10)$$

式(10)称为 IN-CAR 系统的过参数化辨识模型(over-parameterization identification model).

注1 如果将多项式 $B(z)$ 修改为

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

那么规范化假设应该作相应修正, 如令 $b_0 = 1$. 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.

注2 这个过参数化辨识模型的参数向量 θ 包含了系统线性块参数 b_i 与非线性块参数 c_j 的乘积项, 使得参数向量 θ 的维数 $n = n_a + mn_b$ 大于系统的实际参数数目 $n_a + n_b + m$ ($m, n_b \geq 2$ 时), 产生很多冗余估计. 特别地, 当 m 和 n_b 很大时, 导致辨识算法的计算量很大, 这就激发我们研究计算量小的辨识方法, 这也是本文的研究目标.

下面讨论基于过参数化模型的输入非线性方程

误差系统的随机梯度辨识方法、多新息随机梯度辨识方法、递推最小二乘辨识方法、多新息最小二乘辨识方法, 进而从获得的参数估计 $\hat{\theta}(t)$ 中提取出原系统参数 a_i, b_i 和 c_i 的估计.

1.1 基于过参数化模型的随机梯度辨识算法

令 $\hat{\theta}(t)$ 为参数向量 θ 在时刻 t 的估计, $\|X\|^2 := \text{tr}[XX^T]$ 表示矩阵(或向量) X 的范数, $\mathbf{1}_n$ 是元均为 1 的 n 维列向量, p_0 是一个很大的正数, 如 $p_0 = 10^6$. 定义梯度准则函数:

$$J_1(\theta) := \frac{1}{2} [y(t) - \phi^T(t) \theta]^2,$$

参考文献[1-2], 利用负梯度搜索, 极小化梯度准则函数 $J_1(\theta)$, 可得辨识参数向量 θ 的随机梯度算法(Stochastic Gradient algorithm, SG 算法):

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) - \frac{1}{r(t)} \text{grad}[J_1(\hat{\theta}(t-1))] \\
&= \hat{\theta}(t-1) + \frac{\phi(t)}{r(t)} [y(t) - \phi^T(t) \hat{\theta}(t-1)], \quad (11)
\end{aligned}$$

$$r(t) = r(t-1) + \|\phi(t)\|^2, \quad r(0) = 1, \quad (12)$$

$$\phi(t) = [\phi_0^T(t), \phi_1^T(t), \phi_2^T(t), \dots, \phi_{n_b}^T(t)]^T, \quad (13)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (14)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T. \quad (15)$$

由于参数向量 θ 包含了参数 b_i 与 c_j 的乘积项, 所以在利用算法(11)——(15)获得 θ 的估计 $\hat{\theta}(t)$ 后, 需要从中提取出原系统参数 $(a_1, a_2, \dots, a_{n_a}), (b_1, b_2, \dots, b_{n_b})$ 和 (c_1, c_2, \dots, c_m) 的估计.

令 $\hat{\theta}_i(t)$ 代表向量 $\hat{\theta}(t)$ 的第 i 个元. 由于假设 $b_1 = 1$, 根据 θ 的定义和式(6)——(7)可知, 系统原参数 a_i 和 c_i 可从 $\hat{\theta}(t)$ 中读取出来, 即:

$$\hat{a}_i(t) = \hat{\theta}_i(t), \quad i=1, 2, \dots, n_a, \quad (16)$$

$$\hat{c}_j(t) = \hat{\theta}_{n_a+j}(t), \quad j=1, 2, \dots, m, \quad (17)$$

$$\hat{b}_i(t) = \frac{\hat{\theta}_{im-m+n_a+j}(t)}{\hat{c}_j(t)}, \quad i=2, 3, \dots, n_b, \quad j=1, 2, \dots, m.$$

由于 $j=1, 2, \dots, m$, 对任意 i ($1 < i \leq n_b$), b_i 有 m 个估计值, 即冗余估计. 这里取这 m 个估计值的平均值作为 b_i 的估计^[35-38], 即:

$$\hat{b}_i(t) = \frac{1}{m} \sum_{j=1}^m \frac{\hat{\theta}_{im-m+n_a+j}(t)}{\hat{c}_j(t)}, \quad i=2, 3, \dots, n_b. \quad (18)$$

式(11)——(18)构成了基于平均值的随机梯度参数估计算法.

注3 需要说明的是,上述算法分两步进行,第一步用式(11)~(15)辨识过参数化模型(10)的参数向量 θ ,第二步用式(16)~(18)从辨识出的参数估计 $\hat{\theta}(t)$ 中计算出系统参数的估计.后述的基于过参数化模型的辨识算法都可采用式(16)~(18)从估计 $\hat{\theta}(t)$ 中提取出原系统参数估计.

注4 随机梯度算法的收敛速度慢,为了提高收敛速度,可引入遗忘因子 λ ,将式(12)修改为

$$r(t) = \lambda r(t-1) + \|\phi(t)\|^2, \quad 0 \leq \lambda \leq 1, \quad r(0) = 1. \quad (19)$$

值得指出的是,所有随机梯度类算法,包括后述的多新息随机梯度算法、递阶随机梯度算法、递阶多新息随机梯度算法、两阶段和三阶段随机梯度算法等,都可以引入遗忘因子来改进算法收敛速度.

1.2 基于过参数化模型的多新息随机梯度辨识算法

随机梯度辨识算法步骤简单、计算量小,但存在参数收敛速度慢、参数估计精度不理想等问题,这就引出了下面的多新息随机梯度辨识算法.

在SG算法(11)~(15)中, $e(t) := y(t) - \phi^T(t) \cdot \hat{\theta}(t-1) \in \mathbf{R}$ 是标量新息^[1,13].设新息长度为 p .定义堆积输出向量 $Y(p, t)$ 和堆积信息矩阵 $\Phi(p, t)$ 如下:

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

将标量新息 $e(t) \in \mathbf{R}$ 扩展为新息向量

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

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

参考文献[1, 13, 15]中多新息随机梯度辨识方法的推导过程,将式(11)~(15)中的标量输出 $y(t)$ 扩展为输出向量 $Y(p, t)$,将信息向量 $\phi(t)$ 扩展为信息矩阵 $\Phi(p, t)$,可得到辨识参数向量 θ 的多新息随机梯度算法(Multi-Innovation Stochastic Gradient algorithm, MISG算法):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\Phi(p, t)}{r(t)} E(p, t), \quad \hat{\theta}(0) = \mathbf{1}_n / p_0, \quad (20)$$

$$E(p, t) = Y(p, t) - \Phi^T(p, t) \hat{\theta}(t-1), \quad (21)$$

$$r(t) = r(t-1) + \|\phi(t)\|^2, \quad r(0) = 1, \quad (22)$$

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (23)$$

$$\Phi(p, t) = [\phi(t), \phi(t-1), \dots, \phi(t-p+1)], \quad (24)$$

$$\phi(t) = [\phi_0^T(t), \phi_1^T(t), \phi_2^T(t), \dots, \phi_{n_b}^T(t)]^T, \quad (25)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (26)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T. \quad (27)$$

当新息长度 $p=1$ 时, MISG算法退化为SG算法(11)~(15).为便于收敛性分析,式(22)中的 $r(t)$ 也可修改为

$$r(t) = r(t-1) + \|\Phi(p, t)\|^2, \quad r(0) = 1.$$

利用式(20)~(27)估计出 $\hat{\theta}(t)$ 后,可借助于式(16)~(18)计算出原系统参数 a_i, b_i 和 c_i 的估计.式(20)~(27)与(16)~(18)构成了基于平均值的多新息随机梯度参数估计算法,其实现步骤如下:

1) 初始化:令 $t=1$,给定新息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$;设置初值 $\hat{\theta}(0) = \mathbf{1}_n / p_0, r(0) = 1, p_0 = 10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$,用式(23)构造 $Y(p, t)$,用式(26)~(27)依次构造 $\phi_0(t), \phi_1(t), \phi_2(t), \dots, \phi_{n_b}(t)$,进而用式(25)和(24)构造信息向量 $\phi(t)$ 和信息矩阵 $\Phi(p, t)$.

3) 用式(21)计算新息向量 $E(p, t)$,用式(22)计算 $r(t)$.

4) 用式(20)刷新参数向量的估计 $\hat{\theta}(t)$.

5) 如果 $t < L, t$ 增加1,转到第2)步,否则,终止计算,获得参数向量的估计 $\hat{\theta}(L)$.

6) 利用式(16)~(18)从 $\hat{\theta}(L)$ 中提取参数估计 $\hat{a}_i(L), \hat{b}_i(L)$ 和 $\hat{c}_i(L)$.

利用MISG算法(20)~(27)与(16)~(18)计算IN-EE系统参数估计的流程如图4所示.

注5 通过扩展新息长度,多新息随机梯度辨识算法能够改进参数估计精度.理论分析表明当新息长度 $p=t$ 时,多新息参数估计逼近最小二乘参数估计.在收敛速度与计算量方面,多新息随机梯度算法是随机梯度算法与最小二乘算法的折中^[1,9].

下面介绍收敛速度快的递推最小二乘算法.

1.3 基于过参数化模型的递推最小二乘辨识算法

考虑式(10)的过参数化模型,即:

$$y(t) = \phi^T(t) \theta + v(t), \quad (28)$$

其中各变量定义同上.定义输出向量 $Y(t)$,信息矩阵 $H(t)$ 和噪声向量 $V(t)$ 如下:

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

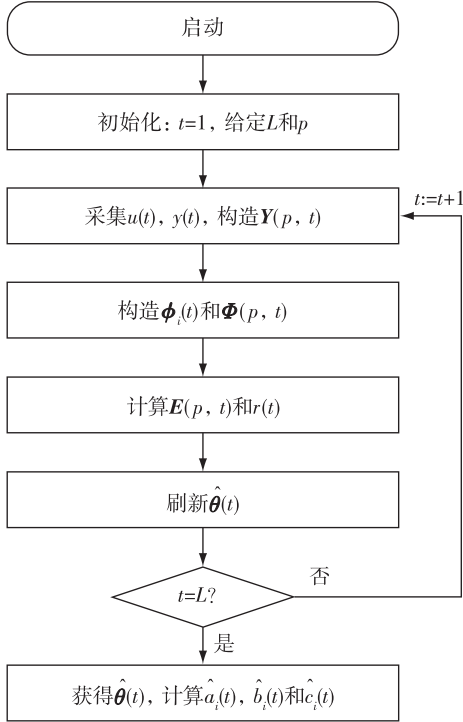


图4 基于过参数化模型的多新息随机梯度算法流程

Fig. 4 The flowchart of the MISG algorithm based on the over-parameterization model

$$\mathbf{H}(t) := \begin{bmatrix} \boldsymbol{\phi}^T(1) \\ \boldsymbol{\phi}^T(2) \\ \vdots \\ \boldsymbol{\phi}^T(t) \end{bmatrix} \in \mathbf{R}^{t \times n},$$

$$\mathbf{V}(t) := \begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(t) \end{bmatrix} \in \mathbf{R}^t,$$

则

$$\mathbf{Y}(t) = \mathbf{H}(t)\boldsymbol{\theta} + \mathbf{V}(t).$$

定义二次准则函数:

$$J_2(\boldsymbol{\theta}) := \frac{1}{2} [\mathbf{Y}(t) - \mathbf{H}(t)\boldsymbol{\theta}]^T [\mathbf{Y}(t) - \mathbf{H}(t)\boldsymbol{\theta}],$$

极小化这个二次准则函数或令 $J_2(\boldsymbol{\theta})$ 对 $\boldsymbol{\theta}$ 的偏导数为零, 可得到估计参数向量 $\boldsymbol{\theta}$ 的递推算法^[1-2]:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}(t)\boldsymbol{\phi}(t) [y(t) - \boldsymbol{\phi}^T(t)\hat{\boldsymbol{\theta}}(t-1)], \quad (29)$$

$$\mathbf{P}^{-1}(t) = \mathbf{P}^{-1}(t-1) + \boldsymbol{\phi}(t)\boldsymbol{\phi}^T(t). \quad (30)$$

根据 $\mathbf{P}^{-1}(t)$ 的定义, 为了实现矩阵 $\mathbf{P}^{-1}(t)$ 的递推计算, 其初值应该设置为 $\mathbf{P}^{-1}(0) = 0$, 导致其逆不存在, 实践中通常给其设置为一个接近于零的正定矩阵, 如 $\mathbf{P}^{-1}(0) = \mathbf{I}_n/p_0$, 即 $\mathbf{P}(0) = p_0\mathbf{I}_n$ 是一个很大的正定矩阵, 其中 \mathbf{I}_n 是一个 n 阶单位阵 (an identity

matrix of order n), p_0 是一个很大的常数, 如取 $p_0 = 10^6$.

为了避免矩阵求逆运算, 将矩阵求逆引理^[1-2]

$$(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \quad (31)$$

应用到式(30), 可以推出辨识参数向量 $\boldsymbol{\theta}$ 的递推最小二乘算法 (Recursive Least Squares algorithm, RLS 算法):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}(t) [y(t) - \boldsymbol{\phi}^T(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \quad (32)$$

$$\mathbf{L}(t) = \mathbf{P}(t-1)\boldsymbol{\phi}(t) [1 + \boldsymbol{\phi}^T(t)\mathbf{P}(t-1)\boldsymbol{\phi}(t)]^{-1}, \quad (33)$$

$$\mathbf{P}(t) = [\mathbf{I}_n - \mathbf{L}(t)\boldsymbol{\phi}^T(t)]\mathbf{P}(t-1), \quad \mathbf{P}(0) = p_0\mathbf{I}_n, \quad (34)$$

$$\boldsymbol{\phi}(t) = [\boldsymbol{\phi}_0^T(t), \boldsymbol{\phi}_1^T(t), \boldsymbol{\phi}_2^T(t), \dots, \boldsymbol{\phi}_{n_b}^T(t)]^T, \quad (35)$$

$$\boldsymbol{\phi}_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n)]^T, \quad (36)$$

$$\boldsymbol{\phi}_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T. \quad (37)$$

利用式(32)–(37)估计出 $\hat{\boldsymbol{\theta}}(t)$ 后, 可借助于式(16)–(18)计算出原系统参数 a_i, b_i 和 c_i 的估计. 式(32)–(37)与(16)–(18)构成了基于平均值的递推最小二乘参数估计算法.

注6 辨识算法的计算量是评价计算效率的一个重要指标. 辨识算法的计算量可用其乘法运算次数和加法运算次数表示. 一次加法运算称为一个 flop, 一次乘法运算也称为一个 flop. 除法作为乘法对待, 减法作为加法对待. 这样就可以用 flop 数, 即浮点运算数来表示计算量的大小. 表1列出了基于过参数化模型的 RLS 算法(32)–(37)的计算量 ($n := n_a + mn_b$).

从表1可以看出, 最小二乘算法每步的计算量是 $N_1 := 4n^2 + 6n$. 简单来说最小二乘算法的计算量是 $O(n^2)$.

1.4 基于过参数化模型的多新息最小二乘辨识算法

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

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

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

借助于多新息辨识理论, 参考多新息最小二乘辨识算法的推导^[1, 14-15], 扩展 RLS 算法(32)–(37)中的标量新息 $e(t) := y(t) - \boldsymbol{\phi}^T(t)\hat{\boldsymbol{\theta}}(t-1)$ 为新息向量

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

表 1 递推最小二乘算法的计算量

Table 1 The computational efficiency of the RLS algorithm

变量	表达式	乘法次数	加法次数
$\hat{\theta}(t)$	$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t)e(t) \in \mathbf{R}^n$	n	n
$e(t)$	$e(t) := y(t) - \hat{\phi}^T(t)\hat{\theta}(t-1) \in \mathbf{R}$	n	n
$L(t)$	$L(t) = \zeta(t) / [1 + \hat{\phi}^T(t)\zeta(t)] \in \mathbf{R}^n$	$2n$	n
$\zeta(t)$	$\zeta(t) := P(t-1)\hat{\phi}(t) \in \mathbf{R}^n$	n^2	$(n-1)n$
$P(t)$	$P(t) = P(t-1) - L(t)\zeta^T(t) \in \mathbf{R}^{n \times n}$	n^2	n^2
总数		$2n^2 + 4n$	$2n^2 + 2n$
总 flop 数		$N_1 := 4n^2 + 6n$	

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

扩展信息向量 $\phi(t)$ 为信息矩阵 $\Phi(p, t)$, 扩展输出 $y(t)$ 为输出向量 $Y(p, t)$, 便得到估计参数向量 θ 的多新息最小二乘算法 (Multi-Innovation Least Squares algorithm, MILS 算法):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t)E(p, t), \quad \hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (38)$$

$$E(p, t) = Y(p, t) - \Phi^T(p, t)\hat{\theta}(t-1), \quad (39)$$

$$L(t) = P(t-1)\Phi(p, t)[I_p + \Phi^T(p, t)P(t-1)\Phi(p, t)]^{-1}, \quad (40)$$

$$P(t) = P(t-1) - L(t)\Phi^T(p, t)P(t-1), \quad P(0) = p_0I_n, \quad (41)$$

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (42)$$

$$\Phi(p, t) = [\phi(t), \phi(t-1), \dots, \phi(t-p+1)], \quad (43)$$

$$\phi(t) = [\phi_0^T(t), \phi_1^T(t), \phi_2^T(t), \dots, \phi_{n_b}^T(t)]^T, \quad (44)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (45)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T. \quad (46)$$

当新息长度 $p = 1$ 时, MILS 算法退化为 RLS 算法 (32) — (37).

利用式 (38) — (46) 估计出 $\hat{\theta}(t)$ 后, 可借助于式 (16) — (18) 计算出原系统参数 a_i, b_i 和 c_i 的估计. 式 (38) — (46) 和 (16) — (18) 构成了基于平均值的多新息最小二乘算法, 算法的实现步骤如下:

1) 初始化: 令 $t = 1$, 给定新息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$; 设置初值 $\hat{\theta}(0) = \mathbf{1}_n/p_0, P(0) = p_0I_n$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式 (42) 构造 $Y(p, t)$, 用式 (45) — (46) 依次构造 $\phi_0(t), \phi_1(t), \phi_2(t), \dots, \phi_{n_b}(t)$, 用式 (44) 和 (43) 构造 $\phi(t)$ 和 $\Phi(p, t)$.

3) 用式 (39) 计算新息向量 $E(p, t)$, 用式 (40) — (41) 计算 $L(t)$ 和 $P(t)$.

4) 用式 (38) 刷新参数向量的估计向量 $\hat{\theta}(t)$.

5) 如果 $t < L, t$ 增加 1 转到第 2) 步, 否则, 获得参数向量的估计 $\hat{\theta}(L)$.

6) 利用式 (16) — (18) 从 $\hat{\theta}(L)$ 中提取参数估计

$\hat{a}_i(L), \hat{b}_i(L)$ 和 $\hat{c}_i(L)$.

基于过参数化模型的 MILS 算法 (38) — (46) 和 (16) — (18) 计算 IN-EE 系统参数估计的流程如图 5 所示, MILS 算法的计算量如表 2 所示 ($n = n_a + mn_b$).

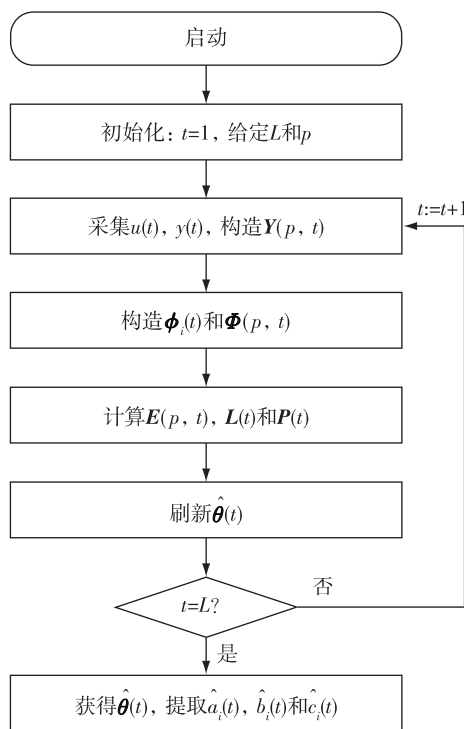


图 5 基于过参数化模型的多新息最小二乘算法流程

Fig. 5 The flowchart of the MILS algorithm based on the over-parameterization model

注 7 值得指出的是, 与多新息随机梯度算法相比, 多新息最小二乘算法改进参数估计精度的能力是非常有限的, 因为最小二乘算法从数据中提取有用信息的能力极其强大, 但是变递推间隔多新息最小二乘算法能很好地提高稀少数据系统和损失数据系统的参数估计精度^[1,9,14].

表 2 多新息最小二乘算法的计算量

Table 2 The computational efficiency of the MILS algorithm

表达式	乘法次数	加法次数
$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t)E(p, t) \in \mathbf{R}^n$	np	np
$E(p, t) = Y(p, t) - \Phi^T(p, t)\hat{\theta}(t-1) \in \mathbf{R}^p$	np	np
$L(t) = \zeta(t)[I_p + \Phi^T(p, t)\zeta(t)]^{-1} \in \mathbf{R}^{n \times p}$	$2np^2 + p^3$	$2np^2 + p^3 - np - 2p^2$
$\zeta(t) := P(t-1)\Phi(p, t) \in \mathbf{R}^{n \times p}$	n^2p	$n^2p - np$
$P(t) = P(t-1) - L(t)\zeta^T(t) \in \mathbf{R}^{n \times n}$	n^2p	n^2p
总数	$2np(n+p+1) + p^3$	$2np(n+p) + p^3 - 2p^2$
总 flop 数	$N_2 := 4n^2p + 4np^2 + 2np + 2p^3 - 2p^2$	

2 基于过参数化模型的递阶多新息辨识方法

输入非线性方程误差系统(输入非线性受控自回归系统)在过参数化后,得到的过参数化辨识模型参数多、维数高,导致辨识算法的计算量大,特别是最小二乘辨识算法涉及计算参数估计误差协方差阵,其计算量就更大.为此,在保持辨识精度或辨识精度不至于牺牲太大的条件下,研究基于过参数化模型分解的辨识方法,即利用递阶辨识原理,研究计算量小的递阶辨识方法,提高计算效率,是十分必要的.本节针对输入非线性方程误差系统,讨论基于过参数化模型分解的递阶随机梯度辨识方法、递阶多新息随机梯度辨识方法、递阶最小二乘辨识方法、递阶多新息最小二乘辨识方法.

2.1 基于过参数化模型的递阶随机梯度辨识算法

利用递阶辨识原理将过参数化模型(28)分解成若干个维数低、变量少的子系统,然后导出这些子系统的辨识方法,最后提取或计算出原系统的参数估计.

根据式(4)–(5)参数向量 θ 和信息向量 $\phi(t)$ 的定义,IN-EE 系统过参数化模型(28)可改写为

$$y(t) = \phi^T(t)\theta + v(t) =$$

$$[\phi_0^T(t), \phi_1^T(t), \phi_2^T(t), \dots, \phi_{n_b}^T(t)] \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_b} \end{bmatrix} + v(t) =$$

$$\phi_0^T(t)\theta_0 + \phi_1^T(t)\theta_1 + \dots + \phi_{n_b}^T(t)\theta_{n_b} + v(t). \quad (47)$$

定义中间变量

$$y_i(t) := y(t) - \sum_{k=0, k \neq i}^{n_b} \phi_k^T(t)\theta_k, \quad i = 0, 1, \dots, n_b. \quad (48)$$

系统(47)可以分解为 $n_b + 1$ 个子系统:

$$y_i(t) = \phi_i^T(t)\theta_i + v(t), \quad i = 0, 1, \dots, n_b. \quad (49)$$

考虑第 i 个子系统,定义梯度准则函数:

$$J_{3i}(\theta_i) := \frac{1}{2} [y_i(t) - \phi_i^T(t)\theta_i]^2.$$

设参数向量 θ_i 在时刻 t 的估计为 $\hat{\theta}_i(t)$.使用负梯度搜索极小化准则函数 $J_{3i}(\theta_i)$,可以得到下列递推关系:

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) - \frac{1}{r_i(t)} \text{grad}[J_{3i}(\hat{\theta}_i(t-1))] =$$

$$\hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t)\hat{\theta}_i(t-1)] =$$

$$\hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i(t)} [y(t) - \phi_i^T(t)\hat{\theta}_i(t-1) - \sum_{k=0, k \neq i}^{n_b} \phi_k^T(t)\theta_k], \quad (50)$$

$$r_i(t) = r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1. \quad (51)$$

算法(50)–(51)不可计算,因为在计算第 i 个子系统的参数向量 θ_i 时,需要用到其他 n_b 个子系统参数向量 $\theta_k (k=0, 1, \dots, i-1, i+1, \dots, n_b)$ 的信息,而它们都是未知的.根据递阶辨识原理,在估计某一个子系统的参数向量时,其余子系统参数向量可以用它们前一时刻的估计代替.因此,将 θ_k 用它在时刻 $t-1$ 的估计 $\hat{\theta}_k(t-1)$ 代替,可得到辨识第 i 个子系统参数向量 θ_i 的随机梯度算法,即递阶随机梯度算法(Hierarchical Stochastic Gradient algorithm, HSG 算法):

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i(t)} [y(t) - \sum_{k=0}^{n_b} \phi_k^T(t)\hat{\theta}_k(t-1)],$$

$$i = 0, 1, \dots, n_b, \quad (52)$$

$$r_i(t) = r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \quad (53)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (54)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T, \quad i = 1, 2, \dots, n_b. \quad (55)$$

当 $i=0, 1, \dots, n_b$ 时,这个算法可以估计出 IN-EE 系统(47)的所有子参数向量 θ_i .在获得所有子系统的参数估计向量 $\hat{\theta}_i(t) (i=0, 1, \dots, n_b)$ 后,可以利用它们提取原系统的参数估计 $\hat{a}_i(t)$, $\hat{b}_i(t)$ 和 $\hat{c}_i(t)$, 或

$\hat{a}(t)$, $\hat{b}(t)$ 和 $\hat{c}(t)$.

根据 $b_1=1$ 的假设, 以及式(6)—(7)关于 θ_i 的定义可得:

$$\theta_0 = [a_1, a_2, \dots, a_{n_a}]^T,$$

$$\theta_1 = [c_1, c_2, \dots, c_m]^T.$$

因此有 $\hat{a}(t) = \hat{\theta}_0(t)$, $\hat{c}(t) = \hat{\theta}_1(t)$. 令 $\hat{\theta}_i(j, t)$ 表示参数估计向量 $\hat{\theta}_i(t)$ 的第 j 个元, 则

$$\hat{a}_j(t) = \hat{\theta}_0(j, t), \quad j=1, 2, \dots, n_a, \quad (56)$$

$$\hat{c}_j(t) = \hat{\theta}_1(j, t), \quad j=1, 2, \dots, m. \quad (57)$$

在获得参数 c_j 的估计 $\hat{c}_j(t)$ 之后, 可用 $\frac{\hat{\theta}_i(j, t)}{\hat{c}_j(t)}$ 计算 $b_i (i=2, 3, \dots, n_b)$ 的估计. 注意到 j 有 m 个取值, $j=1, 2, \dots, m$, 故 b_i 有 m 个不同的估计 $\hat{b}_i(t)$, 这里仍取它们的平均值作为 b_i 的估计, 即:

$$\hat{b}_i(t) = \frac{1}{m} \sum_{j=1}^m \frac{\hat{\theta}_i(j, t)}{\hat{c}_j(t)}, \quad i=2, 3, \dots, n_b. \quad (58)$$

由于基于辨识模型分解的辨识方法称为递阶辨识方法, 因此式(52)—(58)构成了基于平均值的递阶随机梯度辨识算法.

注 8 利用递阶辨识原理, 将过参数化模型(47)分解成 n_b+1 个子系统, 使得信息向量和参数向量的维数从 $n := n_a + mn_b$ 降为 n_a 或 m , 有效地减少了矩阵范数运算量, 提高了辨识效率. 只是随机梯度辨识算法本身计算量较小, 递阶随机梯度算法计算量小的优势不明显.

注 9 如果式(52)中 $r_i(t)$ 都修改为同一 $r(t)$, 即:

$$r(t) = r(t-1) + \sum_{i=0}^{n_b} \|\phi_i(t)\|^2, \quad r(0) = 1, \quad (59)$$

则 HSG 算法等同于 SG 算法(11)—(15). 因此, 对于线性回归模型, 两阶段随机梯度辨识算法、三阶段随机梯度辨识算法、多阶段随机梯度辨识算法、递阶随机梯度辨识算法等与随机梯度算法间只有细微区别.

2.2 基于过参数化模型的递阶多新息随机梯度辨识算法

对第 i 个子系统(49), 定义堆积输出向量 $Y(p, t)$ 及堆积信息矩阵 $\Phi_i(p, t)$ 如下:

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

$$\Phi_0(p, t) := [\phi_0(t), \phi_0(t-1), \dots, \phi_0(t-p+1)] \in \mathbf{R}^{n_a \times p},$$

$$\Phi_i(p, t) := [\phi_i(t), \phi_i(t-1), \dots, \phi_i(t-p+1)] \in \mathbf{R}^{m \times p},$$

$$i=1, 2, \dots, n_b.$$

将式(52)中标量新息

$$e(t) := y(t) - \sum_{k=0}^{n_b} \phi_k^T(t) \hat{\theta}_k(t-1) \in \mathbf{R}$$

扩展为新息向量

$$E(p, t) := \begin{bmatrix} y(t) - \sum_{k=0}^{n_b} \phi_k^T(t) \hat{\theta}_k(t-1) \\ y(t-1) - \sum_{k=0}^{n_b} \phi_k^T(t-1) \hat{\theta}_k(t-1) \\ \vdots \\ y(t-p+1) - \sum_{k=0}^{n_b} \phi_k^T(t-p+1) \hat{\theta}_k(t-1) \end{bmatrix} =$$

$$Y(p, t) - \sum_{k=0}^{n_b} \Phi_k^T(p, t) \hat{\theta}_k(t-1) \in \mathbf{R}^p.$$

借助于多新息辨识理论, 对于式(52)—(55), 将系统输出 $y(t)$ 扩展为输出向量 $Y(p, t)$, 将信息向量 $\phi_0(t)$ 和 $\phi_i(t)$ 扩展为堆积信息矩阵 $\Phi_0(p, t)$ 和 $\Phi_i(p, t)$, 可以得到辨识参数向量 θ_i 的递阶多新息随机梯度算法 (Hierarchical Multi-Innovation Stochastic Gradient algorithm, HMISG 算法):

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + \frac{\Phi_i(p, t)}{r_i(t)} E(p, t), \quad i=0, 1, \dots, n_b, \quad (60)$$

$$E(p, t) = Y(p, t) - \sum_{k=0}^{n_b} \Phi_k^T(p, t) \hat{\theta}_k(t-1), \quad (61)$$

$$r_i(t) = r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \quad (62)$$

$$Y(p, t) := [y(t), y(t-1), \dots, y(t-p+1)], \quad (63)$$

$$\Phi_0(p, t) := [\phi_0(t), \phi_0(t-1), \dots, \phi_0(t-p+1)], \quad (64)$$

$$\Phi_i(p, t) := [\phi_i(t), \phi_i(t-1), \dots, \phi_i(t-p+1)], \quad i=1, 2, \dots, n_b, \quad (65)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n)]^T, \quad (66)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T. \quad (67)$$

当新息长度 $p=1$ 时, HMISG 算法退化为 HSG 算法(52)—(55). 式(62)也可修改为

$$r_i(t) = r_i(t-1) + \|\Phi_i(p, t)\|^2, \quad r_i(0) = 1.$$

利用式(60)—(67)估计出 $\hat{\theta}_i(t)$ 后, 就可借助于式(56)—(58)计算出原系统参数 a_i, b_i 和 c_i 的估计. 式(60)—(67)与(56)—(58)构成了基于平均值的 HMISG 算法, 其实现步骤如下:

1) 初始化: 令 $t=1$, 给定新息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$; 取 $\hat{\theta}_0(0) = \mathbf{1}_{n_a}/p_0, r_0(0) = 1, \hat{\theta}_i(0) = \mathbf{1}_m/p_0, r_i(0) = 1, i=1, 2, \dots, n_b, p_0 = 10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式(63)构造 $Y(p, t)$, 用式(66)—(67)构造 $\phi_0(t), \phi_1(t), \phi_2(t), \dots, \phi_{n_b}(t)$, 用式(64)—(65)构造 $\Phi_0(p, t),$

$\Phi_1(p, t), \Phi_2(p, t), \dots, \Phi_{n_b}(p, t)$.

3) 用式(61)计算新息向量 $\mathbf{E}(p, t)$, 用式(62)计算 $r_i(t)$.

4) 用式(60)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_i(t)$, $i=0, 1, \dots, n_b$.

5) 如果 $t < L$, t 增加 1 转到第 2) 步, 否则获得参数估计向量 $\hat{\boldsymbol{\theta}}_i(L)$.

6) 利用式(56)—(58)从 $\hat{\boldsymbol{\theta}}_i(L)$ 中提取原系统参数估计 $\hat{a}_i(L)$, $\hat{b}_i(L)$ 和 $\hat{c}_i(L)$.

基于过参数化模型的 HMISG 算法(60)—(67)与(56)—(58)计算 IN-EE 系统参数估计的流程如图 6 所示.

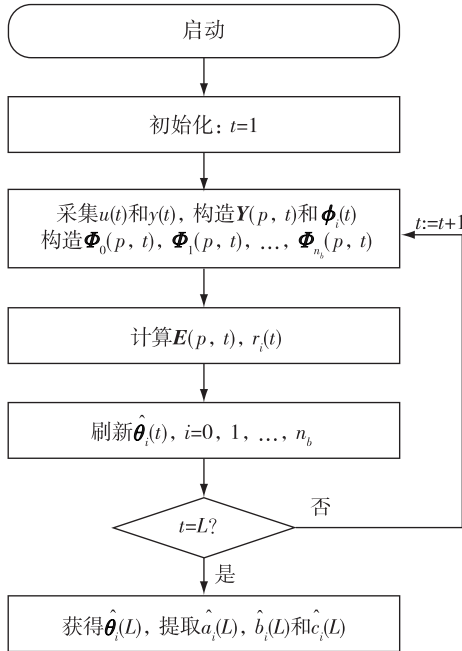


图 6 基于过参数化模型的 HMISG 算法流程
Fig. 6 The flowchart of the HMISG algorithm based on the over-parameterization model

2.3 基于过参数化模型的递阶最小二乘辨识算法

利用递阶辨识原理, 将 IN-EE 系统(47)分解成 n_b+1 个子系统, 然后再利用最小二乘搜索辨识这些子系统的参数向量, 最后提取或计算出系统的原参数.

考虑由 IN-CAR 系统过参数模型分解得到的 n_b+1 个虚拟子系统

$$y_i(t) = \boldsymbol{\phi}_i^T(t) \boldsymbol{\theta}_i + v(t), \quad i=0, 1, \dots, n_b, \quad (68)$$

其中

$$y_i(t) := y(t) - \sum_{k=0, k \neq i}^n \boldsymbol{\phi}_k^T(t) \boldsymbol{\theta}_k,$$

$$\boldsymbol{\phi}_0(t) := \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \vdots \\ -y(t-n_a) \end{bmatrix} \in \mathbf{R}^{n_a},$$

$$\boldsymbol{\phi}_i(t) := \begin{bmatrix} f_1(u(t-i)) \\ f_2(u(t-i)) \\ \vdots \\ f_m(u(t-i)) \end{bmatrix} \in \mathbf{R}^m.$$

令 $\hat{\boldsymbol{\theta}}_i(t)$ 表示第 i 个子系统参数向量 $\boldsymbol{\theta}_i$ 在时刻 t 的估计. 针对第 i 个子系统, 定义并极小化二次准则函数:

$$J_{4i}(\boldsymbol{\theta}_i) := \frac{1}{2} \sum_{j=1}^t [y_i(j) - \boldsymbol{\phi}_i^T(j) \boldsymbol{\theta}_i]^2.$$

令 $J_{4i}(\boldsymbol{\theta}_i)$ 对 $\boldsymbol{\theta}_i$ 的梯度为零, 可得计算 $\hat{\boldsymbol{\theta}}_i(t)$ 的递推算法^[1,39]:

$$\hat{\boldsymbol{\theta}}_i(t) = \hat{\boldsymbol{\theta}}_i(t-1) + \mathbf{P}_i(t) \boldsymbol{\phi}_i(t) [y_i(t) - \boldsymbol{\phi}_i^T(t) \hat{\boldsymbol{\theta}}_i(t-1)] = \hat{\boldsymbol{\theta}}_i(t-1) + \mathbf{P}_i(t) \boldsymbol{\phi}_i(t) [y(t) - \boldsymbol{\phi}_i^T(t) \hat{\boldsymbol{\theta}}_i(t-1) - \sum_{k=0, k \neq i}^{n_b} \boldsymbol{\phi}_k^T(t) \boldsymbol{\theta}_k], \quad (69)$$

$$\mathbf{P}_i^{-1}(t) = \mathbf{P}_i^{-1}(t-1) + \boldsymbol{\phi}_i(t) \boldsymbol{\phi}_i^T(t). \quad (70)$$

注意到在计算第 i 个子系统的参数向量 $\boldsymbol{\theta}_i$ 时, 需要用到其他 n_b 个子系统的参数向量 $\boldsymbol{\theta}_k (k \neq i)$ 的信息, 而它们都是未知待辨识的. 借助递阶辨识原理, 将未知参数向量 $\boldsymbol{\theta}_k$ 用其在时刻 $t-1$ 的估计 $\hat{\boldsymbol{\theta}}_k(t-1)$ 代替, 可得:

$$\hat{\boldsymbol{\theta}}_i(t) = \hat{\boldsymbol{\theta}}_i(t-1) + \mathbf{P}_i(t) \boldsymbol{\phi}_i(t) \left[y(t) - \sum_{k=0}^{n_b} \boldsymbol{\phi}_k^T(t) \hat{\boldsymbol{\theta}}_k(t-1) \right]. \quad (71)$$

为了避免矩阵求逆运算, 将矩阵求逆引理(式(31))应用到式(70)可得

$$\mathbf{P}_i(t) = \mathbf{P}_i(t-1) - \mathbf{L}_i(t) \boldsymbol{\phi}_i^T(t) \mathbf{P}_i(t-1), \quad (72)$$

其中增益向量 $\mathbf{L}_i(t)$ 定义为

$$\mathbf{L}_i(t) := \mathbf{P}_i(t) \boldsymbol{\phi}_i(t) = \frac{\mathbf{P}_i(t-1) \boldsymbol{\phi}_i(t)}{1 + \boldsymbol{\phi}_i^T(t) \mathbf{P}_i(t-1) \boldsymbol{\phi}_i(t)}.$$

由此可得辨识参数向量 $\boldsymbol{\theta}_i$ 的递阶最小二乘算法 (Hierarchical Least Squares algorithm, HLS 算法)^[40]:

$$\hat{\boldsymbol{\theta}}_i(t) = \hat{\boldsymbol{\theta}}_i(t-1) + \mathbf{L}_i(t) \left[y(t) - \sum_{k=0}^{n_b} \boldsymbol{\phi}_k^T(t) \hat{\boldsymbol{\theta}}_k(t-1) \right], \quad i=0, 1, \dots, n_b, \quad (73)$$

$$L_i(t) = \frac{P_i(t-1)\phi_i(t)}{1 + \phi_i^T(t)P_i(t-1)\phi_i(t)}, \quad (74)$$

$$P_i(t) = P_i(t-1) - L_i(t)\phi_i^T(t)P_i(t-1), \quad (75)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (76)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T, \quad (77)$$

$i = 1, 2, \dots, n_b.$

令 $i = 0, 1, \dots, n_b$, 一旦 HLS 算法(73)—(77)可估计出所有参数向量 $\hat{\theta}_i(t)$ 后,可借助于式(56)—(58)计算出原系统参数 a_i, b_i 和 c_i 的估计。

注意到参数向量 θ_0 和信息向量 $\phi_0(t)$ 是 n_a 维,其他 n_b 个参数向量和与其对应的信息向量都是 m 维的.因此,参数向量和协方差矩阵的初始值可设置如下:

$$\hat{\theta}_0(0) = \mathbf{1}_{n_a}/p_0, \quad P_0(0) = p_0 \mathbf{I}_{n_a}, \quad p_0 = 10^6, \quad (78)$$

$$\hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad P_i(0) = p_0 \mathbf{I}_m, \quad i = 1, 2, \dots, n_b. \quad (79)$$

式(73)—(79)与(56)—(58)构成了基于平均值的递阶最小二乘辨识算法,其计算量如表3所示。

2.4 基于过参数化模型的递阶多新息最小二乘辨识算法

定义堆积输出向量 $Y(p, t)$ 和堆积信息矩阵 $\Phi_i(p, t)$ 如下:

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

$$\Phi_0(p, t) := [\phi_0(t), \phi_0(t-1), \dots, \phi_0(t-p+1)] \in \mathbf{R}^{n_a \times p},$$

$$\Phi_i(p, t) := [\phi_i(t), \phi_i(t-1), \dots, \phi_i(t-p+1)] \in \mathbf{R}^{m \times p},$$

$$i = 1, 2, \dots, n_b.$$

借助于多新息辨识理论,将式(73)中的标量新息

$$e(t) := y(t) - \sum_{k=0}^{n_b} \phi_k^T(t) \hat{\theta}_k(t-1) \in \mathbf{R}$$

扩展为新息向量

$$E(p, t) = Y(p, t) - \sum_{k=0}^{n_b} \Phi_k^T(p, t) \hat{\theta}_k(t-1) \in \mathbf{R}^p.$$

扩展 HLS 算法(73)—(79)中系统输出 $y(t)$ 为堆积输出向量 $Y(p, t)$,信息向量 $\phi_0(t)$ 和 $\phi_i(t)$ 为堆积信息矩阵 $\Phi_0(p, t)$ 和 $\Phi_i(p, t)$,可以得到辨识参数向量 θ_i 的递阶多新息最小二乘算法(Hierarchical Multi-Innovation Least Squares algorithm, HMILS 算法):

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + L_i(t)E(p, t), \quad i = 0, 1, \dots, n_b, \quad (80)$$

$$E(p, t) = Y(p, t) - \sum_{k=0}^{n_b} \Phi_k^T(p, t) \hat{\theta}_k(t-1), \quad (81)$$

$$L_i(t) = P_i(t-1) \Phi_i(p, t) [I_p + \Phi_i^T(p, t) P_i(t-1) \Phi_i(p, t)]^{-1}, \quad (82)$$

$$P_i(t) = P_i(t-1) - L_i(t) \Phi_i^T(p, t) P_i(t-1), \quad (83)$$

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (84)$$

$$\Phi_0(p, t) := [\phi_0(t), \phi_0(t-1), \dots, \phi_0(t-p+1)], \quad (85)$$

$$\Phi_i(p, t) := [\phi_i(t), \phi_i(t-1), \dots, \phi_i(t-p+1)],$$

$$i = 1, 2, \dots, n_b, \quad (86)$$

$$\phi_0(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (87)$$

$$\phi_i(t) = [f_1(u(t-i)), f_2(u(t-i)), \dots, f_m(u(t-i))]^T, \quad (88)$$

$$\hat{\theta}_0(0) = \mathbf{1}_{n_a}/p_0, \quad P_0(0) = p_0 \mathbf{I}_{n_a}, \quad (89)$$

$$\hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad P_i(0) = p_0 \mathbf{I}_m, \quad i = 1, 2, \dots, n_b. \quad (90)$$

当新息长度 $p = 1$ 时, HMILS 算法退化为 HLS 算法(73)—(79)。

利用式(80)—(90)估计出 $\hat{\theta}_i(t)$ 后,可借助于式(56)—(58)计算出原系统参数 a_i, b_i 和 c_i 的估计.式(80)—(90)与(56)—(58)构成了基于平均值的递阶多新息最小二乘辨识算法,其计算步骤如下:

表3 递阶最小二乘算法的计算量

Table 3 The computational efficiency of the HLS algorithm

变量	算法	乘法次数	加法次数
$\hat{\theta}_0(t)$	$\hat{\theta}_0(t) = \hat{\theta}_0(t-1) + L_0(t)e(t) \in \mathbf{R}^{n_a}$	n_a	n_a
	$e(t) := y(t) - \sum_{i=0}^{n_b} \phi_i^T(t) \hat{\theta}_i(t-1) \in \mathbf{R}$	$n_a + mn_b$	$n_a + mn_b$
$L_0(t)$	$L_0(t) = \xi_0(t) / [1 + \phi_0^T(t) \xi_0(t)] \in \mathbf{R}^{n_a}$	$2n_a$	n_a
	$\xi_0(t) := P_0(t-1) \phi_0(t) \in \mathbf{R}^{n_a}$	n_a^2	$(n_a - 1)n_a$
$P_0(t)$	$P_0(t) = P_0(t-1) - L_0(t) \xi_0^T(t) \in \mathbf{R}^{n_a \times n_a}$	n_a^2	n_a^2
$\hat{\theta}_i(t)$	$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + L_i(t)e(t) \in \mathbf{R}^m$	m	m
$L_i(t)$	$L_i(t) = \xi_i(t) / [1 + \phi_i^T(t) \xi_i(t)] \in \mathbf{R}^m$	$2m$	m
	$\xi_i(t) := P_i(t-1) \phi_i(t) \in \mathbf{R}^m$	m^2	$m(m-1)$
$P_i(t)$	$P_i(t) = P_i(t-1) - L_i(t) \xi_i^T(t) \in \mathbf{R}^{m \times m}$	m^2	m^2
总数		$2m^2 n_b + 4mn_b + 2n_a^2 + 4n_a$	$2m^2 n_b + 2mn_b + 2n_a^2 + 2n_a$
总 flop 数		$N_3 := 4n_a^2 + 4m^2 n_b + 6n_a + 6mn_b$	

1) 初始化:令 $t=1$, 选择数据长度 L , 给定新息长度 p 和基函数 $f_j(\cdot)$, 根据式 (89) — (90) 对参数估计向量和协方差阵赋初值.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式 (84) 构造 $Y(p, t)$, 用式 (87) — (88) 构造 $\phi_0(t), \phi_1(t), \phi_2(t), \dots, \phi_{n_b}(t)$, 用式 (85) — (86) 构造 $\Phi_0(p, t), \Phi_1(p, t), \Phi_2(p, t), \dots, \Phi_{n_b}(p, t)$.

3) 用式 (81) 计算新息向量 $E(p, t)$, 用式 (82) — (83) 计算 $L_i(t)$ 和 $P_i(t)$.

4) 用式 (80) 刷新参数估计向量 $\hat{\theta}_i(t), i=0, 1, \dots, n_b$.

5) 如果 $t < L$, t 增加 1 转到第 2) 步, 否则获得参数向量的估计 $\hat{\theta}_i(L)$.

6) 利用式 (56) — (58) 从 $\hat{\theta}_i(L)$ 中提取原系统的参数估计 $\hat{a}_i(L), \hat{b}_i(L)$ 和 $\hat{c}_i(L)$.

递阶多新息最小二乘辨识算法 (80) — (90) 与 (56) — (58) 计算 IN-EE 系统参数估计的计算量如表 4 所示 ($n := n_a + mn_b$).

3 基于关键项分离的多新息辨识方法

前两节基于输入非线性方程误差系统的过参数化辨识模型 (10), 研究了多新息辨识方法和递阶多新息辨识方法. 该过参数化辨识模型参数向量 θ 中含有 $n_a + mn_b$ 个参数, 大于系统实际参数 $n_a + n_b + m$ 个 ($m, n_b > 2$). 参数数目增加了, 导致辨识方法计算量增加, 也出现很多冗余参数估计, 而且在辨识出过参数化向量后, 还需从中提取系统的实际参数估计, 又

增加额外计算开支. 为避免过参数模型出现冗余参数估计和辨识方法计算量大的问题 (指同类算法间的比较), 下面借用关键项分离原理, 研究基于关键项分离的随机梯度辨识方法以及多新息随机梯度辨识方法、递推最小二乘辨识方法、多新息最小二乘辨识方法.

考虑输入非线性方程误差系统 (1) — (2), 重写如下:

$$y(t) = - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=0}^{n_b} b_i \bar{u}(t-i) + v(t), \quad (91)$$

$$\bar{u}(t) = \sum_{j=1}^m c_j f_j(u(t)). \quad (92)$$

考虑到系统的可辨识性, 这里假设 $b_0 = 1$. 选择中间变量 $\bar{u}(t)$ 作为关键项, IN-EE 系统 (91) — (92) 可写为

$$\begin{aligned} y(t) &= - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=0}^{n_b} b_i \bar{u}(t-i) + v(t) = \\ &= - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + b_0 \bar{u}(t) + v(t) = \\ &= - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + \sum_{j=1}^m c_j f_j(u(t)) + v(t). \end{aligned} \quad (93)$$

定义线性模块的参数向量 \mathbf{a} 和 \mathbf{b} , 以及非线性模块的参数向量 \mathbf{c} 为

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

表 4 递阶多新息最小二乘算法的计算量

Table 4 The computational efficiency of the HMILS algorithm

表达式	乘法次数	加法次数
$\hat{\theta}_0(t) = \hat{\theta}_0(t-1) + L_0(t)E(p, t) \in \mathbf{R}^{n_a}$	$n_a p$	$n_a p$
$E(p, t) = Y(p, t) - \sum_{i=0}^{n_b} \Phi_i^T(p, t) \hat{\theta}_i(t-1) \in \mathbf{R}^p$	$n_a p + mn_b p$	$n_a p + mn_b p$
$L_0(t) = \zeta_0(t) [I_p + \Phi_0^T(p, t) \zeta_0(t)]^{-1} \in \mathbf{R}^{n_a \times p}$	$2n_a p^2 + p^3$	$2n_a p^2 + p^3 - n_a p - 2p^2$
$\zeta_0(t) := P_0(t-1) \Phi_0(p, t) \in \mathbf{R}^{n_a \times p}$	$n_a^2 p$	$n_a^2 p - n_a p$
$P_0(t) = P_0(t-1) - L_0(t) \zeta_0^T(t) \in \mathbf{R}^{n_a \times n_a}$	$n_a^2 p$	$n_a^2 p$
$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + L_i(t)E(p, t) \in \mathbf{R}^m$	mp	mp
$L_i(t) = \zeta_i(t) [I_p + \Phi_i^T(p, t) \zeta_i(t)]^{-1} \in \mathbf{R}^{m \times p}$	$2mp^2 + p^3$	$2mp^2 + p^3 - mp - 2p^2$
$\zeta_i(t) := P_i(t-1) \Phi_i(p, t) \in \mathbf{R}^{m \times p}$	$m^2 p$	$m^2 p - mp$
$P_i(t) = P_i(t-1) - L_i(t) \zeta_i^T(t) \in \mathbf{R}^{m \times m}$	$m^2 p$	$m^2 p$
总数	$2n_a^2 p + 2n_a p^2 + 2n_a p + p^3 + n_b(2m^2 p + 2mp^2 + 2mp + p^3)$	$2n_a^2 p + 2n_a p^2 + p^3 - 2p^2 + n_b(2m^2 p + 2mp^2 + p^3 - 2p^2)$
总 flop 数	$N_4 := 4p(n_a^2 + mn_b^2) + 4np^2 + 2np + 2(n_b + 1)(p^3 - p^2)$	

$$\mathbf{c} := \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \in \mathbf{R}^m,$$

定义相应的信息向量 $\boldsymbol{\varphi}_a(t)$, $\boldsymbol{\varphi}_b(t)$ 和 $\mathbf{f}(t)$ 为

$$\boldsymbol{\varphi}_a(t) := \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \vdots \\ -y(t-n_a) \end{bmatrix} \in \mathbf{R}^{n_a}, \quad \boldsymbol{\varphi}_b(t) := \begin{bmatrix} \bar{u}(t-1) \\ \bar{u}(t-2) \\ \vdots \\ \bar{u}(t-n_b) \end{bmatrix} \in \mathbf{R}^{n_b},$$

$$\mathbf{f}(t) := \begin{bmatrix} f_1(u(t)) \\ f_2(u(t)) \\ \vdots \\ f_m(u(t)) \end{bmatrix} \in \mathbf{R}^m.$$

进一步,令信息向量

$$\boldsymbol{\varphi}(t) := [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \mathbf{f}^T(t)]^T \in \mathbf{R}^{n_0}, \quad n_0 := n_a + n_b + m,$$

令参数向量

$$\boldsymbol{\vartheta} := [\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T]^T \in \mathbf{R}^{n_0},$$

则模型(93)可改写为

$$y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta} + v(t). \quad (94)$$

注 10 式(94)称为基于关键项分离的辨识模型,其输出 $y(t)$ 表示为未知参数的线性组合,是一个伪线性辨识问题.与前述的过参数化模型(28)相比,辨识模型(94)参数向量包含的都是系统的真实参数 a_i, b_i 和 c_i ,没有出现参数乘积项,不会产生冗余参数估计.

注 11 该模型参数向量维数比过参数化模型参数向量维数小,相应的辨识方法的计算量也会小.基于关键项分离的辨识模型(94)是一个伪线性辨识问题,但是信息向量 $\boldsymbol{\varphi}(t)$ 中的 $\boldsymbol{\varphi}_b(t)$ 包含了未知变量 $\bar{u}(t-i)$,在辨识算法中只能使用其估计代替(使用辅助模型输出代替),这可能影响参数估计精度.基于关键项分离的非线性系统辨识算法的收敛性分析仍是辨识领域有待解决的难题,如:

1) 基于关键项分离的 IN-EE 系统随机梯度算法收敛性分析;

2) 基于关键项分离的 IN-EE 系统多新息随机梯度算法收敛性分析;

3) 基于关键项分离的 IN-EE 系统递推最小二乘算法收敛性分析;

4) 基于关键项分离的 IN-EE 系统多新息最小二乘算法收敛性分析等.

下面讨论基于关键项分离的 IN-EE 系统的随机梯度辨识方法、多新息随机梯度辨识算法、递推最小

二乘辨识算法、多新息最小二乘辨识算法.

3.1 基于关键项分离的随机梯度辨识算法

$$\text{令 } \hat{\boldsymbol{\vartheta}}(t) := \begin{bmatrix} \hat{\mathbf{a}}(t) \\ \hat{\mathbf{b}}(t) \\ \hat{\mathbf{c}}(t) \end{bmatrix} \in \mathbf{R}^{n_0} \text{ 是参数向量 } \boldsymbol{\vartheta} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} \text{ 在}$$

时刻 t 的估计.对于辨识模型(94),定义梯度准则函数:

$$J(\boldsymbol{\vartheta}) := \frac{1}{2} \|y(t) - \boldsymbol{\varphi}^T(t) \boldsymbol{\vartheta}\|^2.$$

参考文献[1-2],利用负梯度搜索,极小化梯度准则函数 $J(\boldsymbol{\vartheta})$,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的随机梯度算法:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r(t)} [y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (95)$$

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2, \quad r(0) = 1. \quad (96)$$

注意到信息向量 $\boldsymbol{\varphi}(t)$ 中的 $\boldsymbol{\varphi}_b(t)$ 含有未知变量 $\bar{u}(t-i)$,算法(95)–(96)是不可实现的.解决方法是按照式(92)构造辅助模型

$$\hat{\bar{u}}(t) = \sum_{j=1}^m \hat{c}_j(t) f_j(u(t)) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t).$$

辅助模型的输出 $\hat{\bar{u}}(t)$ 可以看作是 $\bar{u}(t)$ 的估计.由于采用递推方案,一旦 \mathbf{c} 的估计 $\hat{\mathbf{c}}(t)$ 计算得到,就可用上式计算 $\bar{u}(t)$ 的估计 $\hat{\bar{u}}(t)$.用 $\hat{\bar{u}}(t-i)$ 构造 $\boldsymbol{\varphi}_b(t)$ 的估计

$$\hat{\boldsymbol{\varphi}}_b(t) := [\hat{\bar{u}}(t-1), \hat{\bar{u}}(t-2), \dots, \hat{\bar{u}}(t-n_b)]^T \in \mathbf{R}^{n_b},$$

进而得到信息向量 $\boldsymbol{\varphi}(t)$ 的估计

$$\hat{\boldsymbol{\varphi}}(t) := [\hat{\boldsymbol{\varphi}}_a^T(t), \hat{\boldsymbol{\varphi}}_b^T(t), \mathbf{f}^T(t)]^T \in \mathbf{R}^{n_0}.$$

将式(95)–(96)中未知信息向量 $\boldsymbol{\varphi}(t)$ 用它的估计 $\hat{\boldsymbol{\varphi}}(t)$ 代替,就得到了辨识 IN-EE 系统参数向量 $\boldsymbol{\vartheta}$ 的基于关键项分离的随机梯度算法(Key Term separation based Stochastic Gradient algorithm, KT-SG 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)} [y(t) - \hat{\boldsymbol{\varphi}}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (97)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2, \quad r(0) = 1, \quad (98)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\hat{\boldsymbol{\varphi}}_a^T(t), \hat{\boldsymbol{\varphi}}_b^T(t), \mathbf{f}^T(t)]^T, \quad (99)$$

$$\hat{\boldsymbol{\varphi}}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (100)$$

$$\mathbf{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (101)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{\bar{u}}(t-1), \hat{\bar{u}}(t-2), \dots, \hat{\bar{u}}(t-n_b)]^T, \quad (102)$$

$$\hat{\bar{u}}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t), \quad (103)$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\mathbf{a}}^T(t), \hat{\mathbf{b}}^T(t), \hat{\mathbf{c}}^T(t)]^T. \quad (104)$$

注 12 与基于过参数化模型的随机梯度算法

(11)–(15)相比,KT-SG 算法(97)–(104)的参数估计维数要小,故计算量要小得多.读者可以对两个算法每递推计算一步需要的乘法次数和加法次数进行比较.

3.2 基于关键项分离的多新息随机梯度辨识算法

使用最新的 p 组数据,定义堆积输出向量 $Y(p,t)$ 以及堆积信息矩阵 $\hat{\Phi}(p,t)$ 为

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

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

设整数 $p \geq 1$ 表示新息长度.将 KT-SG 算法(97)–(104)中的标量信息

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

扩展为新息向量

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

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

扩展输出 $y(t)$ 为堆积输出向量 $Y(p,t)$,扩展信息向量 $\hat{\varphi}(t)$ 为堆积信息矩阵 $\hat{\Phi}(p,t)$,便得到辨识 IN-EE 系统参数向量 ϑ 的基于关键项分离的多新息随机梯度算法(Key Term separation based Multi-Innovation Stochastic Gradient algorithm,KT-MISG 算法):

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \frac{\hat{\Phi}(p,t)}{r(t)} E(p,t), \quad (105)$$

$$E(p,t) = Y(p,t) - \hat{\Phi}^T(p,t) \hat{\vartheta}(t-1), \quad (106)$$

$$r(t) = r(t-1) + \|\hat{\varphi}(t)\|^2, \quad r(0) = 1, \quad (107)$$

$$Y(p,t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (108)$$

$$\hat{\Phi}(p,t) = [\hat{\varphi}(t), \hat{\varphi}(t-1), \dots, \hat{\varphi}(t-p+1)], \quad (109)$$

$$\hat{\varphi}(t) = [\varphi_a^T(t), \hat{\varphi}_b^T(t), f^T(t)]^T, \quad (110)$$

$$\varphi_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (111)$$

$$\hat{\varphi}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (112)$$

$$f(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (113)$$

$$\hat{u}(t) = f^T(t) \hat{c}(t), \quad (114)$$

$$\hat{\vartheta}(t) = [\hat{a}^T(t), \hat{b}^T(t), \hat{c}^T(t)]^T. \quad (115)$$

当新息长度 $p = 1$ 时,KT-MISG 算法退化为 KT-SG 算法(97)–(104),可以说 KT-SG 算法是 KT-MISG 算法的特例.为使算法有更快的收敛速度,式(107)中的 $r(t)$ 也可修改为

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

KT-MISG 算法(105)–(115)的实现步骤如下:

1) 初始化: $t = 1$, 给定数据长度 L 、新息长度 p 和

基函数 $f_j(\cdot)$; 令 $\hat{\vartheta}(0) = \mathbf{1}_{n_0}/p_0$, $r(0) = 1$, $\hat{u}(j) = 1/p_0$, $j \leq 0$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式(108)构造 $Y(p,t)$.

3) 用式(111)–(113)构造 $\varphi_a(t)$, $\hat{\varphi}_b(t)$ 和 $f(t)$, 用式(110)和(109)构造信息向量 $\hat{\varphi}(t)$ 和堆积信息矩阵 $\hat{\Phi}(p,t)$.

4) 用式(106)计算新息向量 $E(p,t)$, 用式(107)计算 $r(t)$.

5) 用式(105)刷新 $\hat{\vartheta}(t)$.

6) 从式(115)的 $\hat{\vartheta}(t)$ 中读取 $\hat{c}(t)$, 用式(114)计算中间变量 $\hat{u}(t)$.

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

KT-MISG 算法(105)–(115)计算 IN-EE 系统参数估计向量 $\hat{\vartheta}(t)$ 的流程如图 7 所示.

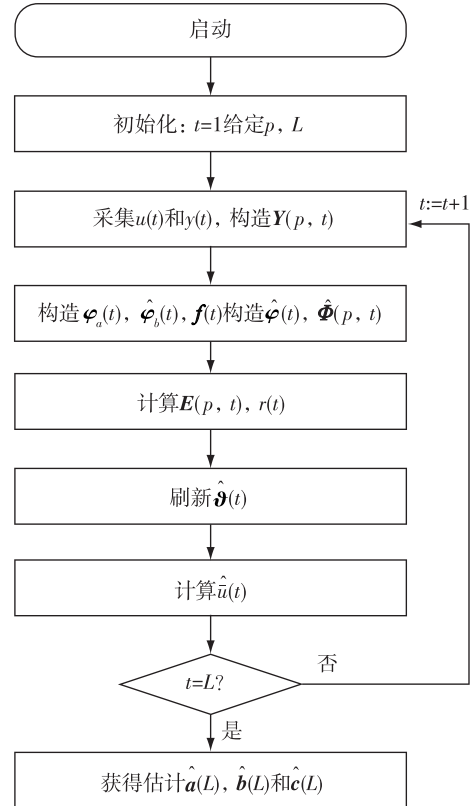


图 7 基于关键项分离的多新息随机梯度算法流程

Fig. 7 The flowchart of the KT-MISG algorithm

注 13 与 KT-SG 算法(97)–(104)相比,KT-MISG 算法(105)–(115)通过新息扩展,由于充分地利用了系统的观测数据和辨识新息,故算法的收敛速度加快,辨识精度提高,不过运算量有所增大.

3.3 基于关键项分离的递推最小二乘辨识算法

针对 IN-EE 系统的关键项分离辨识模型(94), 定义二次准则函数:

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

极小化这个二次准则函数或令 $J(\boldsymbol{\vartheta})$ 对 $\boldsymbol{\vartheta}$ 的偏导数为零, 即:

$$\frac{\partial J(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = - \sum_{j=1}^t \boldsymbol{\varphi}(j) [y(j) - \boldsymbol{\varphi}^T(j)\boldsymbol{\vartheta}] = \mathbf{0},$$

在持续激励条件下, 数据长度 t 足够大, 相关矩阵正定时, 可获得参数向量 $\boldsymbol{\vartheta}$ 的最小二乘估计为

$$\hat{\boldsymbol{\vartheta}}(t) = \left[\sum_{j=1}^t \boldsymbol{\varphi}(j)\boldsymbol{\varphi}^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}(j)y(j), \quad (116)$$

进一步可以得到递推最小二乘估计算法:

$$\boldsymbol{\vartheta}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{L}(t) [y(t) - \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (117)$$

$$\mathbf{L}(t) = \frac{\mathbf{P}(t-1)\boldsymbol{\varphi}(t)}{1 + \boldsymbol{\varphi}^T(t)\mathbf{P}(t-1)\boldsymbol{\varphi}(t)}, \quad (118)$$

$$\mathbf{P}(t) = [\mathbf{I}_{n_0} - \mathbf{L}(t)\boldsymbol{\varphi}^T(t)]\mathbf{P}(t-1). \quad (119)$$

式(117)–(119)的递推最小二乘估计算法不可实现, 因为信息向量 $\boldsymbol{\varphi}(t)$ 中的 $\boldsymbol{\varphi}_b(t)$ 含有未知变量 $\bar{u}(t-i)$. 解决方法是未知量 $\bar{u}(t)$, $\boldsymbol{\varphi}_b(t)$ 和 $\boldsymbol{\varphi}(t)$ 均用其估计 $\hat{u}(t)$, $\hat{\boldsymbol{\varphi}}_b(t)$ 和 $\hat{\boldsymbol{\varphi}}(t)$ 代替, 可以得到辨识 IN-EE 系统参数向量 $\boldsymbol{\vartheta}$ 的基于关键项分离的递推最小二乘算法 (Key Term separation based Recursive Least Squares algorithm, KT-RLS 算法)^[41]:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{L}(t) [y(t) - \hat{\boldsymbol{\varphi}}^T(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (120)$$

$$\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{n_0}/p_0,$$

$$\mathbf{L}(t) = \mathbf{P}(t-1)\hat{\boldsymbol{\varphi}}(t) [1 + \hat{\boldsymbol{\varphi}}^T(t)\mathbf{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \quad (121)$$

$$\mathbf{P}(t) = [\mathbf{I}_{n_0} - \mathbf{L}(t)\hat{\boldsymbol{\varphi}}^T(t)]\mathbf{P}(t-1), \quad \mathbf{P}(0) = p_0\mathbf{I}_{n_0}, \quad (122)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \mathbf{f}^T(t)]^T, \quad (123)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (124)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (125)$$

$$\mathbf{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (126)$$

$$\hat{u}(t) = \mathbf{f}^T(t)\hat{\boldsymbol{\varepsilon}}(t), \quad (127)$$

$$\hat{\boldsymbol{\varepsilon}}(t) = [\hat{\boldsymbol{a}}^T(t), \hat{\boldsymbol{b}}^T(t), \hat{\boldsymbol{\varepsilon}}^T(t)]^T. \quad (128)$$

表 5 列出了 KT-RLS 算法每步的计算量 ($n_0 := n_a + n_b + m$).

3.4 基于关键项分离的多新息最小二乘辨识算法

定义堆积输出向量 $\mathbf{Y}(p, t)$ 和堆积信息矩阵

$\hat{\boldsymbol{\Phi}}(p, t)$ 如下:

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

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

借助多新息辨识理论^[14], 将式(120)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

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

扩展 KT-RLS 算法(120)–(128)中输出 $y(t)$ 为堆积输出向量 $\mathbf{Y}(p, t)$, 信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 为堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p, t)$, 可以得到辨识 IN-EE 系统参数向量 $\boldsymbol{\vartheta}$ 的基于关键项分离的多新息最小二乘算法 (Key Term separation based Multi-Innovation Least Squares algorithm, KT-MILS 算法):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{L}(t)\mathbf{E}(p, t), \quad (129)$$

$$\mathbf{E}(p, t) = \mathbf{Y}(p, t) - \hat{\boldsymbol{\Phi}}^T(p, t)\hat{\boldsymbol{\vartheta}}(t-1), \quad (130)$$

$$\mathbf{L}(t) = \mathbf{P}(t-1)\hat{\boldsymbol{\Phi}}(p, t) [\mathbf{I}_p + \hat{\boldsymbol{\Phi}}^T(p, t)\mathbf{P}(t-1)\hat{\boldsymbol{\Phi}}(p, t)]^{-1}, \quad (131)$$

$$\mathbf{P}(t) = [\mathbf{I}_{n_0} - \mathbf{L}(t)\hat{\boldsymbol{\Phi}}^T(p, t)]\mathbf{P}(t-1), \quad (132)$$

$$\mathbf{Y}(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (133)$$

$$\hat{\boldsymbol{\Phi}}(p, t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \dots, \hat{\boldsymbol{\varphi}}(t-p+1)], \quad (134)$$

$$\hat{\boldsymbol{\varphi}}(t) = [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \mathbf{f}^T(t)]^T, \quad (135)$$

表 5 基于关键项分离的递推最小二乘算法的计算量

Table 5 The computational efficiency of the KT-RLS algorithm

变量	计算步骤	乘法次数	加法次数
$\hat{\boldsymbol{\vartheta}}(t)$	$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{L}(t)e(t) \in \mathbf{R}^{n_0}$	n_0	n_0
	$e(t) := y(t) - \hat{\boldsymbol{\varphi}}^T(t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbf{R}$	n_0	n_0
$\mathbf{L}(t)$	$\mathbf{L}(t) = \boldsymbol{\zeta}(t) / [1 + \hat{\boldsymbol{\varphi}}^T(t)\boldsymbol{\zeta}(t)] \in \mathbf{R}^{n_0}$	$2n_0$	n_0
	$\boldsymbol{\zeta}(t) := \mathbf{P}(t-1)\hat{\boldsymbol{\varphi}}(t) \in \mathbf{R}^{n_0}$	n_0^2	$(n_0 - 1)n_0$
$\mathbf{P}(t)$	$\mathbf{P}(t) = \mathbf{P}(t-1) - \mathbf{L}(t)\boldsymbol{\zeta}^T(t) \in \mathbf{R}^{n_0 \times n_0}$	n_0^2	n_0^2
$\hat{u}(t)$	$\hat{u}(t) = \mathbf{f}^T(t)\hat{\boldsymbol{\varepsilon}}(t) \in \mathbf{R}$	m	$m - 1$
总数		$2n_0^2 + 4n_0 + m$	$2n_0^2 + 2n_0 + m - 1$
总 Flop 数		$N_S := 4n_0^2 + 6n_0 + 2m - 1$	

$$\varphi_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (136)$$

$$\hat{\varphi}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (137)$$

$$f(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (138)$$

$$\hat{u}(t) = f^T(t) \hat{c}(t), \quad (139)$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{a}}^T(t), \hat{\boldsymbol{b}}^T(t), \hat{\boldsymbol{c}}^T(t)]^T. \quad (140)$$

当新息长度 $p=1$ 时, KT-MILS 算法退化为 KT-RLS 算法 (120) — (128), 可以说 KT-RLS 算法是 KT-MILS 算法的特例, 是新息长度 $p=1$ 时的 KT-MILS 算法. KT-MILS 算法 (129) — (140) 的计算步骤如下:

1) 初始化: 令 $t=1$, 给定信息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$; 设置 $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{n_0}/p_0$, $\boldsymbol{P}(0) = p_0 \boldsymbol{I}_{n_0}$, $\hat{u}(j) = 1/p_0, j \leq 0, p_0 = 10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式 (133) 构造 $\boldsymbol{Y}(p, t)$.

3) 用式 (136) — (138) 构造 $\varphi_a(t)$, $\hat{\varphi}_b(t)$ 和 $f(t)$, 用 (135) 和 (134) 构造 $\hat{\varphi}(t)$ 和 $\hat{\boldsymbol{\Phi}}(p, t)$.

4) 用式 (130) 计算新息向量 $\boldsymbol{E}(p, t)$, 用式 (131) — (132) 计算 $\boldsymbol{L}(t)$ 和 $\boldsymbol{P}(t)$.

5) 用式 (129) 刷新 $\hat{\boldsymbol{\theta}}(t)$.

6) 用式 (139) 计算中间变量 $\hat{u}(t)$.

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

利用 KT-MILS 算法 (129) — (140) 计算 IN-EE 系统参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 的流程如图 8 所示, 其每步的计算量如表 6 所示 ($n_0 := n_a + n_b + m$).

4 基于关键项分离的两阶段多新息辨识方法

对于输入非线性方程误差系统, 其关键项分离辨识模型 (94) 参数向量含 $n_a + n_b + m$ 个参数, 相对于过参数化方法来说, 待辨识参数个数减少了, 算法计

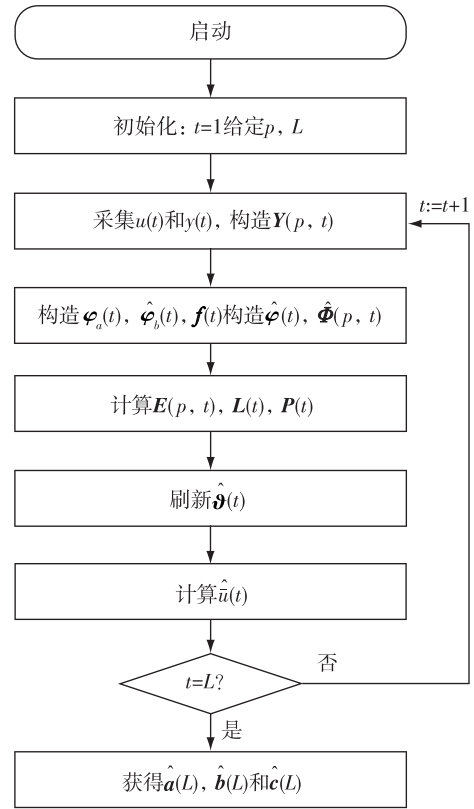


图 8 基于关键项分离的多新息最小二乘算法流程

Fig. 8 The flowchart of the KT-MILS algorithm

算量减小了, 辨识效率提高了. 若线性模块和非线性模块的参数过很多, 计算量还是很大的. 为了在保持辨识精度变化不大的条件下, 进一步提高辨识效率, 可以利用递阶辨识原理, 将关键项分离辨识模型 (94) 分解为 2 个或 3 个子系统, 研究提出相应的辨识方法.

为清晰起见, 将 IN-EE 系统的关键项分离辨识模型 (94) 重写如下:

$$y(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + v(t), \quad (141)$$

表 6 基于关键项分离的多新息最小二乘算法的计算量

Table 6 The computational efficiency of the KT-MILS algorithm

表达式	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{L}(t) \boldsymbol{E}(p, t) \in \mathbf{R}^{n_0}$	$n_0 p$	$n_0 p$
$\boldsymbol{E}(p, t) = \boldsymbol{Y}(p, t) - \boldsymbol{\Phi}^T(p, t) \hat{\boldsymbol{\theta}}(t-1) \in \mathbf{R}^p$	$n_0 p$	$n_0 p$
$\boldsymbol{L}(t) = \boldsymbol{\zeta}(t) [\boldsymbol{I}_p + \boldsymbol{\Phi}^T(p, t) \boldsymbol{\zeta}(t)]^{-1} \in \mathbf{R}^{n_0 \times p}$	$2n_0 p^2 + p^3$	$2n_0 p^2 + p^3 - n_0 p - 2p^2$
$\boldsymbol{\zeta}(t) := \boldsymbol{P}(t-1) \boldsymbol{\Phi}(p, t) \in \mathbf{R}^{n_0 \times p}$	$n_0^2 p$	$n_0^2 p - n_0 p$
$\boldsymbol{P}(t) = \boldsymbol{P}(t-1) - \boldsymbol{L}(t) \boldsymbol{\zeta}^T(t) \in \mathbf{R}^{n_0 \times n_0}$	$n_0^2 p$	$n_0^2 p$
$\hat{u}(t) = f^T(t) \hat{\boldsymbol{c}}(t) \in \mathbf{R}$	m	$m-1$
总数	$2n_0 p(n_0 + p + 1) + p^3 + m$	$2n_0 p(n_0 + p) + p^3 - 2p^2 + m - 1$
总 flops 数	$N_6 := 4n_0^2 p + 4n_0 p^2 + 2n_0 p + 2p^3 - 2p^2 + 2m - 1$	

其中参数向量和信息向量定义为

$$\boldsymbol{\vartheta} := [\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T]^T \in \mathbf{R}^{n_0}, \quad n_0 := n_a + n_b + m,$$

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

$$\mathbf{b} := [b_1, b_2, \dots, b_{n_b}]^T \in \mathbf{R}^{n_b},$$

$$\mathbf{c} := [c_1, c_2, \dots, c_m]^T \in \mathbf{R}^m,$$

$$\boldsymbol{\varphi}(t) := [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \mathbf{f}^T(t)]^T \in \mathbf{R}^{n_0},$$

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

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

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

根据参数向量 $\boldsymbol{\vartheta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 中的定义,式(141)可等价表示为

$$y(t) = \boldsymbol{\varphi}_a^T(t) \mathbf{a} + \boldsymbol{\varphi}_b^T(t) \mathbf{b} + \mathbf{f}^T(t) \mathbf{c} + v(t). \quad (142)$$

式(142)包含了3个子参数向量 \mathbf{a} , \mathbf{b} 和 \mathbf{c} . 可以把任意2个参数向量合在一起,形成一个两阶段的参数线性化模型. 例如,把线性部分参数向量 \mathbf{a} 和 \mathbf{b} 合在一起,则模型(141)也可等价表示为的两阶段参数线性化模型:

$$y(t) = \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta} + \mathbf{f}^T(t) \mathbf{c} + v(t), \quad (143)$$

$$\boldsymbol{\theta} := \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \in \mathbf{R}^{n_a+n_b},$$

$$\boldsymbol{\varphi}_1(t) := \begin{bmatrix} \boldsymbol{\varphi}_a(t) \\ \boldsymbol{\varphi}_b(t) \end{bmatrix} \in \mathbf{R}^{n_a+n_b}.$$

下面介绍基于关键项分离的两阶段随机梯度辨识方法、两阶段多新息随机梯度辨识方法、两阶段递推最小二乘辨识方法、两阶段多新息最小二乘辨识方法.

4.1 基于关键项分离的两阶段随机梯度辨识算法

针对辨识模型(143),定义2个中间变量:

$$y_1(t) := y(t) - \mathbf{f}^T(t) \mathbf{c},$$

$$y_2(t) := y(t) - \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta}.$$

模型(143)可分解为2个子系统:

$$y_1(t) = \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta} + v(t), \quad (144)$$

$$y_2(t) = \mathbf{f}^T(t) \mathbf{c} + v(t). \quad (145)$$

定义2个梯度准则函数:

$$J_5(\boldsymbol{\theta}) := \frac{1}{2} [y_1(t) - \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta}]^2,$$

$$J_6(\mathbf{c}) := \frac{1}{2} [y_2(t) - \mathbf{f}^T(t) \mathbf{c}]^2.$$

利用负梯度搜索原理极小化这2个准则函数,可得辨识子系统(144)和(145)参数向量 $\boldsymbol{\theta}$ 和 \mathbf{c} 的随机梯度算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}_1(t)}{r_1(t)} [y_1(t) - \boldsymbol{\varphi}_1^T(t) \hat{\boldsymbol{\theta}}(t-1)] =$$

$$\hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}_1(t)}{r_1(t)} [y(t) - \mathbf{f}^T(t) \mathbf{c} - \boldsymbol{\varphi}_1^T(t) \hat{\boldsymbol{\theta}}(t-1)], \quad (146)$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\varphi}_1(t)\|^2, \quad r_1(0) = 1, \quad (147)$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \frac{\mathbf{f}(t)}{r_2(t)} [y_2(t) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1)] =$$

$$\hat{\mathbf{c}}(t-1) + \frac{\mathbf{f}(t)}{r_2(t)} [y(t) - \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta} - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1)], \quad (148)$$

$$r_2(t) = r_2(t-1) + \|\mathbf{f}(t)\|^2, \quad r_2(0) = 1. \quad (149)$$

注意到算法(146)~(149)右边包含了未知参数向量 \mathbf{c} 和 $\boldsymbol{\theta}$, 而且信息向量 $\boldsymbol{\varphi}_1(t)$ 包含了未知变量 $\bar{u}(t-i)$, 因此算法不可实现. 借助递阶辨识原理, 这些未知变量可用辅助模型的输出代替, 也可用其估计代替, 细节如下.

$$\text{设 } \hat{\boldsymbol{\theta}}(t) := \begin{bmatrix} \hat{\mathbf{a}}(t) \\ \hat{\mathbf{b}}(t) \end{bmatrix} \in \mathbf{R}^{n_a+n_b} \text{ 是参数向量 } \boldsymbol{\theta} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \text{ 在}$$

时刻 t 的估计, $\hat{\mathbf{c}}(t) \in \mathbf{R}^m$ 是参数向量 \mathbf{c} 在时刻 t 的估计. 根据式(92), $\bar{u}(t)$ 估计为

$$\hat{\bar{u}}(t) = \sum_{j=1}^m \hat{c}_j(t) f_j(u(t)) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t). \quad (150)$$

构造 $\boldsymbol{\varphi}_1(t)$ 的估计为

$$\hat{\boldsymbol{\varphi}}_1(t) := \begin{bmatrix} \boldsymbol{\varphi}_a(t) \\ \hat{\boldsymbol{\varphi}}_b(t) \end{bmatrix} \in \mathbf{R}^{n_a+n_b},$$

$$\hat{\boldsymbol{\varphi}}_b(t) := [\hat{\bar{u}}(t-1), \hat{\bar{u}}(t-2), \dots, \hat{\bar{u}}(t-n_b)]^T \in \mathbf{R}^{n_b}.$$

将算法(146)~(149)中的未知 $\boldsymbol{\varphi}_1(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}_1(t)$ 代替, 未知向量 \mathbf{c} 和 $\boldsymbol{\theta}$ 用其在时刻 $(t-1)$ 的估计 $\hat{\mathbf{c}}(t-1)$ 和 $\hat{\boldsymbol{\theta}}(t-1)$ 代替, 便可以得到辨识 IN-EE 系统参数向量的基于关键项分离的两阶段随机梯度算法 (Key Term separation based Two-Stage Stochastic Gradient algorithm, KT-2S-SG 算法):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}_1(t)}{r_1(t)} [y(t) - \hat{\boldsymbol{\varphi}}_1^T(t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1)], \quad (151)$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\varphi}}_1(t)\|^2 = r_1(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2 + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2, \quad r_1(0) = 1, \quad (152)$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \frac{\mathbf{f}(t)}{r_2(t)} [y(t) - \hat{\boldsymbol{\varphi}}_1^T(t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1)], \quad (153)$$

$$r_2(t) = r_2(t-1) + \|\mathbf{f}(t)\|^2, \quad r_2(0) = 1, \quad (154)$$

$$\hat{\boldsymbol{\varphi}}_1(t) = \begin{bmatrix} \boldsymbol{\varphi}_a(t) \\ \hat{\boldsymbol{\varphi}}_b(t) \end{bmatrix}, \quad (155)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (156)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{\bar{u}}(t-1), \hat{\bar{u}}(t-2), \dots, \hat{\bar{u}}(t-n_b)]^T, \quad (157)$$

$$\mathbf{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (158)$$

$$\hat{\mathbf{u}}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t). \quad (159)$$

如果式(151)和(153)中 $r_1(t)$ 和 $r_2(t)$ 采用一个共同的 $r(t)$, 即:

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2 + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2 + \|\mathbf{f}(t)\|^2, \\ r(0) = 1, \quad (160)$$

则 KT-2S-SG 算法完全等同于 KT-SG 算法(97)—(104).

4.2 基于关键项分离的两阶段多新息随机梯度辨识算法

定义新息长度为 p 的堆积输出向量 $\mathbf{Y}(p, t)$ 及堆积信息矩阵 $\hat{\boldsymbol{\Phi}}_1(p, t)$ 和 $\mathbf{F}(p, t)$ 如下:

$$\mathbf{Y}(p, t) := [y(t), y(t-1), \dots, y(t-p+1)]^T \in \mathbf{R}^p, \\ \hat{\boldsymbol{\Phi}}_1(p, t) := [\hat{\boldsymbol{\varphi}}_1(t), \hat{\boldsymbol{\varphi}}_1(t-1), \dots, \hat{\boldsymbol{\varphi}}_1(t-p+1)] \in \mathbf{R}^{(n_a+n_b) \times p}, \\ \mathbf{F}(p, t) := [\mathbf{f}(t), \mathbf{f}(t-1), \dots, \mathbf{f}(t-p+1)] \in \mathbf{R}^{m \times p}.$$

将式(151)和(153)中的标量新息

$$e(t) := y(t) - \hat{\boldsymbol{\varphi}}_1^T(t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1) \in \mathbf{R}$$

扩展为新息向量

$$\mathbf{E}(p, t) := \begin{bmatrix} y(t) - \hat{\boldsymbol{\varphi}}_1^T(t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1) \\ y(t-1) - \hat{\boldsymbol{\varphi}}_1^T(t-1) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t-1) \hat{\mathbf{c}}(t-1) \\ \vdots \\ y(t-p+1) - \hat{\boldsymbol{\varphi}}_1^T(t-p+1) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t-p+1) \hat{\mathbf{c}}(t-1) \end{bmatrix} =$$

$$\mathbf{Y}(p, t) - \hat{\boldsymbol{\Phi}}_1(p, t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{F}(p, t) \hat{\mathbf{c}}(t-1) \in \mathbf{R}^p,$$

将 KT-2S-SG 算法(151)—(193)中的信息向量 $\hat{\boldsymbol{\varphi}}_1(t)$ 和 $\mathbf{f}(t)$ 扩展为堆积信息矩阵 $\hat{\boldsymbol{\Phi}}_1(p, t)$ 和 $\mathbf{F}(p, t)$, 可以得到辨识 IN-EE 系统参数向量的基于关键项分离的两阶段多新息随机梯度算法(Key Term separation based Two-Stage Multi-Innovation Stochastic Gradient algorithm, KT-2S-MISG 算法):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}_1(p, t)}{r_1(t)} \mathbf{E}(p, t), \quad (161)$$

$$\mathbf{E}(p, t) = \mathbf{Y}(p, t) - \hat{\boldsymbol{\Phi}}_1(p, t) \hat{\boldsymbol{\theta}}(t-1) - \mathbf{F}(p, t) \hat{\mathbf{c}}(t-1), \quad (162)$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2 + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2, \\ r_1(0) = 1, \quad (163)$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \frac{\mathbf{F}(p, t)}{r_2(t)} \mathbf{E}(p, t), \quad (164)$$

$$r_2(t) = r_2(t-1) + \|\mathbf{f}(t)\|^2, \quad r_2(0) = 1, \quad (165)$$

$$\mathbf{Y}(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (166)$$

$$\hat{\boldsymbol{\Phi}}_1(p, t) = [\hat{\boldsymbol{\varphi}}_1(t), \hat{\boldsymbol{\varphi}}_1(t-1), \dots, \hat{\boldsymbol{\varphi}}_1(t-p+1)], \quad (167)$$

$$\mathbf{F}(p, t) = [\mathbf{f}(t), \mathbf{f}(t-1), \dots, \mathbf{f}(t-p+1)], \quad (168)$$

$$\hat{\boldsymbol{\varphi}}_1(t) = \begin{bmatrix} \boldsymbol{\varphi}_a(t) \\ \hat{\boldsymbol{\varphi}}_b(t) \end{bmatrix}, \quad (169)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T \in \mathbf{R}^{n_a}, \quad (170)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (171)$$

$$\mathbf{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (172)$$

$$\hat{u}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t). \quad (173)$$

当新息长度 $p = 1$ 时, KT-2S-MISG 算法退化为 KT-2S-SG 算法(151)—(159). KT-2S-MISG 算法与 KT-MISG 算法只有细微的差别, 如果式(161)和(164)中 $r_1(t)$ 和 $r_2(t)$ 采用同一个 $r(t)$, 即:

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2 + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2 + \|\mathbf{f}(t)\|^2, \\ r(0) = 1, \quad (174)$$

则 KT-2S-MISG 算法完全等同于 KT-MISG 算法(105)—(115). 也可在式(174)中引入遗忘因子 λ 来加快收敛速度,

$$r(t) = \lambda r(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2 + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2 + \|\mathbf{f}(t)\|^2, \\ 0 \leq \lambda \leq 1, \quad r(0) = 1. \quad (175)$$

在收敛性分析中, $r(t)$ 也可取作

$$r(t) = \lambda r(t-1) + \|\boldsymbol{\Phi}_a(p, t)\|^2 + \|\hat{\boldsymbol{\Phi}}_b(p, t)\|^2 + \|\mathbf{F}(p, t)\|^2, \\ 0 \leq \lambda \leq 1, \quad r(0) = 1. \quad (176)$$

注 14 两阶段随机梯度类算法、多阶段随机梯度类算法与随机梯度类算法间的区别不是很大, 只是 $r(t)$ 的微小变化. 同样, 两阶段多新息随机梯度类算法、多阶段多新息随机梯度类算法与多新息随机梯度类算法间的区别也不是很大, 也只是在 $r(t)$ 的小变化上. 如果采用一个共同的 $r(t)$, 它们就完全一样了. 两阶段辨识算法、多阶段辨识算法、递阶辨识算法(两阶段、多阶段也是递阶的意思)的特征主要表现在最小二乘类算法上.

4.3 基于关键项分离的两阶段递推最小二乘辨识算法

针对模型(143)分解得到的子系统(144)和子系统(145), 即:

$$y_1(t) = \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta} + v(t), \quad (177)$$

$$y_2(t) = \mathbf{f}^T(t) \mathbf{c} + v(t), \quad (178)$$

其中

$$y_1(t) := y(t) - \mathbf{f}^T(t) \mathbf{c},$$

$$y_2(t) := y(t) - \boldsymbol{\varphi}_1^T(t) \boldsymbol{\theta}.$$

定义 2 个二次准则函数:

$$J_7(\boldsymbol{\theta}) := \frac{1}{2} \sum_{j=1}^t [y_1(j) - \boldsymbol{\varphi}_1^T(j) \boldsymbol{\theta}]^2,$$

$$J_8(\mathbf{c}) := \frac{1}{2} \sum_{j=1}^l [y_2(j) - \mathbf{f}^T(j)\mathbf{c}]^2.$$

极小化这 2 个准则函数或令它们的偏导数等于零, 可得:

$$\frac{\partial J_7(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2 \sum_{j=1}^l \boldsymbol{\varphi}_1(j) [y_1(j) - \boldsymbol{\varphi}_1^T(j)\boldsymbol{\theta}] = \mathbf{0},$$

$$\frac{\partial J_8(\mathbf{c})}{\partial \mathbf{c}} = -2 \sum_{j=1}^l \mathbf{f}(j) [y_2(j) - \mathbf{f}^T(j)\mathbf{c}] = \mathbf{0}.$$

对于大 t , 假设相关矩阵可逆, 可以求得

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1(t) &= \left[\sum_{j=1}^t \boldsymbol{\varphi}_1(j)\boldsymbol{\varphi}_1^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_1(j)y_1(j) = \\ & \left[\sum_{j=1}^t \boldsymbol{\varphi}_1(j)\boldsymbol{\varphi}_1^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_1(j) [y(j) - \mathbf{f}^T(j)\mathbf{c}], \\ \hat{\mathbf{c}}(t) &= \left[\sum_{j=1}^t \mathbf{f}(j)\mathbf{f}^T(j) \right]^{-1} \sum_{j=1}^t \mathbf{f}(j)y_2(j) = \\ & \left[\sum_{j=1}^t \mathbf{f}(j)\mathbf{f}^T(j) \right]^{-1} \sum_{j=1}^t \mathbf{f}(j) [y(j) - \boldsymbol{\varphi}_1^T(j)\boldsymbol{\theta}]. \end{aligned}$$

上述表达式右边的未知量 $\boldsymbol{\varphi}_1(t)$, \mathbf{c} 和 $\boldsymbol{\theta}$ 分别用其估计 $\hat{\boldsymbol{\varphi}}_1(t)$, $\hat{\mathbf{c}}(t-1)$ 和 $\hat{\boldsymbol{\theta}}(t-1)$ 代替, 得到基于关键项分离的两阶段最小二乘算法 (KT-2S-LS 算法):

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1(t) &= \left[\sum_{j=1}^t \hat{\boldsymbol{\varphi}}_1(j)\hat{\boldsymbol{\varphi}}_1^T(j) \right]^{-1} \sum_{j=1}^t \hat{\boldsymbol{\varphi}}_1(j) [y(j) - \mathbf{f}^T(j)\hat{\mathbf{c}}(t-1)], \\ \hat{\mathbf{c}}(t) &= \left[\sum_{j=1}^t \mathbf{f}(j)\mathbf{f}^T(j) \right]^{-1} \sum_{j=1}^t \mathbf{f}(j) [y(j) - \hat{\boldsymbol{\varphi}}_1^T(j)\hat{\boldsymbol{\theta}}(t-1)], \\ \hat{\boldsymbol{\varphi}}_1(t) &= [-y(t-1), -y(t-2), \dots, -y(t-n_a), \hat{u}(t-1), \\ & \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \\ \hat{u}(t) &= \mathbf{f}^T(u(t))\hat{\mathbf{c}}(t). \end{aligned}$$

为了避免批数据处理算法的矩阵求逆运算, 参考

文献[1,39]中的方法, 能够得到辨识 IN-EE 系统参数向量的基于关键项分离的两阶段递推最小二乘算法 (Key Term separation based Two-Stage Recursive Least Squares algorithm, KT-2S-RLS 算法)^[41]:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}_1(t) [y(t) - \hat{\boldsymbol{\varphi}}_1^T(t)\hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t)\hat{\mathbf{c}}(t-1)], \quad (179)$$

$$\mathbf{L}_1(t) = \mathbf{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t) [1 + \hat{\boldsymbol{\varphi}}_1^T(t)\mathbf{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t)]^{-1}, \quad (180)$$

$$\mathbf{P}_1(t) = [\mathbf{I}_{n_a+n_b} - \mathbf{L}_1(t)\hat{\boldsymbol{\varphi}}_1^T(t)]\mathbf{P}_1(t-1), \quad (181)$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \mathbf{L}_2(t) [y(t) - \hat{\boldsymbol{\varphi}}_1^T(t)\hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t)\hat{\mathbf{c}}(t-1)], \quad (182)$$

$$\mathbf{L}_2(t) = \mathbf{P}_2(t-1)\mathbf{f}(t) [1 + \mathbf{f}^T(t)\mathbf{P}_2(t-1)\mathbf{f}(t)]^{-1}, \quad (183)$$

$$\mathbf{P}_2(t) = [\mathbf{I}_m - \mathbf{L}_2(t)\mathbf{f}^T(t)]\mathbf{P}_2(t-1), \quad (184)$$

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_1(t) &= [-y(t-1), -y(t-2), \dots, -y(t-n_a), \\ & \hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (185) \end{aligned}$$

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

$$\hat{u}(t) = \mathbf{f}^T(t)\hat{\mathbf{c}}(t). \quad (187)$$

算法的初值可以取为

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{n_a+n_b}/p_0, \quad \hat{\mathbf{c}}(0) = \mathbf{1}_m/p_0, \quad \hat{u}(i) = 1/p_0, \quad i \leq 0,$$

$$\mathbf{P}_1(0) = p_0\mathbf{I}_{n_a+n_b}, \quad \mathbf{P}_2(0) = p_0\mathbf{I}_m, \quad p_0 = 10^6.$$

表 7 列出了 KT-2S-RLS 算法每步的计算量 ($n_1 := n_a + n_b$).

注 15 最小二乘算法的计算量是 $O(n^2)$ (n 是参数向量维数, 即协方差阵的维数). KT-2S-RLS 算法通过分解为 2 个子系统, 尽管要计算 2 个协方差阵, 其计算量也比 KT-RLS 算法 (120) — (128) 小, 因为当 $n = n_1 + n_2$ 时, $n_1^2 + n_2^2 < n^2$.

表 7 基于关键项分离的两阶段递推最小二乘算法的计算量

Table 7 The computational efficiency of the KT-2S-RLS algorithm

变量	计算步骤	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}_1(t)e(t) \in \mathbf{R}^{n_1}$	n_1	n_1
	$e(t) := y(t) - \hat{\boldsymbol{\varphi}}_1^T(t)\hat{\boldsymbol{\theta}}(t-1) - \mathbf{f}^T(t)\hat{\mathbf{c}}(t-1) \in \mathbf{R}$	$n_1 + m$	$n_1 + m$
$\mathbf{L}_1(t)$	$\mathbf{L}_1(t) = \boldsymbol{\zeta}_1(t) / [1 + \hat{\boldsymbol{\varphi}}_1^T(t)\boldsymbol{\zeta}_1(t)] \in \mathbf{R}^{n_1}$	$2n_1$	n_1
	$\boldsymbol{\zeta}_1(t) := \mathbf{P}_1(t-1)\hat{\boldsymbol{\varphi}}_1(t) \in \mathbf{R}^{n_1}$	n_1^2	$n_1^2 - n_1$
$\mathbf{P}_1(t)$	$\mathbf{P}_1(t) = \mathbf{P}_1(t-1) - \mathbf{L}_1(t)\boldsymbol{\zeta}_1^T(t) \in \mathbf{R}^{n_1 \times n_1}$	n_1^2	n_1^2
$\hat{\mathbf{c}}(t)$	$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \mathbf{L}_2(t)e(t) \in \mathbf{R}^m$	m	m
	$\mathbf{L}_2(t) = \boldsymbol{\zeta}_2(t) / [1 + \mathbf{f}^T(t)\boldsymbol{\zeta}_2(t)] \in \mathbf{R}^m$	$2m$	m
$\mathbf{P}_2(t)$	$\boldsymbol{\zeta}_2(t) := \mathbf{P}_2(t-1)\mathbf{f}(t) \in \mathbf{R}^m$	m^2	$m^2 - m$
	$\mathbf{P}_2(t) = \mathbf{P}_2(t-1) - \mathbf{L}_2(t)\boldsymbol{\zeta}_2^T(t) \in \mathbf{R}^{m \times m}$	m^2	m^2
$\hat{u}(t)$	$\hat{u}(t) = \mathbf{f}^T(t)\hat{\mathbf{c}}(t) \in \mathbf{R}$	m	$m-1$
总数		$2n_1^2 + 2m^2 + 4n_1 + 5m$	$2n_1^2 + 2m^2 + 2n_1 + 3m - 1$
总 Flop 数		$N_7 := 4n_1^2 + 4m^2 + 6n_1 + 8m - 1$	

4.4 基于关键项分离的两阶段多新息最小二乘辨识算法

定义新息长度为 p 的堆积输出向量 $Y(p, t)$ 及堆积信息矩阵 $\hat{\Phi}_1(p, t)$ 和 $F(p, t)$ 如下:

$$\begin{aligned} Y(p, t) &= [y(t), y(t-1), \dots, y(t-p+1)]^T \in \mathbf{R}^p, \\ \hat{\Phi}_1(p, t) &= [\hat{\varphi}_1(t), \hat{\varphi}_1(t-1), \dots, \hat{\varphi}_1(t-p+1)] \in \mathbf{R}^{(n_a+n_b) \times p}, \\ F(p, t) &= [f(t), f(t-1), \dots, f(t-p+1)] \in \mathbf{R}^{m \times p}. \end{aligned}$$

将式(179)和(182)中的标量新息

$$e(t) := y(t) - \hat{\varphi}_1^T(t) \hat{\theta}(t-1) - f^T(t) \hat{c}(t-1) \in \mathbf{R}$$

扩展为新息向量

$$E(p, t) := Y(p, t) - \hat{\Phi}_1(p, t) \hat{\theta}(t-1) - F(p, t) \hat{c}(t-1) \in \mathbf{R}^p,$$

将KT-2S-RLS算法(179)–(187)中的信息向量 $\hat{\varphi}_1(t)$ 和 $f(t)$ 扩展为堆积信息矩阵 $\hat{\Phi}_1(p, t)$ 和 $F(p, t)$, 可以得到辨识IN-EE系统参数向量的基于关键项分离的两阶段多新息最小二乘算法(Key Term Separation based Two-Stage Multi-Innovation Least Squares algorithm, KT-2S-MILS算法):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L_1(t) E(p, t), \quad \hat{\theta}(0) = \mathbf{1}_{n_a+n_b}/p_0, \quad (188)$$

$$E(p, t) = Y(p, t) - \hat{\Phi}_1(p, t) \hat{\theta}(t-1) - F(p, t) \hat{c}(t-1), \quad (189)$$

$$\begin{aligned} L_1(t) &= P_1(t-1) \hat{\Phi}_1^T(p, t) [I_p + \\ &\quad \hat{\Phi}_1^T(p, t) P_1(t-1) \hat{\Phi}_1(p, t)]^{-1}, \quad (190) \end{aligned}$$

$$\begin{aligned} P_1(t) &= [I_{n_a+n_b} - L_1(t) \hat{\Phi}_1^T(p, t)] P_1(t-1), \\ P_1(0) &= p_0 I_{n_a+n_b}, \quad (191) \end{aligned}$$

$$\hat{c}(t) = \hat{c}(t-1) + L_2(t) E(p, t), \quad \hat{c}(0) = \mathbf{1}_m/p_0 \quad (192)$$

$$L_2(t) = P_2(t-1) F(p, t) [I_p + F^T(p, t) P_2(t-1) F(p, t)]^{-1}, \quad (193)$$

$$\begin{aligned} P_2(t) &= [I_m - L_2(t) F^T(p, t)] P_2(t-1), \\ P_2(0) &= p_0 I_m, \quad (194) \end{aligned}$$

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (195)$$

$$\hat{\Phi}_1(p, t) = [\hat{\varphi}_1(t), \hat{\varphi}_1(t-1), \dots, \hat{\varphi}_1(t-p+1)], \quad (196)$$

$$F(p, t) = [f(t), f(t-1), \dots, f(t-p+1)], \quad (197)$$

$$\begin{aligned} \hat{\varphi}_1(t) &= [-y(t-1), -y(t-2), \dots, -y(t-n_a), \hat{u}(t-1), \\ &\quad \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (198) \end{aligned}$$

$$f(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (199)$$

$$\hat{u}(t) = f^T(t) \hat{c}(t). \quad (200)$$

当新息长度 $p=1$ 时, KT-2S-MILS 算法退化为 KT-2S-RLS 算法(179)–(187). KT-2S-MILS 算法每一步的计算量如表8所示($n_1 := n_a + n_b$).

5 基于关键项分离的三阶段多新息辨识方法

上节将IN-EE系统分解为2个子系统, 讨论了基于关键项分离的辨识算法, 在一定程度上提高了计算效率. 为进一步提高计算效率, 利用递阶辨识原理, 将IN-EE系统的关键项分离辨识模型(94)分解为3个子系统, 一个包含线性块自回归部分 $A(z)$ 的参数向量 a , 一个包括线性块滑动平均部分 $B(z)$ 的参数向量 b , 一个包括非线性部分的参数向量 c , 讨论了相应的辨识方法.

将IN-EE系统的关键项分离辨识模型(94)重写如下:

$$y(t) = \varphi^T(t) \vartheta + v(t), \quad (201)$$

其中参数向量和信息向量定义为

$$\vartheta := [a^T, b^T, c^T]^T \in \mathbf{R}^{n_0}, \quad n_0 := n_a + n_b + m,$$

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

$$b := [b_1, b_2, \dots, b_{n_b}]^T \in \mathbf{R}^{n_b},$$

表8 KT-2S-MILS算法的计算量

Table 8 The computational efficiency of the KT-2S-MILS algorithm

表达式	乘法次数	加法次数
$\hat{\theta}(t) = \hat{\theta}(t-1) + L_1(t) E(p, t) \in \mathbf{R}^{n_1}$	$n_1 p$	$n_1 p$
$E(p, t) = Y(p, t) - \hat{\Phi}_1(p, t) \hat{\theta}(t-1) - F(p, t) \hat{c}(t-1) \in \mathbf{R}^p$	$n_1 p + mp$	$n_1 p + mp$
$L_1(t) = \zeta_1(t) [I_p + \hat{\Phi}_1^T(p, t) \zeta_1(t)]^{-1} \in \mathbf{R}^{n_1 \times p}$	$2n_1 p^2 + p^3$	$2n_1 p^2 + p^3 - n_1 p - 2p^2$
$\zeta_1(t) := P_1(t-1) \hat{\Phi}_1(p, t) \in \mathbf{R}^{n_1 \times p}$	$n_1^2 p$	$n_1^2 p - n_1 p$
$P_1(t) = P_1(t-1) - L_1(t) \zeta_1^T(t) \in \mathbf{R}^{n_1 \times n_1}$	$n_1^2 p$	$n_1^2 p$
$\hat{c}(t) = \hat{c}(t-1) + L_2(t) E(p, t) \in \mathbf{R}^m$	mp	mp
$L_2(t) = \zeta_2(t) [I_p + n F^T(p, t) \zeta_2(t)]^{-1} \in \mathbf{R}^{m \times p}$	$2mp^2 + p^3$	$2mp^2 + p^3 - mp - 2p^2$
$\zeta_2(t) := P_2(t-1) F(p, t) \in \mathbf{R}^{m \times p}$	$m^2 p$	$m^2 p - mp$
$P_2(t) = P_2(t-1) - L_2(t) \zeta_2^T(t) \in \mathbf{R}^{m \times m}$	$m^2 p$	$m^2 p$
$\hat{u}(t) = f^T(t) \hat{c}(t) \in \mathbf{R}$	m	$m-1$
总数	$2n_1^2 p + 2n_1 p^2 + 2n_1 p + 2p^3 +$ $2m^2 p + 2mp^2 + 2mp + m$	$2n_1^2 p + 2n_1 p^2 + 2p^3 +$ $2m^2 p + 2mp^2 - 4p^2 + m - 1$
总 flop 数	$N_8 := 4n_1^2 p + 4m^2 p + 4(n_1 + m)p^2 + 2n_0 p + 4p^3 - 4p^2 + 2m - 1$	

$$\begin{aligned} \boldsymbol{c} &:= [c_1, c_2, \dots, c_m]^T \in \mathbf{R}^m, \\ \boldsymbol{\varphi}(t) &:= [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \boldsymbol{f}^T(t)]^T \in \mathbf{R}^{n_0}, \\ \boldsymbol{\varphi}_a(t) &:= [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T \in \mathbf{R}^{n_a}, \\ \boldsymbol{\varphi}_b(t) &:= [\bar{u}(t-1), \bar{u}(t-2), \dots, \bar{u}(t-n_b)]^T \in \mathbf{R}^{n_b}, \\ \boldsymbol{f}(t) &:= [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T \in \mathbf{R}^m. \end{aligned}$$

根据参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 中的定义,式(201)可等价表示为

$$y(t) = \boldsymbol{\varphi}_a^T(t) \boldsymbol{a} + \boldsymbol{\varphi}_b^T(t) \boldsymbol{b} + \boldsymbol{f}^T(t) \boldsymbol{c} + v(t). \quad (202)$$

式(202)包含了3个子参数向量 \boldsymbol{a} , \boldsymbol{b} 和 \boldsymbol{c} . 下面讨论基于关键项分离的三阶段随机梯度辨识方法、三阶段多新息随机梯度辨识方法、三阶段递推最小二乘辨识方法、三阶段多新息最小二乘辨识方法.

5.1 基于关键项分离的三阶段随机梯度辨识算法

以上给出了 IN-CAR 系统两阶段参数线性化关键项分离模型的递阶辨识算法,下面针对三段参数线性化模型(202)给出相应的辨识算法.

定义3个中间变量

$$y_a(t) := y(t) - \boldsymbol{\varphi}_b^T(t) \boldsymbol{b} - \boldsymbol{f}^T(t) \boldsymbol{c},$$

$$y_b(t) := y(t) - \boldsymbol{\varphi}_a^T(t) \boldsymbol{a} - \boldsymbol{f}^T(t) \boldsymbol{c},$$

$$y_c(t) := y(t) - \boldsymbol{\varphi}_a^T(t) \boldsymbol{a} - \boldsymbol{\varphi}_b^T(t) \boldsymbol{b}.$$

于是,系统(201)可以分解为3个子系统

$$y_a(t) = \boldsymbol{\varphi}_a^T(t) \boldsymbol{a} + v(t), \quad (203)$$

$$y_b(t) = \boldsymbol{\varphi}_b^T(t) \boldsymbol{b} + v(t), \quad (204)$$

$$y_c(t) = \boldsymbol{f}^T(t) \boldsymbol{c} + v(t). \quad (205)$$

针对3个子系统(203)–(205),定义3个梯度准则函数:

$$J_a(\boldsymbol{a}) := \frac{1}{2} [y_a(t) - \boldsymbol{\varphi}_a^T(t) \boldsymbol{a}]^2,$$

$$J_b(\boldsymbol{b}) := \frac{1}{2} [y_b(t) - \boldsymbol{\varphi}_b^T(t) \boldsymbol{b}]^2,$$

$$J_c(\boldsymbol{c}) := \frac{1}{2} [y_c(t) - \boldsymbol{f}^T(t) \boldsymbol{c}]^2.$$

利用负梯度搜索原理极小化准则函数 $J_a(\boldsymbol{a})$, $J_b(\boldsymbol{b})$ 和 $J_c(\boldsymbol{c})$, 参考 KT-2S-SG 算法(151)–(159)的推导,可以得到辨识 IN-EE 系统参数向量的基于关键项分离的三阶段随机梯度算法(Key Term separation based Three-Stage Stochastic Gradient algorithm, KT-3S-SG 算法)^[29]:

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \frac{\boldsymbol{\varphi}_a(t)}{r_1(t)} e(t), \quad (206)$$

$$e(t) = y(t) - \boldsymbol{\varphi}_a^T(t) \hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t-1), \quad (207)$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2, \quad r_1(0) = 1, \quad (208)$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}_b(t)}{r_2(t)} e(t), \quad (209)$$

$$r_2(t) = r_2(t-1) + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2, \quad r_2(0) = 1, \quad (210)$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{\boldsymbol{f}(t)}{r_3(t)} e(t), \quad (211)$$

$$r_3(t) = r_3(t-1) + \|\boldsymbol{f}(t)\|^2, \quad r_3(0) = 1, \quad (212)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (213)$$

$$\boldsymbol{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (214)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (215)$$

$$\hat{u}(t) = \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t). \quad (216)$$

如果式(206), (209)和(211)中 $r_1(t)$, $r_2(t)$ 和 $r_3(t)$ 采用一个共同的 $r(t)$, 即

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

则 KT-3S-SG 算法完全等同于 KT-SG 算法(97)–(104).

5.2 基于关键项分离的三阶段多新息随机梯度辨识算法

定义新息长度为 p 的堆积输出向量 $\boldsymbol{Y}(p, t)$ 及堆积信息矩阵 $\boldsymbol{\Phi}_a(p, t)$, $\hat{\boldsymbol{\Phi}}_b(p, t)$ 和 $\boldsymbol{F}(p, t)$ 如下:

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

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

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

$$\boldsymbol{F}(p, t) = [\boldsymbol{f}(t), \boldsymbol{f}(t-1), \dots, \boldsymbol{f}(t-p+1)] \in \mathbf{R}^{m \times p}.$$

将式(206), (209)和(211)中标量新息 $e(t) = y(t) - \boldsymbol{\varphi}_a^T(t) \hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t-1) \in \mathbf{R}$ 扩展为新息向量

$$\begin{aligned} \boldsymbol{E}(p, t) &:= \boldsymbol{Y}(p, t) - \boldsymbol{\Phi}_a(p, t) \hat{\boldsymbol{a}}(t-1) - \\ &\quad \hat{\boldsymbol{\Phi}}_b(p, t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{F}(p, t) \hat{\boldsymbol{c}}(t-1) \in \mathbf{R}^p. \end{aligned}$$

扩展 KT-3S-SG 算法(206)–(216)中信息向量 $\boldsymbol{\varphi}_a(t)$, $\hat{\boldsymbol{\varphi}}_b(t)$ 和 $\boldsymbol{f}(t)$ 为堆积信息矩阵 $\boldsymbol{\Phi}_a(p, t)$, $\hat{\boldsymbol{\Phi}}_b(p, t)$ 和 $\boldsymbol{F}(p, t)$, 可以得到辨识 IN-EE 系统参数向量的基于关键项分离的三阶段多新息随机梯度算法(Key Term separation based Three-Stage Multi-Innovation Stochastic Gradient algorithm, KT-3S-MISG 算法)^[29]:

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \frac{\boldsymbol{\Phi}_a(p, t)}{r_1(t)} \boldsymbol{E}(p, t), \quad \hat{\boldsymbol{a}}(0) = \mathbf{1}_{n_a}/p_0, \quad (217)$$

$$\begin{aligned} \boldsymbol{E}(p, t) &= \boldsymbol{Y}(p, t) - \boldsymbol{\Phi}_a^T(p, t) \hat{\boldsymbol{a}}(t-1) - \\ &\quad \hat{\boldsymbol{\Phi}}_b^T(p, t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{F}^T(p, t) \hat{\boldsymbol{c}}(t-1), \quad (218) \end{aligned}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2, \quad r_1(0) = 1, \quad (219)$$

$$\hat{\mathbf{b}}(t) = \hat{\mathbf{b}}(t-1) + \frac{\hat{\Phi}_b(p, t)}{r_2(t)} \mathbf{E}(p, t), \quad \hat{\mathbf{b}}(0) = \mathbf{1}_{n_b}/p_0, \quad (220)$$

$$r_2(t) = r_2(t-1) + \|\hat{\varphi}_b(t)\|^2, \quad r_2(0) = 1, \quad (221)$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \frac{\mathbf{F}(p, t)}{r_3(t)} \mathbf{E}(p, t), \quad \hat{\mathbf{c}}(0) = \mathbf{1}_m/p_0, \quad (222)$$

$$r_3(t) = r_3(t-1) + \|\mathbf{f}(t)\|^2, \quad r_3(0) = 1, \quad (223)$$

$$\mathbf{Y}(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (224)$$

$$\Phi_a(p, t) = [\varphi_a(t), \varphi_a(t-1), \dots, \varphi_a(t-p+1)], \quad (225)$$

$$\hat{\Phi}_b(p, t) = [\hat{\varphi}_b(t), \hat{\varphi}_b(t-1), \dots, \hat{\varphi}_b(t-p+1)], \quad (226)$$

$$\mathbf{F}(p, t) = [\mathbf{f}(t), \mathbf{f}(t-1), \dots, \mathbf{f}(t-p+1)], \quad (227)$$

$$\varphi_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (228)$$

$$\hat{\varphi}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (229)$$

$$\mathbf{f}(t) = [f_1(u(t)), f_2(u(t)), \dots, f_m(u(t))]^T, \quad (230)$$

$$\hat{u}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t). \quad (231)$$

当新息长度 $p = 1$ 时, KT-3S-MISG 算法退化为 KT-3S-SG 算法(206)–(216). 如果式(217), (220)和(222)中 $r_1(t)$, $r_2(t)$ 和 $r_3(t)$ 采用一个共同的 $r(t)$, 即:

$$r(t) = r(t-1) + \|\varphi_a(t)\|^2 + \|\hat{\varphi}_b(t)\|^2 + \|\mathbf{f}(t)\|^2, \quad r(0) = 1, \quad (232)$$

则 KT-3S-MISG 算法完全等同于 KT-MISG 算法(105)–(115). 在收敛性分析中, $r(t)$ 也可修改为

$$r(t) = \lambda r(t-1) + \|\Phi_a(p, t)\|^2 + \|\hat{\Phi}_b(p, t)\|^2 + \|\mathbf{F}(p, t)\|^2, \quad 0 \leq \lambda \leq 1, \quad r(0) = 1. \quad (233)$$

KT-3S-MISG 算法(217)–(231) 辨识参数 a_i , b_i 和 c_i 的计算步骤如下:

1) 初始化: $t = 1$, 给定信息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$; 令 $\hat{\mathbf{a}}(0) = \mathbf{1}_{n_a}/p_0$, $\hat{\mathbf{b}}(0) = \mathbf{1}_{n_b}/p_0$, $\hat{\mathbf{c}}(0) = \mathbf{1}_m/p_0$, $r_1(0) = r_2(0) = r_3(0) = 1$, $\hat{u}(t) = 1/p_0$, $t \leq 0, p_0 = 10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 利用式(224)构造输出向量 $\mathbf{Y}(p, t)$, 利用式(228)–(230)依次构造信息向量 $\varphi_a(t)$, $\hat{\varphi}_b(t)$ 和 $\mathbf{f}(t)$, 用式(225)–(227)构造堆积信息矩阵 $\Phi_a(p, t)$, $\hat{\Phi}_b(p, t)$ 和 $\mathbf{F}(p, t)$.

3) 利用式(218)计算新息向量 $\mathbf{E}(p, t)$, 利用式(219), (221)和(223)计算 $r_1(t)$, $r_2(t)$ 和 $r_3(t)$.

4) 利用式(217), (220)和(222)刷新参数估计向量 $\hat{\mathbf{a}}(t)$, $\hat{\mathbf{b}}(t)$ 和 $\hat{\mathbf{c}}(t)$.

5) 如果 $t < L$, t 增加 1, 转到第 2) 步, 否则获得参数估计向量 $\hat{\mathbf{a}}(t)$, $\hat{\mathbf{b}}(t)$ 和 $\hat{\mathbf{c}}(t)$.

利用 KT-3S-MISG 算法(217)–(231) 计算 IN-

EE 系统参数估计的流程如图 9 所示.

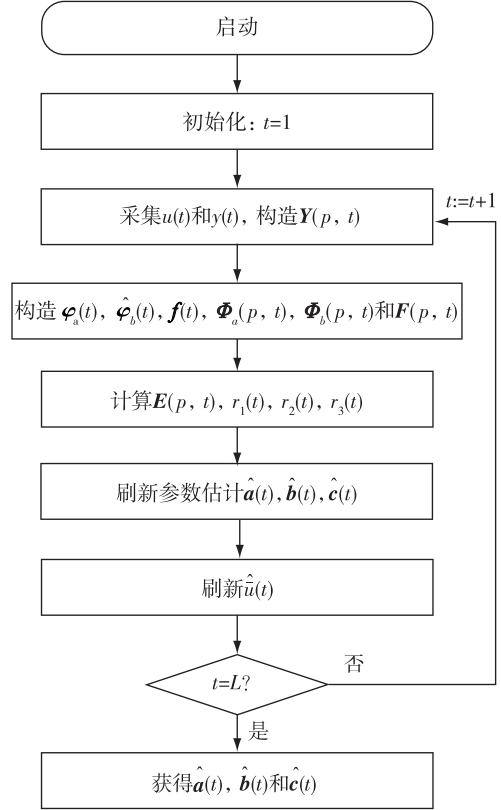


图 9 基于关键项分离的三阶段多新息随机梯度算法流程
Fig. 9 The flowchart of the KT-3S-MISG algorithm

5.3 基于关键项分离的三阶段递推最小二乘辨识算法

将 3 个子系统(203)–(205)重写如下:

$$y_a(t) = \varphi_a^T(t) \mathbf{a} + v(t), \quad (234)$$

$$y_b(t) = \varphi_b^T(t) \mathbf{b} + v(t), \quad (235)$$

$$y_c(t) = \mathbf{f}^T(t) \mathbf{c} + v(t), \quad (236)$$

其中

$$y_a(t) := y(t) - \varphi_b^T(t) \mathbf{b} - \mathbf{f}^T(t) \mathbf{c},$$

$$y_b(t) := y(t) - \varphi_a^T(t) \mathbf{a} - \mathbf{f}^T(t) \mathbf{c},$$

$$y_c(t) := y(t) - \varphi_a^T(t) \mathbf{a} - \varphi_b^T(t) \mathbf{b}.$$

定义 3 个二次准则函数:

$$J_a(\mathbf{a}) := \frac{1}{2} \sum_{j=1}^t [y_a(j) - \varphi_a^T(j) \mathbf{a}]^2,$$

$$J_b(\mathbf{b}) := \frac{1}{2} \sum_{j=1}^t [y_b(j) - \varphi_b^T(j) \mathbf{b}]^2,$$

$$J_c(\mathbf{c}) := \frac{1}{2} \sum_{j=1}^t [y_c(j) - \mathbf{f}^T(j) \mathbf{c}]^2.$$

分别令 $J_a(\mathbf{a})$, $J_b(\mathbf{b})$ 和 $J_c(\mathbf{c})$ 对 \mathbf{a} , \mathbf{b} 和 \mathbf{c} 的导数为零, 假设对于大 t , 相关矩阵可逆时, 可以求得

$$\begin{aligned}\hat{\boldsymbol{a}}(t) &= \left[\sum_{j=1}^t \boldsymbol{\varphi}_a(j) \boldsymbol{\varphi}_a^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_a(j) y_a(j) = \\ & \left[\sum_{j=1}^t \boldsymbol{\varphi}_a(j) \boldsymbol{\varphi}_a^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_a(j) [y(j) - \boldsymbol{\varphi}_b^T(j) \boldsymbol{b} - \boldsymbol{f}^T(j) \boldsymbol{c}], \\ \hat{\boldsymbol{b}}(t) &= \left[\sum_{j=1}^t \boldsymbol{\varphi}_b(j) \boldsymbol{\varphi}_b^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_b(j) y_b(j) = \\ & \left[\sum_{j=1}^t \boldsymbol{\varphi}_b(j) \boldsymbol{\varphi}_b^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_b(j) [y(j) - \boldsymbol{\varphi}_a^T(j) \boldsymbol{a} - \boldsymbol{f}^T(j) \boldsymbol{c}], \\ \hat{\boldsymbol{c}}(t) &= \left[\sum_{j=1}^t \boldsymbol{f}(j) \boldsymbol{f}^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{f}(j) y_c(j) = \\ & \left[\sum_{j=1}^t \boldsymbol{f}(j) \boldsymbol{f}^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{f}(j) [y(j) - \boldsymbol{\varphi}_a^T(j) \boldsymbol{a} - \boldsymbol{\varphi}_b^T(j) \boldsymbol{b}].\end{aligned}$$

上述表达式右边包含未知变量,故算法不可实现.解决的办法是未知量 $\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\varphi}_b(t)$ 均用其估计 $\hat{\boldsymbol{b}}(t-1), \hat{\boldsymbol{c}}(t-1), \hat{\boldsymbol{a}}(t-1), \hat{\boldsymbol{\varphi}}_b(t)$ 代替,便得到辨识 IN-EE 系统参数向量的基于关键项分离的三阶段最小二乘算法(Key Term separation based Three-Stage Least Squares algorithm, KT-3S-LS 算法):

$$\hat{\boldsymbol{a}}(t) = \left[\sum_{j=1}^t \boldsymbol{\varphi}_a(j) \boldsymbol{\varphi}_a^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{\varphi}_a(j) [y(j) - \hat{\boldsymbol{\varphi}}_b^T(j) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^T(j) \hat{\boldsymbol{c}}(t-1)], \quad (237)$$

$$\hat{\boldsymbol{b}}(t) = \left[\sum_{j=1}^t \hat{\boldsymbol{\varphi}}_b(j) \hat{\boldsymbol{\varphi}}_b^T(j) \right]^{-1} \sum_{j=1}^t \hat{\boldsymbol{\varphi}}_b(j) [y(j) - \boldsymbol{\varphi}_a^T(j) \hat{\boldsymbol{a}}(t-1) - \boldsymbol{f}^T(j) \hat{\boldsymbol{c}}(t-1)], \quad (238)$$

$$\hat{\boldsymbol{c}}(t) = \left[\sum_{j=1}^t \boldsymbol{f}(j) \boldsymbol{f}^T(j) \right]^{-1} \sum_{j=1}^t \boldsymbol{f}(j) [y(j) - \boldsymbol{\varphi}_a^T(j) \hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\boldsymbol{b}}(t-1)], \quad (239)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (240)$$

$$\hat{u}(t) = \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t). \quad (241)$$

算法(237)–(241)涉及到计算矩阵的逆,可以运用矩阵求逆引理(式(31)),推导出辨识 IN-EE 系统参数向量的基于关键项分离的三阶段递推最小二乘算法(Key Term separation based Three-Stage Recursive Least Squares algorithm, KT-3S-RLS 算法):

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \boldsymbol{L}_a(t) \boldsymbol{e}(t), \quad \hat{\boldsymbol{a}}(0) = \mathbf{1}_{n_a}/p_0, \quad (242)$$

$$\boldsymbol{e}(t) = [y(t) - \boldsymbol{\varphi}_a^T(t) \hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t-1)], \quad (243)$$

$$\boldsymbol{L}_a(t) = \frac{\boldsymbol{P}_a(t-1) \boldsymbol{\varphi}_a(t)}{1 + \boldsymbol{\varphi}_a^T(t) \boldsymbol{P}_a(t-1) \boldsymbol{\varphi}_a(t)}, \quad (244)$$

$$\boldsymbol{P}_a(t) = [\boldsymbol{I}_{n_a} - \boldsymbol{L}_a(t) \boldsymbol{\varphi}_a^T(t)] \boldsymbol{P}_a(t-1), \quad \boldsymbol{P}_a(0) = p_0 \boldsymbol{I}_{n_a}, \quad (245)$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \boldsymbol{L}_b(t) \boldsymbol{e}(t), \quad \hat{\boldsymbol{b}}(0) = \mathbf{1}_{n_b}/p_0, \quad (246)$$

$$\boldsymbol{L}_b(t) = \frac{\boldsymbol{P}_b(t-1) \hat{\boldsymbol{\varphi}}_b(t)}{1 + \hat{\boldsymbol{\varphi}}_b^T(t) \boldsymbol{P}_b(t-1) \hat{\boldsymbol{\varphi}}_b(t)}, \quad (247)$$

$$\boldsymbol{P}_b(t) = [\boldsymbol{I}_{n_b} - \boldsymbol{L}_b(t) \hat{\boldsymbol{\varphi}}_b^T(t)] \boldsymbol{P}_b(t-1), \quad \boldsymbol{P}_b(0) = p_0 \boldsymbol{I}_{n_b}, \quad (248)$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \boldsymbol{L}_c(t) \boldsymbol{e}(t), \quad \hat{\boldsymbol{c}}(0) = \mathbf{1}_m/p_0, \quad (249)$$

$$\boldsymbol{L}_c(t) = \frac{\boldsymbol{P}_c(t-1) \boldsymbol{f}(t)}{1 + \boldsymbol{f}^T(t) \boldsymbol{P}_c(t-1) \boldsymbol{f}(t)}, \quad (250)$$

$$\boldsymbol{P}_c(t) = [\boldsymbol{I}_m - \boldsymbol{L}_c(t) \boldsymbol{f}^T(t)] \boldsymbol{P}_c(t-1), \quad \boldsymbol{P}_c(0) = p_0 \boldsymbol{I}_m, \quad (251)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (252)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (253)$$

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

$$\hat{u}(t) = \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t). \quad (255)$$

KT-3S-RLS 算法每步的计算量如表 9 所示 ($n_0 := n_a + n_b + m$).

5.4 基于关键项分离的三阶段多新息最小二乘辨识算法

定义新息长度为 p 的堆积输出向量 $\boldsymbol{Y}(p, t)$ 及堆积信息矩阵 $\boldsymbol{\Phi}_a(p, t), \hat{\boldsymbol{\Phi}}_b(p, t)$ 和 $\boldsymbol{F}(p, t)$ 如下:

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

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

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

$$\boldsymbol{F}(p, t) := [\boldsymbol{f}(t), \boldsymbol{f}(t-1), \dots, \boldsymbol{f}(t-p+1)] \in \mathbf{R}^{m \times p}.$$

将式(243)中标量新息

$$\boldsymbol{e}(t) = y(t) - \boldsymbol{\varphi}_a^T(t) \hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^T(t) \hat{\boldsymbol{c}}(t-1) \in \mathbf{R}$$

扩展为新息向量

$$\boldsymbol{E}(p, t) := \boldsymbol{Y}(p, t) - \boldsymbol{\Phi}_a(p, t) \hat{\boldsymbol{a}}(t-1) -$$

$$\hat{\boldsymbol{\Phi}}_b(p, t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{F}(p, t) \hat{\boldsymbol{c}}(t-1) \in \mathbf{R}^p,$$

将 KT-3S-RLS 算法(242)–(255)中信息向量 $\boldsymbol{\varphi}_a(t), \hat{\boldsymbol{\varphi}}_b(t)$ 和 $\boldsymbol{f}(t)$ 扩展为堆积信息矩阵 $\boldsymbol{\Phi}_a(p, t), \hat{\boldsymbol{\Phi}}_b(p, t)$ 和 $\boldsymbol{F}(p, t)$,可以得到辨识 IN-EE 系统参数向量的基于关键项分离的三阶段多新息最小二乘算法(Key Term separation based Three-Stage Multi-Innovation Least Squares algorithm, KT-3S-MILS 算法):

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \boldsymbol{L}_a(t) \boldsymbol{E}(p, t), \quad (256)$$

$$\boldsymbol{E}(p, t) = \boldsymbol{Y}(p, t) - \boldsymbol{\Phi}_a(p, t) \hat{\boldsymbol{a}}(t-1) -$$

$$\hat{\boldsymbol{\Phi}}_b(p, t) \hat{\boldsymbol{b}}(t-1) - \boldsymbol{F}(p, t) \hat{\boldsymbol{c}}(t-1), \quad (257)$$

$$\boldsymbol{L}_a(t) = \boldsymbol{P}_a(t-1) \boldsymbol{\Phi}_a(p, t) [\boldsymbol{I}_p +$$

$$\boldsymbol{\Phi}_a^T(p, t) \boldsymbol{P}_a(t-1) \boldsymbol{\Phi}_a^T(p, t)]^{-1}, \quad (258)$$

$$\boldsymbol{P}_a(t) = [\boldsymbol{I}_{n_a} - \boldsymbol{L}_a(t) \boldsymbol{\Phi}_a^T(p, t)] \boldsymbol{P}_a(t-1), \quad (259)$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \boldsymbol{L}_b(t) \boldsymbol{E}(p, t), \quad (260)$$

$$\boldsymbol{L}_b(t) = \boldsymbol{P}_b(t-1) \hat{\boldsymbol{\Phi}}_b(p, t) [\boldsymbol{I}_p +$$

$$\hat{\boldsymbol{\Phi}}_b^T(p, t) \boldsymbol{P}_b(t-1) \hat{\boldsymbol{\Phi}}_b^T(p, t)]^{-1}, \quad (261)$$

$$\boldsymbol{P}_b(t) = [\boldsymbol{I}_{n_b} - \boldsymbol{L}_b(t) \hat{\boldsymbol{\Phi}}_b^T(p, t)] \boldsymbol{P}_b(t-1), \quad (262)$$

表9 基于关键项分离的递推最小二乘算法的计算量

Table 9 The computational efficiency of the KT-3S-RLS algorithm

变量	计算步骤	乘法次数	加法次数
$\hat{\mathbf{a}}(t)$	$\hat{\mathbf{a}}(t) = \hat{\mathbf{a}}(t-1) + \mathbf{L}_a(t) e(t) \in \mathbf{R}^{n_a}$	n_a	n_a
	$e(t) := y(t) - \boldsymbol{\varphi}_a^T(t) \hat{\mathbf{a}}(t-1) - \mathbf{f}^T(t) \hat{\mathbf{c}}(t-1) - \hat{\boldsymbol{\varphi}}_b^T(t) \hat{\mathbf{b}}(t-1) \in \mathbf{R}$	n_0	n_0
$\mathbf{L}_a(t)$	$\mathbf{L}_a(t) = \boldsymbol{\zeta}_1(t) / [1 + \boldsymbol{\varphi}_a^T(t) \boldsymbol{\zeta}_1(t)] \in \mathbf{R}^{n_a}$	$2n_a$	n_a
	$\boldsymbol{\zeta}_1(t) := \mathbf{P}_a(t-1) \boldsymbol{\varphi}_a(t) \in \mathbf{R}^{n_a}$	n_a^2	$(n_a - 1)n_a$
$\mathbf{P}_a(t)$	$\mathbf{P}_a(t) = \mathbf{P}_a(t-1) - \mathbf{L}_a(t) \boldsymbol{\zeta}_1^T(t) \in \mathbf{R}^{n_a \times n_a}$	n_a^2	n_a^2
$\hat{\mathbf{b}}(t)$	$\hat{\mathbf{b}}(t) = \hat{\mathbf{b}}(t-1) + \mathbf{L}_b(t) e(t) \in \mathbf{R}^{n_b}$	n_b	n_b
$\mathbf{L}_b(t)$	$\mathbf{L}_b(t) = \boldsymbol{\zeta}_2(t) / [1 + \hat{\boldsymbol{\varphi}}_b^T(t) \boldsymbol{\zeta}_2(t)] \in \mathbf{R}^{n_b}$	$2n_b$	n_b
	$\boldsymbol{\zeta}_2(t) := \mathbf{P}_b(t-1) \hat{\boldsymbol{\varphi}}_b(t) \in \mathbf{R}^{n_b}$	n_b^2	$(n_b - 1)n_b$
$\mathbf{P}_b(t)$	$\mathbf{P}_b(t) = \mathbf{P}_b(t-1) - \mathbf{L}_b(t) \boldsymbol{\zeta}_2^T(t) \in \mathbf{R}^{n_b \times n_b}$	n_b^2	n_b^2
$\hat{\mathbf{c}}(t)$	$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \mathbf{L}_c(t) e(t) \in \mathbf{R}^m$	m	m
$\mathbf{L}_c(t)$	$\mathbf{L}_c(t) = \boldsymbol{\zeta}_3(t) / [1 + \mathbf{f}^T(t) \boldsymbol{\zeta}_3(t)] \in \mathbf{R}^m$	$2m$	m
	$\boldsymbol{\zeta}_3(t) := \mathbf{P}_c(t-1) \mathbf{f}(t) \in \mathbf{R}^m$	m^2	$(m-1)m$
$\mathbf{P}_c(t)$	$\mathbf{P}_c(t) = \mathbf{P}_c(t-1) - \mathbf{L}_c(t) \boldsymbol{\zeta}_3^T(t) \in \mathbf{R}^{m \times m}$	m^2	m^2
$\hat{\mathbf{u}}(t)$	$\hat{\mathbf{u}}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t) \in \mathbf{R}$	m	$m-1$
总数		$2(n_a^2 + n_b^2 + m^2) + 4n_0 + m$	$2(n_a^2 + n_b^2 + m^2) + 2n_0 + m - 1$
总 flop 数		$N_9 := 4(n_a^2 + n_b^2 + m^2) + 6n_0 + 2m - 1$	

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t-1) + \mathbf{L}_c(t) \mathbf{E}(p, t), \quad \hat{\mathbf{c}}(0) = \mathbf{1}_m / p_0, \quad (263)$$

$$\mathbf{L}_c(t) = \mathbf{P}_c(t-1) \mathbf{F}(p, t) [\mathbf{I}_p + \mathbf{F}^T(p, t) \mathbf{P}_c(t-1) \mathbf{F}(p, t)]^{-1}, \quad (264)$$

$$\mathbf{P}_c(t) = [\mathbf{I}_m - \mathbf{L}_c(t) \mathbf{F}^T(p, t)] \mathbf{P}_c(t-1), \quad (265)$$

$$\mathbf{Y}(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \quad (266)$$

$$\boldsymbol{\Phi}_a(p, t) = [\boldsymbol{\varphi}_a(t), \boldsymbol{\varphi}_a(t-1), \dots, \boldsymbol{\varphi}_a(t-p+1)], \quad (267)$$

$$\hat{\boldsymbol{\Phi}}_b(p, t) = [\hat{\boldsymbol{\varphi}}_b(t), \hat{\boldsymbol{\varphi}}_b(t-1), \dots, \hat{\boldsymbol{\varphi}}_b(t-p+1)], \quad (268)$$

$$\mathbf{F}(p, t) = [\mathbf{f}(t), \mathbf{f}(t-1), \dots, \mathbf{f}(t-p+1)], \quad (269)$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \quad (270)$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\hat{u}(t-1), \hat{u}(t-2), \dots, \hat{u}(t-n_b)]^T, \quad (271)$$

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

$$\hat{u}(t) = \mathbf{f}^T(t) \hat{\mathbf{c}}(t). \quad (273)$$

当新息长度 $p = 1$ 时, KT-3S-MILS 算法退化为 KT-3S-RLS 算法(242)—(255). 利用 KT-3S-MILS 算法(256)—(273) 估计 IN-EE 系统参数向量的计算步骤如下:

1) 初始化: $t = 1$, 给定信息长度 p 、数据长度 L 和基函数 $f_j(\cdot)$; 令 $\hat{\mathbf{a}}(0) = \mathbf{1}_{n_a} / p_0$, $\hat{\mathbf{b}}(0) = \mathbf{1}_{n_b} / p_0$ 和 $\hat{\mathbf{c}}(0) = \mathbf{1}_m / p_0$, $\mathbf{P}_a(0) = p_0 \mathbf{I}_{n_a}$, $\mathbf{P}_b(0) = p_0 \mathbf{I}_{n_b}$, $\mathbf{P}_c(0) = p_0 \mathbf{I}_m$, $\hat{u}(j) = 1 / p_0, j \leq 0, p_0 = 10^6$.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$, 用式(266)构造输出向量 $\mathbf{Y}(p, t)$, 用式(270)—(272)构造信息向量 $\boldsymbol{\varphi}_a(t)$, $\hat{\boldsymbol{\varphi}}_b(t)$ 和 $\mathbf{f}(t)$, 用式(267)—(269)构造 $\boldsymbol{\Phi}_a(p, t)$, $\hat{\boldsymbol{\Phi}}_b(p, t)$ 和 $\mathbf{F}(p, t)$.

3) 用式(257)计算新息向量 $\mathbf{E}(p, t)$.

4) 用式(258)—(259)计算 $\mathbf{L}_a(t)$ 和 $\mathbf{P}_a(t)$, 用式(256)刷新估计 $\hat{\mathbf{a}}(t)$, 用式(261)—(262)计算 $\mathbf{L}_b(t)$ 和 $\mathbf{P}_b(t)$, 用式(260)刷新估计 $\hat{\mathbf{b}}(t)$, 用式(264)—(265)计算 $\mathbf{L}_c(t)$ 和 $\mathbf{P}_c(t)$, 用式(263)刷新估计 $\hat{\mathbf{c}}(t)$.

5) 用式(273)计算中间变量 $\hat{u}(t)$.

6) 如果 $t < L$, t 增加 1, 转到第 2) 步, 否则获得参数估计 $\hat{\mathbf{a}}(L)$, $\hat{\mathbf{b}}(L)$ 和 $\hat{\mathbf{c}}(L)$.

KT-3S-MILS 算法(256)—(273) 计算参数估计的流程如图 10 所示, 算法每步的计算量如表 10 所示 ($n_0 := n_a + n_b + m$).

6 辨识算法的计算量比较

下面对同类算法的计算量进行比较, 简单讨论基于过参数化模型的 RLS 算法与 HLS 算法、基于过参数化模型 MILS 算法与 HMILS 算法, 基于关键项分离的 KT-RLS 算法、KT-2S-RLS 算法、KT-3S-RLS 算法, 基于关键项分离的 KT-MILS 算法、KT-2S-MILS 算法、KT-3S-MILS 算法, 它们每步的计算量(总 flop 数)如表 11 所示. 为了更直观看出辨识算法计算量的大小, 表中给出了 $n_a = 10, n_b = 10, m = 10, p = 8$ 的 flop 数. 可以从表 11 中的数据看出同类辨识算法的计算量大小.

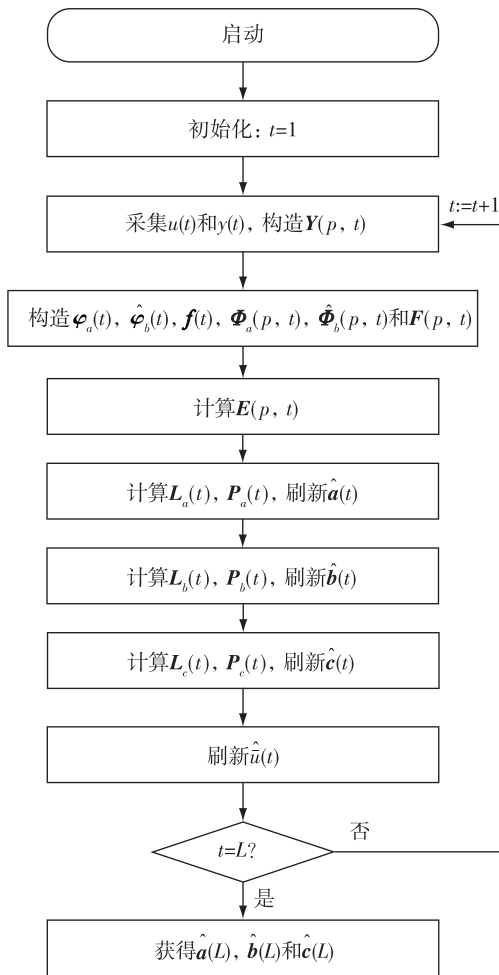


图 10 基于关键项分离的三阶段多新息最小二乘算法流程

Fig. 10 The flowchart of the KT-3S-MILS algorithm

7 结语

针对输入非线性方程误差系统(即输入非线性受控自回归系统),首先,研究了基于过参数化模型的辨识方法,包括随机梯度算法、多新息随机梯度算法、最小二乘算法、多新息最小二乘算法,以及递阶随机梯度算法、递阶多新息随机梯度算法、递阶最小二乘算法、递阶多新息最小二乘算法。

其次,由于过参数化模型产生很多冗余参数,需要辨识的参数比系统实际参数数目多,因而算法的计算量大(是指同类算法之间的比较),特点是不涉及估算未知中间变量(非线性块的输出).为了减小过参数化模型辨识算法计算量大的问题,利用关键项分离原理,研究了 IN-EE 系统的基于关键项分离的随机梯度算法、多新息随机梯度算法、递推最小二乘算法、多新息最小二乘算法。

再次,为减小基于关键项分离辨识算法的计算量,研究了 IN-EE 系统的基于关键项分离的两阶段随机梯度算法、两阶段多新息随机梯度算法、两阶段递推最小二乘算法、两阶段多新息最小二乘算法。

最后,为进一步减小基于关键项分离的两阶段辨识算法的计算量,研究了 IN-EE 系统的基于关键项分离的三阶段随机梯度算法、三阶段多新息随机梯度算法、三阶段递推最小二乘算法、三阶段多新息最小二乘算法。

表 10 基于关键项分离的三阶段多新息最小二乘算法的计算量

Table 10 The computational efficiency of the KT-3S-MILS algorithm

表达式	乘法次数	加法次数
$\hat{a}(t) = \hat{a}(t-1) + L_a(t)E(p, t) \in \mathbf{R}^{n_a}$	$n_a p$	$n_a p$
$E(p, t) = Y(p, t) - \Phi_a(p, t)\hat{a}(t-1) - \Phi_b(p, t)\hat{b}(t-1) - F(p, t)\hat{c}(t-1) \in \mathbf{R}^p$	$n_0 p$	$n_0 p$
$L_a(t) = \zeta_a(t) [I_p + \Phi_a^T(p, t)\zeta_a(t)]^{-1} \in \mathbf{R}^{n_a \times p}$	$2n_a p^2 + p^3$	$2n_a p^2 + p^3 - n_a p - 2p^2$
$\zeta_a(t) := P_a(t-1)\Phi_a(p, t) \in \mathbf{R}^{n_a \times p}$	$n_a^2 p$	$n_a^2 p - n_a p$
$P_a(t) = P_a(t-1) - L_a(t)\zeta_a^T(t) \in \mathbf{R}^{n_a \times n_a}$	$n_a^2 p$	$n_a^2 p$
$\hat{b}(t) = \hat{b}(t-1) + L_b(t)E(p, t) \in \mathbf{R}^{n_b}$	$n_b p$	$n_b p$
$L_b(t) = \zeta_b(t) [I_p + \Phi_b^T(p, t)\zeta_b(t)]^{-1} \in \mathbf{R}^{n_b \times p}$	$2n_b p^2 + p^3$	$2n_b p^2 + p^3 - n_b p - 2p^2$
$\zeta_b(t) := P_b(t-1)F(p, t) \in \mathbf{R}^{n_b \times p}$	$n_b^2 p$	$n_b^2 p - n_b p$
$P_b(t) = P_b(t-1) - L_b(t)\zeta_b^T(t) \in \mathbf{R}^{n_b \times n_b}$	$n_b^2 p$	$n_b^2 p$
$\hat{c}(t) = \hat{c}(t-1) + L_c(t)E(p, t) \in \mathbf{R}^m$	$m p$	$m p$
$L_c(t) = \zeta_c(t) [I_p + F^T(p, t)\zeta_c(t)]^{-1} \in \mathbf{R}^{m \times p}$	$2m p^2 + p^3$	$2m p^2 + p^3 - m p - 2p^2$
$\zeta_c(t) := P_c(t-1)F(p, t) \in \mathbf{R}^{m \times p}$	$m^2 p$	$m^2 p - m p$
$P_c(t) = P_c(t-1) - L_c(t)\zeta_c^T(t) \in \mathbf{R}^{m \times m}$	$m^2 p$	$m^2 p$
$\hat{u}(t) = f^T(t)\hat{c}(t) \in \mathbf{R}$	m	$m - 1$
总数	$2n_a^2 p + 2n_b^2 p + 2m^2 p +$ $2n_0 p^2 + 2n_0 p + 3p^3 + m$	$2n_a^2 p + 2n_b^2 p + 2m^2 p +$ $2n_0 p^2 + 3p^3 - 6p^2 + m - 1$
总 flop 数	$N_{10} := 4p(n_a^2 + n_b^2 + m^2) + 4n_0 p^2 + 2n_0 p + 6p^3 - 6p^2 + 2m - 1$	

表 11 最小二乘类辨识算法每步计算量的比较

Table 11 The computational efficiency of the least squares type identification algorithms

算法	总 Flop 数	例子 flop 数
RLS	$N_1 := 4(n_a + mn_b)^2 + 6(n_a + mn_b)$	49 060
HLS	$N_3 := 4n_a^2 + 4m^2 n_b + 6n_a + 6mn_b$	5 060
MILS	$N_2 := 4(n_a + mn_b)^2 p + 4(n_a + mn_b)p^2 + 2(n_a + mn_b)p + 2p^3 - 2p^2$	418 016
HMILS	$N_4 := 4(n_a^2 + mn_b^2)p + 4(n_a + mn_b)p^2 + 2(n_a + mn_b)p + 2(n_b + 1)(p^3 - p^2)$	74 976
KT-RLS	$N_5 := 4(n_a + n_b + m)^2 + 6(n_a + n_b + m) + 2m - 1$	3 799
KT-2S-RLS	$N_7 := 4(n_a + n_b)^2 + 4m^2 + 6(n_a + n_b + m) + 2m - 1$	2 199
KT-3S-RLS	$N_9 := 4(n_a^2 + n_b^2 + m^2) + 6(n_a + n_b + m) + 2m - 1$	1 399
KT-MILS	$N_6 := 4(n_a + n_b + m)^2 p + 4(n_a + n_b + m)p^2 + 2(n_a + n_b + m)p + 2p^3 - 2p^2 + 2m - 1$	37 875
KT-2S-MILS	$N_8 := 4(n_a + n_b)^2 p + 4m^2 p + 4(n_a + n_b + m)p^2 + 2(n_a + n_b + m)p + 4p^3 - 4p^2 + 2m - 1$	25 971
KT-3S-MILS	$N_{10} := 4(n_a^2 + n_b^2 + m^2)p + 4(n_a + n_b + m)p^2 + 2(n_a + n_b + m)p + 6p^3 - 6p^2 + 2m - 1$	20 467

这些算法的收敛性分析都是需要研究的辨识课题.尽管本文的方法是针对输入非线性系统提出的,但是可以用于线性系统,也可以推广到输出非线性系统、反馈非线性系统等.

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] Vörös J. Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities [J]. IEEE Transactions on Automatic Control, 1999, 44(11): 2145-2149
- [11] Vörös J. Modeling and parameter identification of systems with multi-segment piecewise-linear characteristics [J]. IEEE Transactions on Automatic Control, 2002, 47(1): 184-188
- [12] Vörös J. Identification of Hammerstein systems with time-

- varying piecewise-linear characteristics [J]. IEEE Transaction on Circuit Systems II, 2005, 52(12): 865-869
- [13] Ding F, Chen T. Performance analysis of multi-innovation gradient type identification methods [J]. Automatica, 2007, 43(1): 1-14
- [14] Ding F, Liu X P, Liu G. Multi-innovation least squares identification for system modeling [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(3): 767-778
- [15] Ding F. Several multi-innovation identification methods [J]. Digital Signal Processing, 2010, 20(4): 1027-1039
- [16] Ding F, Liu G, Liu X P. Parameter estimation with scarce measurements [J]. Automatica, 2011, 47(8): 1646-1655
- [17] Ding F. Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling [J]. Applied Mathematical Modelling, 2013, 37(4): 1694-1704
- [18] Ding F, Liu X P, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises [J]. Signal Processing, 2009, 89(10): 1883-1890
- [19] Ding F, Chen H B, Li M. Multi-innovation least squares identification methods based on the auxiliary model for MISO systems [J]. Applied Mathematics and Computation, 2007, 187(2): 658-668
- [20] Zhang J B, Ding F, Shi Y. Self-tuning control based on multi-innovation stochastic gradient parameter estimation [J]. Systems & Control Letters, 2009, 58(1): 69-75
- [21] Han L L, Ding F. Multi-innovation stochastic gradient algorithms for multi-input multi-output systems [J]. Digital Signal Processing, 2009, 19(4): 545-554
- [22] Wang D Q, Ding F. Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems [J]. Digital Signal Processing, 2010, 20(3): 750-762
- [23] 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
- [24] Liu Y J, Yu L, Ding F. Multi-innovation extended stochastic gradient algorithm and its performance analysis [J]. Circuits, Systems and Signal Processing, 2010, 29(4): 649-667
- [25] Han L L, Ding F. Identification for multirate multi-input systems using the multi-innovation identification theory [J]. Computers and Mathematics with Applications, 2009, 57(9): 1438-1449
- [26] Xie L, Yang H Z, Ding F. Modeling and identification for non-uniformly periodically sampled-data systems [J]. IET Control Theory and Applications, 2010, 4(5): 784-794
- [27] Wang D Q, Chu Y Y, Ding F. Auxiliary model-based RELS and MI-ELS algorithms for Hammerstein OEMA systems [J]. Computers and Mathematics with Applications, 2010, 59(9): 3092-3098
- [28] Chen J, Ding F. Least squares and stochastic gradient parameter estimation for multivariable nonlinear Box-Jenkins models based on the auxiliary model and multi-innovation identification theory [J]. Engineering Computations, 2012, 29(8): 907-921.
- [29] 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.
- [30] 丁锋, 谢新民, 方崇智. 时变系统辨识的多新息方法 [J]. 自动化学报, 1996, 22(1): 85-91
DING Feng, XIE Xinmin, FANG Chongzhi. Multi-innovation identification methods for time-varying systems [J]. Acta Automatica Sinica, 1996, 22(1): 85-91
- [31] 丁锋, 萧德云, 丁韬. 多新息随机梯度辨识方法 [J]. 控制理论与应用, 2003, 20(6): 870-874
DING Feng, XIAO Deyun, DING Tao. Multi-innovation stochastic gradient identification methods [J]. Control Theory and Application, 2003, 20(6): 870-874
- [32] 王冬青, 丁锋. Box-Jenkins 模型的基于辅助模型的多新息广义增广随机梯度算法 [J]. 控制与决策, 2008, 23(9): 999-1003, 1010
WANG Dongqing, DING Feng. Auxiliary model based multi-innovation generalized extended stochastic gradient algorithms for Box-Jenkins models [J]. Control and Decision, 2008, 23(9): 999-1003, 1010
- [33] 丁洁, 谢莉, 丁锋. 非均匀采样系统多新息随机梯度辨识性能分析 [J]. 控制与决策, 2011, 26(9): 1338-1342.
DING Jie, XIE Li, DING Feng. Performance analysis of multi-innovation stochastic gradient identification for non-uniformly sampled systems [J]. Control and Decision, 2011, 26(9): 1338-1342.
- [34] Mao Y W, Ding F. Multi-innovation stochastic gradient identification for Hammerstein controlled autoregressive autoregressive systems based on the filtering technique [J]. Nonlinear Dynamics, 2015, 79(3): 1745-1755.
- [35] Ding F, Chen T. Identification of Hammerstein nonlinear ARMAX systems [J]. Automatica, 2005, 41(9): 1479-1489
- [36] Ding F, Shi Y, Chen T. Gradient-based identification methods for Hammerstein nonlinear ARMAX models [J]. Nonlinear Dynamics, 2006, 45(1-2): 31-43
- [37] 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
- [38] 范伟, 丁锋. Hammerstein 非线性系统参数估计分离的三种方法 [J]. 控制科学与工程, 2008, 8(6): 1586-1589.
FAN Wei, DING Feng. Three methods of separating parameters for Hammerstein nonlinear systems [J]. Science Technology and Engineering, 2008, 8(6): 1586-1589
- [39] Ding F, Duan H H. Two-stage parameter estimation algorithms for Box-Jenkins systems [J]. IET Signal Processing, 2013, 7(2): 176-184
- [40] 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.
- [41] 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.

Multi-innovation identification methods for input nonlinear equation-error systems

DING Feng^{1,2,3} CHEN Huibo¹

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 For input nonlinear equation-error systems (namely the input nonlinear controlled autoregressive (IN-CAR) systems), this paper studies and presents the over-parameterization model based multi-innovation identification (MI) methods, the over-parameterization model based hierarchical MI methods and the key term separation based MI methods, and uses the decomposition technique to present the key term separation based two-stage MI methods and the key term separation based three-stage MI methods. These methods can be extended to other input nonlinear equation-error systems, input nonlinear output-error type systems, output nonlinear equation-error type systems and output nonlinear output-error systems, and feedback nonlinear systems. Finally, the computational efficiency, the computational steps and the flowcharts of several typical identification algorithms are discussed.

Key words parameter estimation; recursive identification; gradient search; least squares; over-parameterization model; key term separation principle; model decomposition technique; auxiliary model identification ideal; multi-innovation identification theory; hierarchical identification principle; coupling identification concept; input nonlinear system; output nonlinear system