文章编号:1674-7070(2014)01-0001-16

丁锋[1,2,3]　汪菲菲[1]

多元系统耦合多新息随机梯度类辨识方法

摘要

针对多元线性回归系统,利用耦合辨识概念和多新息辨识理论,讨论了多元随机梯度算法、多元多新息随机梯度算法,以及变递推间隔多元多新息梯度算法,进一步分解多元系统为一些子系统,给出了耦合子系统随机梯度算法、耦合随机梯度算法、耦合子系统多新息随机梯度算法、耦合多新息随机梯度算法,并将这些方法推广到多元伪线性滑动平均系统和多元伪线性自回归滑动平均系统.文中给出了几个典型耦合随机梯度算法、耦合多新息随机梯度算法的计算步骤和示意图.

关键词

参数估计;递推辨识;梯度搜索;最小二乘;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;多元系统

中图分类号　TP273

文献标志码　A

收稿日期　2014-01-06

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

作者简介

丁锋,男,博士,教授,博士生导师,主要从事系统辨识、过程建模、自适应控制方面的研究. fding@ jiangnan.edu.cn

1　江南大学　物联网工程学院,无锡,214122

2　江南大学　控制科学与工程研究中心,无锡,214122

3　江南大学　教育部轻工过程先进控制重点实验室,无锡,214122

0　引言

辅助模型辨识思想、迭代搜索辨识原理、多新息辨识理论、递阶辨识原理、耦合辨识概念是新近提出的辨识理念和辨识方法研究思想[1-2],已在《南京信息工程大学学报》连载论文中进行了讨论[3-13].本文将耦合辨识概念与多新息辨识理论相结合,研究多元系统的辨识问题.

耦合辨识概念是笔者提出的一种新的辨识理念,它主要被用于研究各子系统间存在相同参数的多变量系统或多元系统的辨识问题[1,10,14-15],是实现子系统辨识方法间共同参数估计的连接而产生的耦合辨识方法.它的研究对象可以是结构复杂的参数耦合线性和非线性多变量系统,能够减小参数耦合大规模多变量系统辨识算法的计算量.耦合辨识方法分为部分参数耦合辨识方法和全部参数耦合辨识方法,简称为部分耦合辨识方法和全耦合辨识方法.全耦合辨识方法有时也简称为耦合辨识方法.耦合辨识概念可以与辅助模型辨识思想、多新息辨识理论、递阶辨识原理、迭代搜索原理、牛顿搜索原理等相结合,来研究线性或非线性多变量系统的耦合辨识问题.

耦合辨识概念是 20 余年前,笔者于 1988 年在清华大学攻读硕士学位时,学习萧德云教授讲授的《过程辨识》[16]中多变量线性过程的参数估计方法时,受到具有公分母特征值多项式传递函数描述的多变量系统递推最小二乘估计方法,以及子系统辨识方法的启发(后来建立了递阶辨识原理,并提出了一系列递阶辨识方法[17-21]),经过多年的思考和研究,提炼出多元系统辨识模型,在国际期刊《IEEE Transactions on Automatic Control》2010 年第 8 期上 发表论文“Partially coupled stochastic gradient identification methods for non-uniformly sampled systems(非均匀采样系统的部分参数耦合随机梯度辨识方法)”[14]后,逐步提炼和形成的.在此基础上,撰写了论文“Coupled least squares identification for multivariable systems(多变量系统的耦合最小二乘辨识)”,发表在国际期刊《IET Control Theory and Applications》2013 年第 1 期上[15],并在《南京信息工程大学学报》连载出版了题为“系统辨识(8):耦合辨识概念与方法”的论文[10].

多新息辨识方法能改进参数估计精度,通过扩展辨识新息,即从标量新息到新息向量,从向量新息到新息矩阵,是一种基于新息的辨识方法[1,22-23].根据多新息辨识理论,可以衍生出许多辨识方法,如多

新息随机梯度算法[23-26]、多新息最小二乘算法[27]以及多新息投影算法等,并且可以将它应用到观测器设计和卡尔曼滤波,导出多新息观测器和多新息卡尔曼滤波方法等[1,8],也同样可以将它应用于有色噪声干扰的方程误差类系统、输出误差类系统及非线性系统等[1,9].

文献[13]对多元系统和多变量系统耦合辨识模型进行了分类,并研究了多变量受控自回归系统的子系统递推最小二乘辨识算法和联合递推最小二乘辨识算法,研究了部分信息向量耦合型多变量系统的子系统最小二乘算法、基于块矩阵求逆的最小二乘辨识算法、子系统递推最小二乘辨识算法、基于辨识模型分解的递推最小二乘辨识算法.文献[1,10]研究了全部参数耦合多元系统的耦合最小二乘辨识方法和耦合随机梯度辨识方法、部分参数耦合多变量系统的部分耦合最小二乘辨识算法和部分耦合随机梯度辨识算法[14].文献[15]证明了多元系统全耦合最小二乘辨识方法的等价性和收敛性.进一步可以研究多元系统全耦合随机梯度算法、部分参数耦合最小二乘算法和部分耦合随机梯度算法及其收敛性.

本文针对全参数耦合的多元线性回归系统,讨论了多元随机梯度辨识算法、多元多新息随机梯度辨识算法,以及变递推间隔多元多新息梯度辨识算法;进一步将全参数耦合多元系统分解成 m 个子系统(m 为输出数目),从其子系统随机梯度辨识方法入手,推导出耦合子系统随机梯度辨识算法、耦合随机梯度辨识算法、耦合子系统多新息随机梯度辨识算法、耦合多新息随机梯度辨识算法;利用耦合辨识概念和多新息辨识理论,研究了多元伪线性滑动平均系统和多元伪线性自回归滑动平均系统的多元随机梯度类辨识算法、多元多新息随机梯度类辨识算法、耦合随机梯度类辨识算法和耦合多新息随机梯度类辨识算法.文中给出了耦合随机梯度算法、耦合增广随机梯度算法、耦合多新息广义增广随机梯度算法的计算步骤,给出了耦合随机梯度算法和参数耦合随机梯度算法示意图,使读者更加清晰理解辨识方法的机理.进一步可以研究多元系统耦合随机梯度算法、耦合最小二乘算法、耦合多新息随机梯度算法、耦合多新息最小二乘算法的计算效率及收敛性.

1 多元线性回归系统

1.1 多元随机梯度辨识算法

考虑下列多元线性回归系统(multivariate linear regressive system)[1,10,15]

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\theta}+\boldsymbol{v}(t), \tag{1}$$

其中 $\boldsymbol{y}(t):=[y_1(t),y_2(t),\cdots,y_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为 m 维系统输出向量,$\boldsymbol{\Phi}(t)\in\mathbf{R}^{m\times n}$ 是由系统输入输出数据构成的回归信息矩阵,$\boldsymbol{\theta}\in\mathbf{R}^n$ 是待辨识的系统参数向量,$\boldsymbol{v}(t):=[v_1(t),v_2(t),\cdots,v_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 是零均值白噪声向量.假设 $t\leqslant 0$ 时,$\boldsymbol{y}(t)=\mathbf{0}$,$\boldsymbol{\Phi}(t)=\mathbf{0}$ 和 $\boldsymbol{v}(t)=\mathbf{0}$.令 $\hat{\boldsymbol{\theta}}(t)$ 为参数向量 $\boldsymbol{\theta}$ 在时刻 t 的估计.

1) 多元随机梯度辨识算法

定义和极小化梯度准则函数(gradient criterion function)

$$J_1(\boldsymbol{\theta}):=\|\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\boldsymbol{\theta}\|^2,$$

可以得到估计参数向量 $\boldsymbol{\theta}$ 的多元随机梯度算法(Multivariate Stochastic Gradient algorithm, M-SG 算法)[2]:

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

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{2}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Phi}(t)\|^2,\quad r(0)=1. \tag{3}$$

2) 修正多元随机梯度辨识算法

为了提高收敛速度,引入收敛指数 ε,能够得到估计参数向量 $\boldsymbol{\theta}$ 的修正多元随机梯度算法(Modified Multivariate Stochastic Gradient algorithm, M-M-SG 算法)[2]:

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

$$\frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{4}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Phi}(t)\|^2,\quad r(0)=1. \tag{5}$$

当 $\varepsilon=1$ 时,修正多元随机梯度算法退化为 M-SG 算法.

3) 遗忘因子多元随机梯度辨识算法

为了提高随机梯度算法的暂态性能,引入遗忘因子 λ,能够得到带遗忘因子的多元随机梯度算法(FF-M-SG 算法)[2]:

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

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{6}$$

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

$$r(0)=1. \tag{7}$$

当 $\lambda=1$ 时,FF-M-SG 算法即为 M-SG 算法;当 $\lambda=0$ 时,FF-M-SG 算法退化为投影算法.

文献[2]研究了 M-SG 辨识算法参数估计的一致

收敛性,读者可以研究 M-M-SG 算法参数估计的一致收敛性和 FF-M-SG 算法参数估计误差的有界收敛性.

1.2 多元多新息随机梯度辨识算法

与多元递推最小二乘算法相比,多元随机梯度算法的计算量小,但是收敛速度慢.为了改进随机梯度算法的收敛速度,可引入新息长度,沿用标量代数到向量代数,从向量代数到矩阵论的发展历程.通过这种“类比”,将标量新息扩展为向量得到多新息随机梯度辨识方法,将新息向量扩展为一个“大新息向量”或新息矩阵,从多元随机梯度辨识算法可以得到多元多新息随机梯度辨识算法.

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

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

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

在系统辨识中 $\boldsymbol{e}(t):=\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}^m$ 称为新息向量.令 $p\geqslant 1$ 为新息长度,应用多新息辨识理论,将多元随机梯度算法中新息向量 $\boldsymbol{e}(t)\in\mathbf{R}^m$ 扩展为一个大新息向量:

$$\boldsymbol{E}(p,t):=\begin{bmatrix}\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\\ \boldsymbol{y}(t-1)-\boldsymbol{\Phi}(t-1)\hat{\boldsymbol{\theta}}(t-1)\\ \vdots\\ \boldsymbol{y}(t-p+1)-\boldsymbol{\Phi}(t-p+1)\hat{\boldsymbol{\theta}}(t-1)\end{bmatrix}\quad(8)$$

$$=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\theta}(t-1)\in\mathbf{R}^{mp}.\quad(9)$$

参考文献[1,23,25],可以得到新息长度为 p 的多元多新息随机梯度算法(Multivariate Multi-Innovation Stochastic Gradient algorithm,M-MISG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Gamma}(p,t)}{r(t)}\boldsymbol{E}(p,t),$$

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\quad(10)$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1),\quad(11)$$

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

$$\boldsymbol{Y}(p,t)=[\boldsymbol{y}^{\mathrm{T}}(t)\boldsymbol{y}^{\mathrm{T}}(t-1),\cdots\boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}},\quad(13)$$

$$\boldsymbol{\Gamma}(p,t)=[\boldsymbol{\Phi}^{\mathrm{T}}(t),\boldsymbol{\Phi}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{\Phi}^{\mathrm{T}}(t-p+1)].\quad(14)$$

当新息长度 $p=1$ 时,多元多新息随机梯度算法就退化为多元随机梯度算法.与多元随机梯度算法相比,多元多新息随机梯度算法在每步递推计算参数估计时,不但使用了当前的数据和新息,而且使用了过去的数据和新息.重复使用系统数据信息,这是多新息随机梯度算法能够改善算法收敛性和提高参数估计精度的原因.

当然,还可以推导修正多元多新息随机梯度辨识算法和遗忘因子多元多新息随机梯度辨识算法.

1.3 变递推间隔多元多新息梯度辨识算法

为了处理数据丢失的情况,定义一个整数序列(integer sequence)$\{t_s,s=0,1,2,\cdots\}$满足:

$$0=t_0<t_1<t_2<t_3<\cdots<t_{s-1}<t_s<\cdots,$$

且 $t_s^*:=t_s-t_{s-1}\geqslant 1$,使得 $t=t_s(s=1,2,\cdots)$ 时,$\boldsymbol{y}(t)$ 和 $\boldsymbol{\Phi}(t)$ 都可得到,即对任意 $s=1,2,3,\cdots,\boldsymbol{y}(t_s)$ 和 $\boldsymbol{\Phi}(t_s)$ 都可得到.

1) 变递推间隔多新息投影算法

参考文献[1,25],定义和使用梯度搜索极小化准则函数(criterion function)

$$J_2(\boldsymbol{\theta}):=\|\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\boldsymbol{\theta}\|^2,$$

能够得到变递推间隔多元多新息投影算法(interval-Varying Multivariate Multi-Innovation Projection algorithm,V-M-MI-Proj 算法)[1,25,27]:

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\mu(t_s)\boldsymbol{\Gamma}(p,t_s)\boldsymbol{E}(p,t_s),$$

$$s=1,2,3,\cdots,\quad(15)$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\quad(16)$$

$$\mu(t_s)=\frac{\|\boldsymbol{\Gamma}(p,t_s)\boldsymbol{E}(p,t_s)\|^2}{\|\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\boldsymbol{\Gamma}(p,t_s)\boldsymbol{E}(p,t_s)\|^2},\quad(17)$$

$$\boldsymbol{E}(p,t_s)=\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1}),\quad(18)$$

$$\boldsymbol{\Gamma}(p,t_s)=[\boldsymbol{\Phi}^{\mathrm{T}}(t_s),\boldsymbol{\Phi}^{\mathrm{T}}(t_s-1),\cdots,\boldsymbol{\Phi}^{\mathrm{T}}(t_s-p+1)],\quad(19)$$

$$\boldsymbol{Y}(p,t_s)=[\boldsymbol{y}^{\mathrm{T}}(t_s)\boldsymbol{y}^{\mathrm{T}}(t_s-1),\cdots\boldsymbol{y}^{\mathrm{T}}(t_s-p+1)]^{\mathrm{T}}.\quad(20)$$

式(16)表示在数据丢失的区间内,参数估计保持不变.这个算法计算收敛因子(步长)$\mu(t_s)$比较复杂.下面给出简化的变递推间隔多元多新息投影算法(V-M-MI-Proj)[1,25]:

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\Gamma}(p,t_s)}{\|\boldsymbol{\Gamma}(p,t_s)\|^2}\boldsymbol{E}(p,t_s),$$

$$s=1,2,3,\cdots,\quad(21)$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\quad(22)$$

$$\boldsymbol{E}(p,t_s)=\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1}),\quad(23)$$

$$\boldsymbol{\Gamma}(p,t_s)=[\boldsymbol{\Phi}^{\mathrm{T}}(t_s),\boldsymbol{\Phi}^{\mathrm{T}}(t_s-1),\cdots,\boldsymbol{\Phi}^{\mathrm{T}}(t_s-p+1)],\quad(24)$$

$$\boldsymbol{Y}(p,t_s)=[\boldsymbol{y}^{\mathrm{T}}(t_s)\boldsymbol{y}^{\mathrm{T}}(t_s-1),\cdots\boldsymbol{y}^{\mathrm{T}}(t_s-p+1)]^{\mathrm{T}}.\quad(25)$$

为了防止式(21)右边第二项分母为零,解决的方法之一是,当 $\|\boldsymbol{\Gamma}(p,t_s)\|^2=0$ 时,令 $\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})$,或者在式(21)右边第二项分母上加上一个正常数,将式(21)修改为

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\Gamma}(p,t_s)}{1+\|\boldsymbol{\Gamma}(p,t_s)\|^2}[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})].$$

变递推间隔多元多新息投影辨识算法对噪声敏感.下面是变递推间隔多新息广义投影辨识算法.

2）变递推间隔多新息广义投影算法

变递推间隔多元多新息投影辨识算法(21)—(25)可以推广为变递推间隔多元多新息广义投影算法(interval-Varying Multivariate Multi-Innovation Generalized Projection algorithm, V-M-MIGP 算法)[1,22,28]：

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\Gamma}(p,t_s)}{r(q,t_s)}[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],\quad s=1,2,3,\cdots,\tag{26}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\tag{27}$$

$$r(q,t_s)=\mathrm{tr}[\boldsymbol{\Gamma}(q,t_s)\boldsymbol{\Gamma}^{\mathrm{T}}(q,t_s)],\quad q\geqslant p,\tag{28}$$

$$\boldsymbol{\Gamma}(p,t_s)=[\boldsymbol{\Phi}^{\mathrm{T}}(t_s),\boldsymbol{\Phi}^{\mathrm{T}}(t_s-1),\cdots,\boldsymbol{\Phi}^{\mathrm{T}}(t_s-p+1)],\tag{29}$$

$$\boldsymbol{Y}(p,t_s)=[\boldsymbol{y}^{\mathrm{T}}(t_s),\boldsymbol{y}^{\mathrm{T}}(t_s-1),\cdots,\boldsymbol{y}^{\mathrm{T}}(t_s-p+1)]^{\mathrm{T}}.\tag{30}$$

变递推间隔多新息广义投影辨识算法可以通过加大 q 来克服噪声敏感性,选择适当的 q 可以减小时变多元系统参数估计误差上界.

3）变递推间隔多新息随机梯度算法

进一步可推广为变递推间隔多元多新息随机梯度算法(interval-Varying Multivariate Multi-Innovation Stochastic Gradient algorithm, V-M-MISG 算法)[1,22,25,28]：

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\boldsymbol{\Gamma}(p,t_s)}{r(t_s)}[\boldsymbol{Y}(p,t_s)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],\quad s=1,2,3,\cdots,\tag{31}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s=\{t_s,t_s+1,\cdots,t_{s+1}-1\},\tag{32}$$

$$r(t_s)=\|\boldsymbol{\Gamma}(s,t_s)\|^2=\sum_{i=0}^{s-1}\|\boldsymbol{\Phi}(t_s-i)\|^2,\tag{33}$$

$$\boldsymbol{\Gamma}(p,t_s)=[\boldsymbol{\Phi}^{\mathrm{T}}(t_s),\boldsymbol{\Phi}^{\mathrm{T}}(t_s-1),\cdots,\boldsymbol{\Phi}^{\mathrm{T}}(t_s-p+1)],\tag{34}$$

$$\boldsymbol{Y}(p,t_s)=[\boldsymbol{y}^{\mathrm{T}}(t_s),\boldsymbol{y}^{\mathrm{T}}(t_s-1),\cdots\boldsymbol{y}^{\mathrm{T}}(t_s-p+1)]^{\mathrm{T}}.\tag{35}$$

随着递推步数 s 的增加, $r(t_s)\to\infty$,算法增益矩阵 $\boldsymbol{\Gamma}(p,t_s)/r(t_s)$ 趋于零,故变递推间隔多新息随机梯度辨识算法没有跟踪时变参数的能力.

当然,还可以推导变递推间隔修正多元多新息随机梯度辨识算法和变递推间隔遗忘因子多元多新息随机梯度辨识算法.

2 耦合随机梯度算法与耦合多新息随机梯度算法

2.1 耦合子系统随机梯度辨识算法

令 $\boldsymbol{\phi}_i^{\mathrm{T}}(t)\in\mathbf{R}^{1\times n}$ 是 $\boldsymbol{\Phi}(t)$ 的第 i 行,即

$$\boldsymbol{\Phi}(t):=\begin{bmatrix}\boldsymbol{\phi}_1^{\mathrm{T}}(t)\\ \boldsymbol{\phi}_2^{\mathrm{T}}(t)\\ \vdots\\ \boldsymbol{\phi}_m^{\mathrm{T}}(t)\end{bmatrix}\in\mathbf{R}^{m\times n}.$$

式(1)可以分解为 m 个辨识模型(子系统)：

$$y_i(t)=\boldsymbol{\phi}_i^{\mathrm{T}}(t)\boldsymbol{\theta}+v_i(t),\quad i=1,2,\cdots,m,\tag{36}$$

所有子系统都包含一个共同的参数向量 $\boldsymbol{\theta}\in\mathbf{R}^n$.一般来说,其中一个子系统就能够辨识出参数向量 $\boldsymbol{\theta}$,但是,为了提高参数估计精度,必须利用所有子系统的观测数据来辨识 $\boldsymbol{\theta}$, 就导出了耦合子系统随机梯度算法.

1）耦合子系统随机梯度算法

根据梯度搜索原理,定义一个准则函数,从式(36)可以得到 m 个随机梯度算法(Stochastic Gradient algorithm, SG 算法),即子系统随机梯度算法(Subsystem Stochastic Gradient algorithm, S-SG 算法)：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{37}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1.\tag{38}$$

注意到每个子系统随机梯度算法(37)—(38)都计算了一次参数估计向量 $\hat{\boldsymbol{\theta}}(t)$,但是每个子系统的参数估计 $\hat{\boldsymbol{\theta}}(t)$ 是独立的,即子系统 i 的参数估计与子系统 j 的参数估计 $\hat{\boldsymbol{\theta}}(t)$ 无关($i\neq j$).为清晰起见,把式(37) 中第 i 个子系统的参数向量 $\hat{\boldsymbol{\theta}}(t)$ 记作 $\hat{\boldsymbol{\theta}}_i(t)$,因此第 i 个 S-SG 算法(37)—(38) 可以等价写为

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_i(t-1)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_i(t-1)],\quad \hat{\boldsymbol{\theta}}_i(0)=\mathbf{1}_n/p_0,\tag{39}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1.\tag{40}$$

由此可以看出:各个子系统参数估计 $\hat{\boldsymbol{\theta}}_i(t)$ 间没有任何耦合. 对于 $i=1,2,\cdots,m$, 可以从式(39)—(40) 获得 m 个参数估计向量 $\hat{\boldsymbol{\theta}}_i(t)$,它们都是同一参数向量 $\boldsymbol{\theta}$ 的估计,即每个 S-SG 算法都对 $\boldsymbol{\theta}$ 进行了估计,这导致大量冗余的参数估计 $\hat{\boldsymbol{\theta}}_i(t)$.一种办法是采用其平均估计作为 $\boldsymbol{\theta}$ 的估计,即：

$$\hat{\boldsymbol{\theta}}(t):=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}\in\mathbf{R}^n.\tag{41}$$

如果把式(41)的参数估计 $\hat{\boldsymbol{\theta}}(t)$ 仅仅作为参数向量 $\boldsymbol{\theta}$ 的输出,那么各 S-SG 辨识算法(39)—(40)仍是独立的.根据耦合辨识概念[1,10,14-15],用 $\hat{\boldsymbol{\theta}}(t-1)$ 代替 S-SG 算法中的 $\hat{\boldsymbol{\theta}}_i(t-1)$,就得到耦合子系统随

机梯度算法(Coupled Subsystem Stochastic Gradient algorithm,C-S-SG 算法):

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

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{42}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1, \tag{43}$$

$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{44}$$

2) 遗忘因子耦合子系统随机梯度算法

为了提高耦合子系统随机梯度算法(42)—(44)的收敛速度,引入遗忘因子(Forgetting Factor, FF) $0\leqslant\lambda\leqslant 1$,就得到遗忘因子耦合子系统随机梯度算法(遗忘因子 C-S-SG 算法)[17,23,29]:

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

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{45}$$

$$r_i(t)=\lambda r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1,$$

$$0\leqslant\lambda\leqslant 1, \tag{46}$$

$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{47}$$

3) 修正耦合子系统随机梯度算法

引入收敛指数(convergence index)ε,就得到修正耦合子系统随机梯度算法(修正 C-S-SG 算法)[30-31]:

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\phi}_i(t)}{r_i^{\varepsilon}(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{48}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1, \tag{49}$$

$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{50}$$

算法(39)—(40)的收敛性研究是简单的,因为它与标量系统随机梯度辨识算法性能类似. C-S-SG 算法(42)—(44)和修正 C-S-SG 算法(48)—(50)的参数估计一致收敛性,和遗忘因子 C-S-SG 算法(45)—(47)的参数估计误差有界收敛性都是有待研究的辨识课题.

2.2 耦合随机梯度辨识算法

1) 耦合随机梯度辨识算法

下面利用耦合辨识概念,基于耦合子系统递推随机梯度算法来推导耦合随机梯度算法,以避免冗余的参数估计.

对于递推参数估计算法,可以期望参数估计是收敛的,即假设参数估计随着数据长度 t 的增大而收敛于真参数.也就是说,可以认为第 $i-1$ 个子系统在时刻 t 的参数估计 $\hat{\boldsymbol{\theta}}_{i-1}(t)$ 比第 i 个子系统在时刻 $(t-1)$ 的参数估计 $\hat{\boldsymbol{\theta}}_i(t-1)$ 更接近真参数 $\boldsymbol{\theta}$[1,9,23,25,27,32-33].参考文献[1,15]中的耦合最小二乘辨识方法和文献[14]中的部分耦合随机梯度辨识方法,用 $\hat{\boldsymbol{\theta}}_m(t-1)$ 代替式(39)中 $i=1$ 时右边的 $\hat{\boldsymbol{\theta}}_1(t-1)$,用 $\hat{\boldsymbol{\theta}}_{i-1}(t)$ 代替式(39)右边的 $\hat{\boldsymbol{\theta}}_i(t-1)$,则得到下列耦合随机梯度算法(Coupled Stochastic Gradient algorithm,C-SG 算法)[1,15]:

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

$$\hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0, \tag{51}$$

$$r_1(t)=r_m(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad r_m(0)=1, \tag{52}$$

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{i-1}(t)], \tag{53}$$

$$r_i(t)=r_{i-1}(t)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m. \tag{54}$$

上述算法中,$\hat{\boldsymbol{\theta}}_i(t)\in\mathbf{R}^n$ 为第 i 个子系统在时刻 t 的参数估计向量,子系统辨识算法间参数估计 $\hat{\boldsymbol{\theta}}_i(t)$ 和 $r_i(t)$ 都是耦合的,其示意如图 1 所示.时刻 t 第 m 个子系统的参数估计定义为系统的参数估计: $\hat{\boldsymbol{\theta}}(t):=\hat{\boldsymbol{\theta}}_m(t)$.

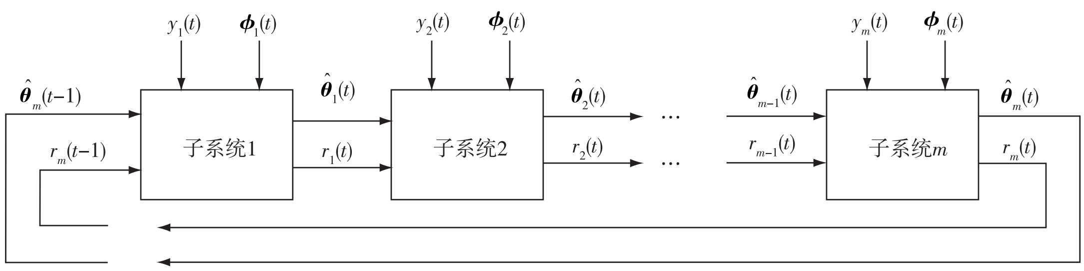

图 1 耦合随机梯度算法的示意

Fig. 1 The schematic diagram of the coupled stochastic gradient algorithm

C-SG 算法(51)—(54)计算参数估计向量 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}_m(t)$ 的步骤如下:

① 置初值:令 $t=1$, $\hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0$, $r_m(0)=1$, $p_0=10^6$.

② 收集观测数据 $\boldsymbol{y}(t)$ 和 $\boldsymbol{\Phi}(t)$,设 $\boldsymbol{\phi}_i^{\mathrm{T}}(t)\in\mathbf{R}^{1\times n}$ 为 $\boldsymbol{\Phi}(t)$ 的第 i 行.

③ 用式(52)计算数据 $r_1(t)$,用式(51)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_1(t)$.

④ 依次当 $i=2,3,\cdots,m$ 时,用式(54)计算 $r_i(t)$,用式(53)刷新参数估计向量 $\hat{\boldsymbol{\theta}}_i(t)$.

⑤ 获得参数估计 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}_m(t)$. t 增 1,转到第②步.

如果耦合随机梯度辨识算法(51)—(54)采用相同的 $r_i(t)=r(t)$,就得到

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\phi}_1(t)}{r(t)}[y_1(t)-\boldsymbol{\phi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{55}$$

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{i-1}(t)],\quad i=2,3,\cdots,m,\tag{56}$$

$$r(t)=r(t-1)+\sum_{i=1}^{m}\|\boldsymbol{\phi}_i(t)\|^2,\quad r(0)=1.\tag{57}$$

读者可以研究这个算法与多元随机梯度算法(2)—(3)之间的联系与区别.与多元随机梯度算法(2)—(3)等价的算法如下[15]:

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\phi}_1(t)}{r(t)}[y_1(t)-\boldsymbol{\phi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{58}$$

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],\quad i=2,3,\cdots,m,\tag{59}$$

$$r(t)=r(t-1)+\sum_{i=1}^{m}\|\boldsymbol{\phi}_i(t)\|^2,\quad r(0)=1.\tag{60}$$

注意到耦合随机梯度辨识算法(51)—(54)中,参数估计 $\hat{\boldsymbol{\theta}}_i(t)$ 和 $r_i(t)$ 都是耦合的,去掉 $r_i(t)$ 间的耦合,便得到一个简单的耦合随机梯度算法:

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\phi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\phi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{61}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad r_1(0)=1,\tag{62}$$

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{i-1}(t)],\tag{63}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1,\quad i=2,3,\cdots,m.\tag{64}$$

因为去掉了 $r_i(t)$ 间的耦合,这个耦合算法的收敛速度会更快,源于增益向量 $\boldsymbol{\phi}_1(t)/r_1(t)$ 和 $\boldsymbol{\phi}_i(t)/r_i(t)$ 趋于零的速度会减慢.耦合随机梯度辨识算法(61)—(64)的示意如图 2 所示,输出的参数估计为 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}_m(t)$.

2)遗忘因子耦合随机梯度辨识算法

在耦合随机梯度算法(51)—(54)中引入遗忘因子 λ,就得到遗忘因子耦合随机梯度算法(遗忘因子 C-SG 算法):

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\phi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\phi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{65}$$

$$r_1(t)=\lambda r_m(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r_m(0)=1,\tag{66}$$

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{i-1}(t)],\tag{67}$$

$$r_i(t)=\lambda r_{i-1}(t)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m.\tag{68}$$

在简单的耦合随机梯度算法(61)—(64)中引入遗忘因子 λ,就得到简单的遗忘因子耦合随机梯度算法(遗忘因子 C-SG 算法):

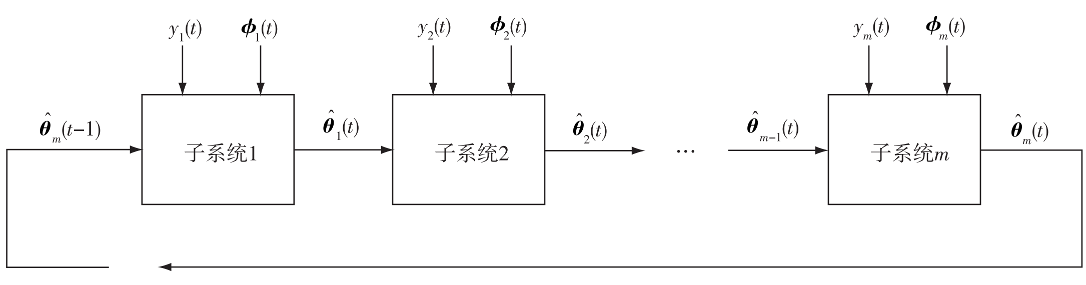

图 2 参数估计耦合随机梯度辨识算法示意

Fig. 2 The schematic diagram of the parameter estimation coupled least squares algorithm

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\phi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\phi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_m(t-1)],$$
$$\hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0, \tag{69}$$
$$r_1(t)=\lambda r_1(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,$$
$$r_1(0)=1, \tag{70}$$
$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\phi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}_{i-1}(t)], \tag{71}$$
$$r_i(t)=\lambda r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1,$$
$$i=2,3,\cdots,m. \tag{72}$$

当然,也可采用不同的遗忘因子 λ_i,也可在耦合随机梯度算法(51)—(54)和(61)—(64)中引入收敛指数,得到相应的修正 C-SG 算法.这些算法的收敛性都是有待研究的课题.

2.3 耦合子系统多新息随机梯度辨识算法

1) 耦合子系统多新息随机梯度算法

借助于多新息辨识理论[1,23],比较多新息随机梯度辨识算法与耦合子系统随机梯度辨识算法的结构形式,从耦合子系统随机梯度(C-S-SG)算法(42)—(44)可以得到耦合子系统多新息随机梯度辨识算法.定义子系统堆积信息矩阵 $\boldsymbol{\Gamma}_i(p,t)$,堆积输出向量 $\boldsymbol{Y}_i(p,t)$ 和新息向量 $\boldsymbol{E}_i(p,t)$ 如下:

$$\boldsymbol{\Gamma}_i(p,t):=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)]\in\mathbf{R}^{n\times p},$$
$$\boldsymbol{Y}_i(p,t):=\begin{bmatrix}y_i(t)\\ y_i(t-1)\\ \vdots\\ y_i(t-p+1)\end{bmatrix}\in\mathbf{R}^p,$$
$$\boldsymbol{E}_i(p,t):=\begin{bmatrix}y_i(t)-\boldsymbol{\phi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\\ y_i(t-1)-\boldsymbol{\phi}_i^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}(t-1)\\ \vdots\\ y_i(t-p+1)-\boldsymbol{\phi}_i^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}(t-1)\end{bmatrix}\in\mathbf{R}^p.$$

则可以提出耦合子系统多新息随机梯度算法(Coupled Subsystem Multi-Innovation Stochastic Gradient algorithm,C-S-MISG 算法):

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t),\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{73}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{74}$$
$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1, \tag{75}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)], \tag{76}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{77}$$
$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{78}$$

当然,也可以进一步引入遗忘因子或收敛指数,得到相应的遗忘因子耦合子系统多新息随机梯度辨识算法、修正耦合子系统多新息随机梯度辨识算法等.下面直接给出这些算法.

2) 遗忘因子耦合子系统多新息随机梯度算法(FF-C-S-MISG 算法)

$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t),\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{79}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{80}$$
$$r_i(t)=\lambda r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,$$
$$r_i(0)=1, \tag{81}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)], \tag{82}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{83}$$
$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{84}$$

3) 修正耦合子系统多新息随机梯度算法(M-C-S-MISG 算法)

$$\boldsymbol{\theta}_i(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i^{\varepsilon}(t)}\boldsymbol{E}_i(p,t),\quad \frac{1}{2}<\varepsilon\leqslant 1,$$
$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{85}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{86}$$
$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad r_i(0)=1, \tag{87}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)], \tag{88}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{89}$$
$$\hat{\boldsymbol{\theta}}(t)=\frac{\hat{\boldsymbol{\theta}}_1(t)+\hat{\boldsymbol{\theta}}_2(t)+\cdots+\hat{\boldsymbol{\theta}}_m(t)}{m}. \tag{90}$$

2.4 耦合多新息随机梯度辨识算法

1) 耦合多新息随机梯度算法

借助于多新息辨识理论,从耦合随机梯度辨识算法(61)—(64)可以得到耦合多新息随机梯度算法(Coupled Multi-Innovation Stochastic Gradient algorithm,C-MISG 算法):

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\Gamma}_1(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t),\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0, \tag{91}$$
$$\boldsymbol{E}_1(p,t)=\boldsymbol{Y}_1(p,t)-\boldsymbol{\Gamma}_1^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_m(t-1), \tag{92}$$
$$r_1(t)=r_m(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad r_m(0)=1, \tag{93}$$
$$\boldsymbol{\Gamma}_1(p,t)=[\boldsymbol{\phi}_1(t),\boldsymbol{\phi}_1(t-1),\cdots,\boldsymbol{\phi}_1(t-p+1)], \tag{94}$$
$$\boldsymbol{Y}_1(p,t)=[y_1(t),y_1(t-1),\cdots,y_1(t-p+1)]^{\mathrm{T}}, \tag{95}$$
$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t), \tag{96}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{i-1}(t), \tag{97}$$

$$r_i(t)=r_{i-1}(t)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m,\tag{98}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)],\tag{99}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}.\tag{100}$$

这里 $\hat{\boldsymbol{\theta}}_i(t)\in\mathbf{R}^n$ 为第 i 个子系统在时刻 t 的参数估计向量,系统的参数估计定义为 $\hat{\boldsymbol{\theta}}(t):=\hat{\boldsymbol{\theta}}_m(t)$. 将式(93)和(98)分别修改为

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad r_1(0)=1,\tag{101}$$
$$r_i(t)=r_i(t-1)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m,\tag{102}$$

就得到收敛更快的简单耦合多新息随机梯度算法.

进一步引入遗忘因子或收敛指数,得到相应的遗忘因子耦合多新息随机梯度辨识算法、修正耦合多新息随机梯度辨识算法等.下面直接给出这些算法.

2) 遗忘因子耦合多新息随机梯度算法(FF-C-MISG 算法)

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\Gamma}_1(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t),\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{103}$$
$$\boldsymbol{E}_1(p,t)=\boldsymbol{Y}_1(p,t)-\boldsymbol{\Gamma}_1^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_m(t-1),\tag{104}$$
$$r_1(t)=\lambda r_m(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r_m(0)=1,\tag{105}$$
$$\boldsymbol{\Gamma}_1(p,t)=[\boldsymbol{\phi}_1(t),\boldsymbol{\phi}_1(t-1),\cdots,\boldsymbol{\phi}_1(t-p+1)],\tag{106}$$
$$\boldsymbol{Y}_1(p,t)=[y_1(t),y_1(t-1),\cdots,y_1(t-p+1)]^{\mathrm{T}},\tag{107}$$
$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t),\tag{108}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{i-1}(t),\tag{109}$$
$$r_i(t)=\lambda r_{i-1}(t)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m,\tag{110}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)],\tag{111}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}.\tag{112}$$

3) 修正耦合多新息随机梯度算法(M-C-MISG 算法)

$$\hat{\boldsymbol{\theta}}_1(t)=\hat{\boldsymbol{\theta}}_m(t-1)+\frac{\boldsymbol{\Gamma}_1(p,t)}{r_1^{\varepsilon}(t)}\boldsymbol{E}_1(p,t),\quad \frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\theta}}_m(0)=\mathbf{1}_n/p_0,\tag{113}$$
$$\boldsymbol{E}_1(p,t)=\boldsymbol{Y}_1(p,t)-\boldsymbol{\Gamma}_1^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_m(t-1),\tag{114}$$
$$r_1(t)=r_m(t-1)+\|\boldsymbol{\phi}_1(t)\|^2,\quad r_m(0)=1,\tag{115}$$
$$\boldsymbol{\Gamma}_1(p,t)=[\boldsymbol{\phi}_1(t),\boldsymbol{\phi}_1(t-1),\cdots,\boldsymbol{\phi}_1(t-p+1)],\tag{116}$$
$$\boldsymbol{Y}_1(p,t)=[y_1(t),y_1(t-1),\cdots,y_1(t-p+1)]^{\mathrm{T}},\tag{117}$$
$$\hat{\boldsymbol{\theta}}_i(t)=\hat{\boldsymbol{\theta}}_{i-1}(t)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i^{\varepsilon}(t)}\boldsymbol{E}_i(p,t),\tag{118}$$
$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}_{i-1}(t),\tag{119}$$
$$r_i(t)=r_{i-1}(t)+\|\boldsymbol{\phi}_i(t)\|^2,\quad i=2,3,\cdots,m,\tag{120}$$
$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\phi}_i(t),\boldsymbol{\phi}_i(t-1),\cdots,\boldsymbol{\phi}_i(t-p+1)],\tag{121}$$
$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}.\tag{122}$$

3 多元伪线性滑动平均系统

一个有色噪声干扰的多变量系统可以参数化为多变量伪线性回归模型,其特点是信息矩阵中包含不可测噪声项.在辨识算法中,这些未知噪声项一般使用系统参数估计和系统的输入输出数据进行估算,并用估计的噪声项代替信息向量(或信息矩阵)中的不可测噪声项[2].基于这一辨识思想,这里研究多元伪线性滑动平均系统的多元增广随机梯度辨识算法、多元多新息增广随机梯度辨识算法、耦合增广随机梯度辨识算法、耦合多新息增广随机梯度算法等.

3.1 多元增广随机梯度辨识算法

考虑多元伪线性滑动平均系统[1,2,15]

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta}+d(z)\boldsymbol{v}(t),\tag{123}$$

其中 $\boldsymbol{y}(t)=[y_1(t),y_2(t),\cdots,y_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为观测输出向量,$\boldsymbol{\Phi}_{\mathrm{s}}(t)\in\mathbf{R}^{m\times n}$ 是由输入输出数据构成的回归信息矩阵,$\boldsymbol{\theta}\in\mathbf{R}^n$ 为系统参数向量,$\boldsymbol{v}(t)\in\mathbf{R}^m$ 为 m 维干扰白噪声向量,$d(z)$ 为单位后移算子 $z^{-1}[z^{-1}y(t)=y(t-1)]$ 的首一多项式:

$$d(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d}\in\mathbf{R}.$$

假设 m,n 和 n_d 已知,且当 $t\leqslant 0$ 时,$\boldsymbol{\Phi}_{\mathrm{s}}(t)=\mathbf{0}$, $\boldsymbol{v}(t)=\mathbf{0}$. 辨识的目标就是利用观测数据 $\{\boldsymbol{y}(t),\boldsymbol{\Phi}_{\mathrm{s}}(t):t=1,2,3,\cdots\}$ 估计系统模型参数向量 $\boldsymbol{\theta}$ 以及噪声模型的参数 d_i.

伪线性滑动平均系统(123)可以写为多元伪线性辨识模型(multivariate pseudo-linear identification model):

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\vartheta}+\boldsymbol{v}(t),\tag{124}$$

其中信息矩阵 $\boldsymbol{\Phi}(t)$ 和参数向量 $\boldsymbol{\vartheta}$ 定义如下:

$$\boldsymbol{\Phi}(t):=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\boldsymbol{v}(t-1),\boldsymbol{v}(t-2),\cdots,\boldsymbol{v}(t-n_d)]\in\mathbf{R}^{m\times n_0},\quad n_0:=n+n_d,\tag{125}$$
$$\boldsymbol{\vartheta}:=[\boldsymbol{\theta}^{\mathrm{T}},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_0}.\tag{126}$$

在辨识模型(124)中,新的参数向量 $\boldsymbol{\vartheta}$ 不仅包含系统模型参数向量 $\boldsymbol{\theta}$,而且包括噪声模型参数 d_i $(i=1,2,\cdots,n_d)$.

伪线性回归模型是指 $\boldsymbol{\Phi}(t)$ 中除了已知的输入输出数据外,还包含一些未知变量,如上面 $\boldsymbol{\Phi}(t)$ 中的未知噪声项 $\boldsymbol{v}(t-i)$,而这些未知变量可以通过输入输出数据估算得到.

1) 多元增广随机梯度辨识算法

$\boldsymbol{\Phi}(t)$ 中包含的未知噪声回归项 $\boldsymbol{v}(t-i)$ 用其估计值 $\hat{\boldsymbol{v}}(t-i)$ 代替,代替后的信息矩阵记作 $\boldsymbol{\Psi}(t)$,在辨识算法中使用 $\boldsymbol{\Psi}(t)$ 代替未知的 $\boldsymbol{\Phi}(t)$,能够获得估计伪线性回归辨识模型(124)参数向量 $\boldsymbol{\vartheta}$ 的多元增广随机梯度算法(Multivariate Extended Stochastic Gradient algorithm,M-ESG 算法)[2]:

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Psi}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{127}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{128}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{129}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0. \tag{130}$$

2)修正多元增广随机梯度辨识算法

引入收敛指数 ε,可以得到估计伪线性回归辨识模型(124)参数向量 $\boldsymbol{\vartheta}$ 的修正多元增广随机梯度算法(Modified Multivariate Extended Stochastic Gradient algorithm,M-M-ESG 算法)[2]:

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Psi}^{\mathrm{T}}(t)}{r^{\varepsilon}(t)}[\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\quad \frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{131}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{132}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{133}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0. \tag{134}$$

收敛指数的引入使得算法增益矩阵 $\boldsymbol{\Psi}^{\mathrm{T}}(t)/r^{\varepsilon}(t)$ 减小变慢,从而提高算法的收敛速度.当收敛指数 $\varepsilon=1$ 时,M-M-ESG 算法退化为 M-ESG 辨识算法.

3)遗忘因子多元增广随机梯度辨识算法

引入遗忘因子 λ,可以得到估计伪线性回归辨识模型(124)参数向量 $\boldsymbol{\vartheta}$ 的遗忘因子多元增广随机梯度算法(Forgetting Factor Multivariate Extended Stochastic Gradient algorithm,FF-M-ESG 算法)[2]:

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Psi}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{135}$$

$$r(t)=\lambda r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r(0)=1, \tag{136}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{137}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0. \tag{138}$$

遗忘因子的引入使得旧数据信息对 $r(t)$ 的贡献减小,算法增益矩阵 $\boldsymbol{\Psi}^{\mathrm{T}}(t)/r(t)$ 衰减变慢,因而可以提高算法启动阶段暂态收敛速度.当遗忘因子 $\lambda=0$ 时,FF-M-ESG 算法退化为多元增广投影辨识算法;当遗忘因子 $\lambda=1$ 时,FF-M-ESG 算法退化为 M-ESG 算法.

3.2 多元多新息增广随机梯度辨识算法

1)多元多新息增广随机梯度辨识算法

借助于多新息辨识理论[1,23,25],从多元增广随机梯度辨识算法(127)—(130)可以得到新息长度为 p 的多元多新息增广随机梯度算法(Multivariate Multi-Innovation Extended Stochastic Gradient algorithm,M-MI-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{139}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{140}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{141}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{142}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{143}$$

$$\boldsymbol{Y}(p,t)=[\boldsymbol{y}^{\mathrm{T}}(t),\boldsymbol{y}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}}, \tag{144}$$

$$\boldsymbol{\Gamma}(p,t)=[\boldsymbol{\Psi}^{\mathrm{T}}(t),\boldsymbol{\Psi}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)]. \tag{145}$$

当新息长度 $p=1$ 时,多元多新息增广随机梯度算法退化为多元增广随机梯度算法.

2)修正多元多新息增广随机梯度辨识算法

引入收敛指数 ε,从多元多新息增广随机梯度辨识算法(139)—(145),可以得到估计伪线性回归辨识模型(124)参数向量 $\boldsymbol{\vartheta}$ 的修正多元多新息增广随机梯度算法(Modified Multivariate Multi-Innovation Extended Stochastic Gradient algorithm,M-M-MI-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}(p,t)}{r^{\varepsilon}(t)}\boldsymbol{E}(p,t),\quad \frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{146}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{147}$$

$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{148}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{149}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{150}$$

$$\boldsymbol{Y}(p,t)=[\boldsymbol{y}^{\mathrm{T}}(t),\boldsymbol{y}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}}, \tag{151}$$

$$\boldsymbol{\Gamma}(p,t)=[\boldsymbol{\Psi}^{\mathrm{T}}(t),\boldsymbol{\Psi}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)]. \tag{152}$$

3)遗忘因子多元多新息增广随机梯度辨识算法

引入遗忘因子 λ,从多元多新息增广随机梯度辨识算法(139)—(145),可以得到估计伪线性回归辨识模型(124)参数向量 $\boldsymbol{\vartheta}$ 的遗忘因子多元多新息增广随机梯度算法(Forgetting Factor Multivariate Multi-Innovation Extended Stochastic Gradient algorithm,FF-M-MI-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0,\tag{153}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1),\tag{154}$$

$$r(t)=\lambda r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r(0)=1,\tag{155}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)],\tag{156}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0,\tag{157}$$

$$\boldsymbol{Y}(p,t)=[\boldsymbol{y}^{\mathrm{T}}(t),\boldsymbol{y}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}},\tag{158}$$

$$\boldsymbol{\Gamma}(p,t)=[\boldsymbol{\Psi}^{\mathrm{T}}(t),\boldsymbol{\Psi}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)].\tag{159}$$

4 耦合增广随机梯度算法与耦合多新息增广随机梯度算法

考虑多元伪线性滑动平均系统(123)对应的多元伪线性辨识模型(124),重写如下:

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\vartheta}+\boldsymbol{v}(t),\tag{160}$$

其中各变量的定义同上.

4.1 耦合子系统增广随机梯度辨识算法

令 $\boldsymbol{\psi}_i^{\mathrm{T}}(t)\in\mathbf{R}^{1\times n_0}$ 为 $\boldsymbol{\Psi}(t)$ 的第 i 行,即

$$\boldsymbol{\Psi}(t):=\begin{bmatrix}\boldsymbol{\psi}_1^{\mathrm{T}}(t)\\ \boldsymbol{\psi}_2^{\mathrm{T}}(t)\\ \vdots\\ \boldsymbol{\psi}_m^{\mathrm{T}}(t)\end{bmatrix}\in\mathbf{R}^{m\times n_0}.$$

仿照耦合子系统随机梯度算法(42)—(45)的推导,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的耦合子系统增广随机梯度算法(Coupled Subsystem Extended Stochastic Gradient algorithm, C-S-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0,\tag{161}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2,\quad r_i(0)=1,\tag{162}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)],\tag{163}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}},\tag{164}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0,\tag{165}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\frac{\hat{\boldsymbol{\vartheta}}_1(t)+\hat{\boldsymbol{\vartheta}}_2(t)+\cdots+\hat{\boldsymbol{\vartheta}}_m(t)}{m}.\tag{166}$$

值得指出的是,本文的算法都可以引入遗忘因子和收敛指数,得到相应的遗忘因子辨识算法和修正辨识算法,这里不一一讨论.

4.2 耦合增广随机梯度辨识算法

为避免冗余的参数估计,下面利用耦合辨识概念,基于耦合子系统递推随机梯度算法,类比耦合随机梯度算法(具体步骤从略),得到下列耦合增广随机梯度算法(Coupled Extended Stochastic Gradient algorithm, C-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}_1(t)=\hat{\boldsymbol{\vartheta}}_m(t-1)+\frac{\boldsymbol{\psi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\psi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_m(t-1)],\quad \hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0,\tag{167}$$

$$r_1(t)=r_m(t-1)+\|\boldsymbol{\psi}_1(t)\|^2,\quad r_m(0)=1,\tag{168}$$

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_{i-1}(t)],\tag{169}$$

$$r_i(t)=r_{i-1}(t)+\|\boldsymbol{\psi}_i(t)\|^2,\quad i=2,3,\cdots,m,\tag{170}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)],\tag{171}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}},\tag{172}$$

$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0.\tag{173}$$

这里 $\hat{\boldsymbol{\vartheta}}_i(t)\in\mathbf{R}^{n_0}$ 为第 i 个子系统在时刻 t 的参数估计向量,系统的参数估计定义为 $\hat{\boldsymbol{\vartheta}}(t):=\hat{\boldsymbol{\vartheta}}_m(t)$.

C-ESG 辨识算法(167)—(173)计算参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t)$ 的步骤如下:

① 置初值:令 $t=1$, $\hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0$, $r_m(0)=1$, $\hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0\,(j\leqslant 0)$, $p_0=10^6$.

② 收集观测数据 $\boldsymbol{y}(t)$ 和 $\boldsymbol{\Phi}_{\mathrm{s}}(t)$,根据式(171)构造 $\boldsymbol{\Psi}(t)$,从式(172) $\boldsymbol{\Psi}(t)$ 中提取信息向量 $\boldsymbol{\psi}_i(t)$, $i=1,2,\cdots,m$.

③ 用式(168)计算数据 $r_1(t)$,用式(167)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}_1(t)$.

④ 依次当 $i=2,3,\cdots,m$ 时,用式(170)计算数据 $r_i(t)$,用式(169)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}_i(t)$.

⑤ 获得参数估计 $\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t)$,用式(173)计算估计残差 $\hat{\boldsymbol{v}}(t)$. t 增 1,转到第 ② 步.

注意到耦合增广随机梯度辨识算法(167)—(173)中,参数估计 $\hat{\boldsymbol{\vartheta}}_i(t)$ 和 $r_i(t)$ 都是耦合的,去掉 $r_i(t)$ 间的耦合,便得到一个简单的耦合增广随机梯度算法:

$$\hat{\boldsymbol{\vartheta}}_1(t)=\hat{\boldsymbol{\vartheta}}_m(t-1)+\frac{\boldsymbol{\psi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\psi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_m(t-1)],\quad \hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0,\tag{174}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\psi}_1(t)\|^2,\quad r_1(0)=1,\tag{175}$$

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_{i-1}(t)],\tag{176}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2,\quad r_i(0)=1,\quad i=2,3,\cdots,m,\tag{177}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)],\tag{178}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{179}$$

$$\hat{v}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{v}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0. \tag{180}$$

4.3 耦合子系统多新息增广随机梯度辨识算法

借助于多新息辨识理论,从耦合子系统增广随机梯度算法(161)—(166),可以得到耦合子系统多新息增广随机梯度算法(Coupled Subsystem Multi-Innovation Extended Stochastic Gradient algorithm,C-S-MI-ESG 算法):

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t),$$

$$\hat{\boldsymbol{\vartheta}}_i(0)=\mathbf{1}_{n_0}/p_0, \tag{181}$$

$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{182}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2,\quad r_i(0)=1, \tag{183}$$

$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\psi}_i(t),\boldsymbol{\psi}_i(t-1),\cdots,\boldsymbol{\psi}_i(t-p+1)], \tag{184}$$

$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{185}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)], \tag{186}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{187}$$

$$\hat{v}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{v}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{188}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\frac{\hat{\boldsymbol{\vartheta}}_1(t)+\hat{\boldsymbol{\vartheta}}_2(t)+\cdots+\hat{\boldsymbol{\vartheta}}_m(t)}{m}. \tag{189}$$

当然,也可以进一步引入遗忘因子或收敛指数,得到相应的遗忘因子耦合子系统多新息增广随机梯度辨识算法、修正耦合子系统多新息增广随机梯度辨识算法等.

4.4 耦合多新息增广随机梯度辨识算法

借助于多新息辨识理论,从耦合增广随机梯度辨识算法(174)—(180)可以得到耦合多新息增广随机梯度算法(Coupled Multi-Innovation Extended Stochastic Gradient algorithm,C-MI-ESG 算法):

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

$$\hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0, \tag{190}$$

$$\boldsymbol{E}_1(p,t)=\boldsymbol{Y}_1(p,t)-\boldsymbol{\Gamma}_1^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}_m(t-1), \tag{191}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\psi}_1(t)\|^2,\quad r_1(0)=1, \tag{192}$$

$$\boldsymbol{\Gamma}_1(p,t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_1(t-1),\cdots,\boldsymbol{\psi}_1(t-p+1)], \tag{193}$$

$$\boldsymbol{Y}_1(p,t)=[y_1(t),y_1(t-1),\cdots,y_1(t-p+1)]^{\mathrm{T}}, \tag{194}$$

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t), \tag{195}$$

$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}_{i-1}(t), \tag{196}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2,\quad r_i(0)=1,$$

$$i=2,3,\cdots,m, \tag{197}$$

$$\boldsymbol{\Gamma}_i(p,t)=[\boldsymbol{\psi}_i(t),\boldsymbol{\psi}_i(t-1),\cdots,\boldsymbol{\psi}_i(t-p+1)], \tag{198}$$

$$\boldsymbol{Y}_i(p,t)=[y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{199}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)], \tag{200}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{201}$$

$$\hat{v}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{v}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0. \tag{202}$$

这里 $\hat{\boldsymbol{\vartheta}}_i(t)\in\mathbf{R}^{n_0}$ 为第 i 个子系统在时刻 t 的参数估计向量,

系统的参数估计定义为 $\hat{\boldsymbol{\vartheta}}(t):=\hat{\boldsymbol{\vartheta}}_m(t)$.

5 多元伪线性自回归滑动平均系统

本节简单讨论多元伪线性自回归滑动平均系统的多元广义增广随机梯度辨识算法、多元多新息广义增广随机梯度辨识算法、耦合广义增广随机梯度辨识算法、耦合多新息广义增广随机梯度算法等.

5.1 多元广义增广随机梯度辨识算法

考虑多元伪线性自回归滑动平均系统[1,2,15]:

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta}+\frac{d(z)}{c(z)}\boldsymbol{v}(t), \tag{203}$$

其中 $\boldsymbol{y}(t)=[y_1(t),y_2(t),\cdots,y_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为观测输出向量,$\boldsymbol{\Phi}_{\mathrm{s}}(t)\in\mathbf{R}^{m\times n}$ 是由输入输出数据构成的回归信息矩阵,$\boldsymbol{\theta}\in\mathbf{R}^n$ 为系统参数向量,$\boldsymbol{v}(t)\in\mathbf{R}^m$ 为 m 维干扰白噪声向量,$c(z)$ 和 $d(z)$ 均为单位后移算子 z^{-1} 的首一多项式:

$$c(z):=1+c_1z^{-1}+c_2z^{-2}+\cdots+c_{n_c}z^{-n_c}\in\mathbf{R},$$

$$d(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d}\in\mathbf{R}.$$

假设 m,n,n_c 和 n_d 已知,记 $n_0:=n+n_c+n_d$,且当 $t\leqslant 0$ 时,$\boldsymbol{y}(t)=\mathbf{0},\boldsymbol{\Phi}_{\mathrm{s}}(t)=\mathbf{0},\boldsymbol{v}(t)=\mathbf{0}$.辨识的目标就是利用观测数据 $\{\boldsymbol{y}(t),\boldsymbol{\Phi}_{\mathrm{s}}(t):t=1,2,3,\cdots\}$ 估计系统模型参数向量 $\boldsymbol{\theta}$,以及噪声模型的参数 c_i 和 d_i.
令

$$\boldsymbol{w}(t):=\frac{d(z)}{c(z)}\boldsymbol{v}(t)\in\mathbf{R}^m.$$

定义信息矩阵 $\boldsymbol{\Phi}(t)$ 和参数向量 $\boldsymbol{\vartheta}$ 如下:

$$\boldsymbol{\Phi}(t):=[\boldsymbol{\Phi}_{\mathrm{s}}(t),-\boldsymbol{w}(t-1),-\boldsymbol{w}(t-2),\cdots,-\boldsymbol{w}(t-n_c),$$
$$\boldsymbol{v}(t-1),\boldsymbol{v}(t-2),\cdots,\boldsymbol{v}(t-n_d)]\in\mathbf{R}^{m\times n_0},$$

$$\boldsymbol{\vartheta}:=[\boldsymbol{\theta}^{\mathrm{T}},c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_0}.$$

由伪线性自回归滑动平均系统(203)可以得到多元伪线性辨识模型:

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\vartheta}+\boldsymbol{v}(t). \tag{204}$$

这里信息矩阵 $\boldsymbol{\Phi}(t)$ 中除了已知的输入输出数据 $\boldsymbol{\Phi}(t)$ 外,还包含未知有色噪声项 $\boldsymbol{w}(t-i)$ 和不可测白噪声项 $\boldsymbol{v}(t-i)$,而这些未知变量可以通过输入输

出数据估算得到.

$\boldsymbol{\Phi}(t)$ 中的未知项 $\boldsymbol{w}(t-i)$ 和 $\boldsymbol{v}(t-i)$ 分别用其估计值 $\hat{\boldsymbol{w}}(t-i)$ 和 $\hat{\boldsymbol{v}}(t-i)$ 代替,代替后的信息矩阵记作 $\boldsymbol{\Psi}(t)$, 在辨识算法中使用 $\boldsymbol{\Psi}(t)$ 代替未知的 $\boldsymbol{\Phi}(t)$,能够获得估计伪线性回归辨识模型(204)参数向量 $\boldsymbol{\vartheta}$ 的多元广义增广随机梯度算法(Multivariate Generalized Extended Stochastic Gradient algorithm, M-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Psi}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t-1)],$$
$$\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{205}$$
$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{206}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{\boldsymbol{w}}(t-1),-\hat{\boldsymbol{w}}(t-2),\cdots,-\hat{\boldsymbol{w}}(t-n_c),$$
$$\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{207}$$
$$\hat{\boldsymbol{w}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{\boldsymbol{w}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{208}$$
$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{209}$$
$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),$$
$$\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{210}$$

5.2 多元多新息广义增广随机梯度辨识算法

借助于多新息辨识理论,从多元广义增广随机梯度辨识算法(205)—(210)可以得到新息长度为 p 的多元多新息广义增广随机梯度算法(Multivariate Multi-Innovation Generalized Extended Stochastic Gradient algorithm, M-MI-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{211}$$
$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{212}$$
$$r(t)=r(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\quad r(0)=1, \tag{213}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{\boldsymbol{w}}(t-1),-\hat{\boldsymbol{w}}(t-2),\cdots,-\hat{\boldsymbol{w}}(t-n_c),$$
$$\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{214}$$
$$\hat{\boldsymbol{w}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{\boldsymbol{w}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{215}$$
$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{216}$$
$$\boldsymbol{Y}(p,t)=[\boldsymbol{y}^{\mathrm{T}}(t),\boldsymbol{y}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}}, \tag{217}$$
$$\boldsymbol{\Gamma}(p,t)=[\boldsymbol{\Psi}^{\mathrm{T}}(t),\boldsymbol{\Psi}^{\mathrm{T}}(t-1),\cdots,\boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)], \tag{218}$$
$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),$$
$$\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{219}$$

当新息长度 $p=1$ 时,多元多新息广义增广随机梯度算法退化为多元广义增广随机梯度算法.

6 耦合广义增广随机梯度算法与耦合多新息广义增广随机梯度算法

考虑多元伪线性自回归滑动平均系统(203)对应的多元伪线性辨识模型(204),重写如下:

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\vartheta}+\boldsymbol{v}(t), \tag{220}$$

其中各变量的定义同上.

6.1 耦合子系统广义增广随机梯度辨识算法

令 $\boldsymbol{\psi}_i^{\mathrm{T}}(t)\in\mathbf{R}^{1\times n_0}$ 是 $\boldsymbol{\Psi}(t)$ 的第 i 行,即

$$\boldsymbol{\Psi}(t):=\begin{bmatrix}\boldsymbol{\psi}_1^{\mathrm{T}}(t)\\ \boldsymbol{\psi}_2^{\mathrm{T}}(t)\\ \vdots\\ \boldsymbol{\psi}_m^{\mathrm{T}}(t)\end{bmatrix}\in\mathbf{R}^{m\times n_0}.$$

利用耦合辨识概念,可以得到估计参数向量 $\boldsymbol{\vartheta}$ 的耦合子系统广义增广随机梯度算法(Coupled Subsystem Generalized Extended Stochastic Gradient algorithm, C-S-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],$$
$$\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{221}$$
$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2,\quad r_i(0)=1, \tag{222}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{\boldsymbol{w}}(t-1),-\hat{\boldsymbol{w}}(t-2),\cdots,-\hat{\boldsymbol{w}}(t-n_c),$$
$$\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{223}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{224}$$
$$\hat{\boldsymbol{w}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t),\quad \hat{\boldsymbol{w}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{225}$$
$$\hat{\boldsymbol{v}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t),\quad \hat{\boldsymbol{v}}(j)=\mathbf{1}_m/p_0,\quad j\leqslant 0, \tag{226}$$
$$\hat{\boldsymbol{\vartheta}}(t)=\frac{\hat{\boldsymbol{\vartheta}}_1(t)+\hat{\boldsymbol{\vartheta}}_2(t)+\cdots+\hat{\boldsymbol{\vartheta}}_m(t)}{m}, \tag{227}$$
$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),$$
$$\hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{228}$$

6.2 耦合广义增广随机梯度辨识算法

利用耦合辨识概念和多新息辨识理论,可以得到耦合广义增广随机梯度算法(Coupled Generalized Extended Stochastic Gradient algorithm, C-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}_1(t)=\hat{\boldsymbol{\vartheta}}_m(t-1)+\frac{\boldsymbol{\psi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\psi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_m(t-1)],$$
$$\hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0, \tag{229}$$
$$r_1(t)=r_m(t-1)+\|\boldsymbol{\psi}_1(t)\|^2,\quad r_m(0)=1, \tag{230}$$
$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_{i-1}(t)], \tag{231}$$
$$r_i(t)=r_{i-1}(t)+\|\boldsymbol{\psi}_i(t)\|^2,\quad i=2,3,\cdots,m, \tag{232}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{\boldsymbol{w}}(t-1),-\hat{\boldsymbol{w}}(t-2),\cdots,-\hat{\boldsymbol{w}}(t-n_c),$$
$$\hat{\boldsymbol{v}}(t-1),\hat{\boldsymbol{v}}(t-2),\cdots,\hat{\boldsymbol{v}}(t-n_d)], \tag{233}$$
$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{234}$$

$$\hat{w}(t)= \boldsymbol{y}(t) -\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t), \quad \hat{w}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{235}$$

$$\hat{v}(t)= \boldsymbol{y}(t) -\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t), \quad \hat{v}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{236}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t), \tag{237}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t), \hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{238}$$

注意到耦合广义增广随机梯度辨识算法(229)—(238)中,参数估计 $\hat{\boldsymbol{\vartheta}}_i(t)$ 和 $r_i(t)$ 都是耦合的,去掉 $r_i(t)$ 间的耦合,便得到一个简单的耦合广义增广随机梯度算法:

$$\hat{\boldsymbol{\vartheta}}_1(t)= \hat{\boldsymbol{\vartheta}}_m(t-1)+\frac{\boldsymbol{\psi}_1(t)}{r_1(t)}[y_1(t)-\boldsymbol{\psi}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_m(t-1)], \quad \hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0, \tag{239}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\psi}_1(t)\|^2, \quad r_1(0)=1, \tag{240}$$

$$\hat{\boldsymbol{\vartheta}}_i(t)= \hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\psi}_i(t)}{r_i(t)}[y_i(t)-\boldsymbol{\psi}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}_{i-1}(t)], \tag{241}$$

$$r_i(t)= r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2, \quad r_i(0)= 1, \quad i= 2,3,\cdots,m, \tag{242}$$

$$\boldsymbol{\Psi}(t)= [\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c), \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)], \tag{243}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{244}$$

$$\hat{w}(t)= \boldsymbol{y}(t) -\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t), \quad \hat{w}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{245}$$

$$\hat{v}(t)= \boldsymbol{y}(t) -\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t), \quad \hat{v}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{246}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t), \tag{247}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t), \hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{248}$$

6.3 耦合子系统多新息广义增广随机梯度辨识算法

借助于多新息辨识理论,从耦合子系统广义增广随机梯度算法(221)—(228),可以得到耦合子系统多新息广义增广随机梯度算法(Coupled Subsystem Multi-Innovation Generalized Extended Stochastic Gradient algorithm,C-S-MI-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t), \quad \hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0, \tag{249}$$

$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{250}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2, \quad r_i(0)=1, \tag{251}$$

$$\boldsymbol{\Gamma}_i(p,t)= [\boldsymbol{\psi}_i(t),\boldsymbol{\psi}_i(t-1),\cdots,\boldsymbol{\psi}_i(t-p+1)], \tag{252}$$

$$\boldsymbol{Y}_i(p,t)= [y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{253}$$

$$\boldsymbol{\Psi}(t)= [\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c), \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)], \tag{254}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{255}$$

$$\hat{w}(t)= \boldsymbol{y}(t) -\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t), \quad \hat{w}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{256}$$

$$\hat{v}(t)= \boldsymbol{y}(t) -\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t), \quad \hat{v}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{257}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\frac{\hat{\boldsymbol{\vartheta}}_1(t)+\hat{\boldsymbol{\vartheta}}_2(t)+\cdots+\hat{\boldsymbol{\vartheta}}_m(t)}{m}, \tag{258}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t), \hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{259}$$

6.4 耦合多新息广义增广随机梯度辨识算法

借助于多新息辨识理论,从耦合广义增广随机梯度辨识算法(239)—(248),可以得到耦合多新息广义增广随机梯度算法(Coupled Multi-Innovation Generalized Extended Stochastic Gradient algorithm,C-MI-GESG 算法):

$$\hat{\boldsymbol{\vartheta}}_1(t)=\hat{\boldsymbol{\vartheta}}_m(t-1)+\frac{\boldsymbol{\Gamma}_1(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t), \quad \hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0, \tag{260}$$

$$\boldsymbol{E}_1(p,t)=\boldsymbol{Y}_1(p,t)-\boldsymbol{\Gamma}_1^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}_m(t-1), \tag{261}$$

$$r_1(t)= r_1(t-1)+\|\boldsymbol{\psi}_1(t)\|^2, \quad r_1(0)= 1, \tag{262}$$

$$\boldsymbol{\Gamma}_1(p,t)= [\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_1(t-1),\cdots,\boldsymbol{\psi}_1(t-p+1)], \tag{263}$$

$$\boldsymbol{Y}_1(p,t)= [y_1(t),y_1(t-1),\cdots,y_1(t-p+1)]^{\mathrm{T}}, \tag{264}$$

$$\hat{\boldsymbol{\vartheta}}_i(t)=\hat{\boldsymbol{\vartheta}}_{i-1}(t)+\frac{\boldsymbol{\Gamma}_i(p,t)}{r_i(t)}\boldsymbol{E}_i(p,t), \tag{265}$$

$$\boldsymbol{E}_i(p,t)=\boldsymbol{Y}_i(p,t)-\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}_{i-1}(t), \tag{266}$$

$$r_i(t)=r_i(t-1)+\|\boldsymbol{\psi}_i(t)\|^2, \quad r_i(0)=1, \quad i=2,3,\cdots,m, \tag{267}$$

$$\boldsymbol{\Gamma}_i(p,t)= [\boldsymbol{\psi}_i(t),\boldsymbol{\psi}_i(t-1),\cdots,\boldsymbol{\psi}_i(t-p+1)], \tag{268}$$

$$\boldsymbol{Y}_i(p,t)= [y_i(t),y_i(t-1),\cdots,y_i(t-p+1)]^{\mathrm{T}}, \tag{269}$$

$$\boldsymbol{\Psi}(t)= [\boldsymbol{\Phi}_{\mathrm{s}}(t),-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n_c), \hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n_d)], \tag{270}$$

$$\boldsymbol{\Psi}(t)=[\boldsymbol{\psi}_1(t),\boldsymbol{\psi}_2(t),\cdots,\boldsymbol{\psi}_m(t)]^{\mathrm{T}}, \tag{271}$$

$$\hat{w}(t)= \boldsymbol{y}(t) -\boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t), \quad \hat{w}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{272}$$

$$\hat{v}(t)= \boldsymbol{y}(t) -\boldsymbol{\Psi}(t)\hat{\boldsymbol{\vartheta}}(t), \quad \hat{v}(j)=\mathbf{1}_m/p_0, \quad j\leqslant 0, \tag{273}$$

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t), \tag{274}$$

$$\hat{\boldsymbol{\vartheta}}(t)=[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t),\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t), \hat{d}_1(t),\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{275}$$

系统的参数估计定义为第 i 个子系统的参数估计向量:$\hat{\boldsymbol{\vartheta}}(t):=\hat{\boldsymbol{\vartheta}}_m(t)$.

C-MI-GESG 辨识算法(260)—(275)计算参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}_m(t)$ 的步骤如下:

① 赋初值:令 $t=1$,给定新息长度 p,置 $\hat{\boldsymbol{\vartheta}}_m(0)=\mathbf{1}_{n_0}/p_0$,$r_i(0)=1(i=1,2,\cdots,m)$,$\hat{w}(j)=\mathbf{1}_m/p_0$,$\hat{v}(j)=$

$\mathbf{1}_m/p_0(j \leqslant 0)$,$p_0 = 10^6$.

② 收集观测数据 $\boldsymbol{y}(t)$ 和 $\boldsymbol{\Phi}_s(t)$,根据式(270)构造 $\boldsymbol{\Psi}(t)$,根据式(271)从 $\boldsymbol{\Psi}(t)$ 中提取信息向量 $\boldsymbol{\psi}_i(t)$,$i = 1,2,\cdots,m$.

③ 用式(263)构造堆积信息矩阵 $\boldsymbol{\Gamma}_1(p,t)$,用式(264)构造堆积信息向量 $\boldsymbol{Y}_1(p,t)$.

④ 用式(262)计算 $r_1(t)$,用式(261)计算新息向量 $\boldsymbol{E}_1(p,t)$,用式(260)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}_1(t)$.

⑤ 依次当 $i = 2,3,\cdots,m$ 时,用式(268)构造堆积信息矩阵 $\boldsymbol{\Gamma}_i(p,t)$,用式(269)构造堆积信息向量 $\boldsymbol{Y}_i(p,t)$,用式(267)计算 $r_i(t)$,用式(266)计算新息向量 $\boldsymbol{E}_i(p,t)$,用式(265)刷新参数估计向量 $\hat{\boldsymbol{\vartheta}}_i(t)$.

⑥ 根据式(274)获得参数估计向量 $\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}_m(t)$,根据式(275)从参数估计向量 $\hat{\boldsymbol{\vartheta}}(t)$ 中读出 $\hat{\boldsymbol{\theta}}(t)$.

⑦ 分别用式(272)和(273)计算 $\hat{\boldsymbol{w}}(t)$ 和 $\hat{\boldsymbol{v}}(t)$.t 增1,转到第②步.

7 结语

本文将耦合辨识概念与多新息辨识理论相结合,研究了多元线性回归系统、多元伪线性滑动平均系统、多元伪线性自回归滑动平均系统的多元随机梯度类辨识算法、多元多新息随机梯度类辨识算法、耦合子系统随机梯度类辨识算法、耦合随机梯度类辨识算法、耦合子系统多新息随机梯度类辨识算法、耦合多新息随机梯度类辨识算法等.

这些研究针对各子系统全参数耦合的多元系统,提出的方法都是全耦合随机梯度类辨识方法、全耦合多新息随机梯度类辨识方法等,这些研究思想可推广到各子系统部分参数耦合的多元系统,研究和提出部分耦合随机梯度类辨识方法、部分耦合多新息随机梯度类辨识方法等,研究和提出部分耦合最小二乘类辨识方法、部分耦合多新息最小二乘类辨识方法等,进一步可研究这些辨识方法的计算效率及收敛性能.

参考文献
References

[1] 丁锋.系统辨识新论[M].北京:科学出版社,2013
DING Feng. System identification: New theory and methods[M].Beijing:Science Press,2013
[2] 丁锋.系统辨识:辨识方法性能分析[M].北京:科学出版社,2014
DING Feng. System identification: Performance analysis for identification methods [M]. Beijing: Science Press,2014
[3] 丁锋.系统辨识(1):辨识导引[J].南京信息工程大学学报:自然科学版,2011,3(1):1-22
DING Feng. System identification. Part A: Introduction to the identification[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2011,3(1):1-22
[4] 丁锋.系统辨识(2):系统描述的基本模型[J].南京信息工程大学学报:自然科学版,2011,3(2):97-117
DING Feng.System identification.Part B:Basic models for system description[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2011,3(2):97-117
[5] 丁锋.系统辨识(3):辨识精度与辨识基本问题[J].南京信息工程大学学报:自然科学版,2011,3(3):193-226
DING Feng. System identification. Part C: Identification accuracy and basic problems[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2011,3(3):193-226
[6] 丁锋.系统辨识(4):辅助模型辨识思想与方法[J].南京信息工程大学学报:自然科学版,2011,3(4):289-318
DING Feng.System identification.Part D:Auxiliary model identification idea and methods[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2011,3(4):289-318
[7] 丁锋.系统辨识(5):迭代搜索原理与辨识方法[J].南京信息工程大学学报:自然科学版,2011,3(6):481-510
DING Feng.System identification.Part E:Iterative search principle and identification methods [J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2011, 3 (6): 481-510
[8] 丁锋.系统辨识(6):多新息辨识理论与方法[J].南京信息工程大学学报:自然科学版,2012,4(1):1-28
DING Feng.System identification.Part F:Multi-innovation identification theory and methods[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2012,4(1):1-28
[9] 丁锋.系统辨识(7):递阶辨识原理与方法[J].南京信息工程大学学报:自然科学版,2012,4(2):97-124
DING Feng. System identification. Part G: Hierarchical identification principle and methods [J]. Journal of Nanjing University of Information Science and Technology:Natural Science Edition,2012,4(2):97-124
[10] 丁锋.系统辨识(8):耦合辨识概念与方法[J].南京信息工程大学学报:自然科学版,2012,4(3):193-212
DING Feng.System identification.Part H:Coupling identification concept and methods[J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2012,4(3):193-212
[11] 丁锋.辨识方法的计算效率(1):递推算法[J].南京信息工程大学学报:自然科学版,2012,4(4):289-300
DING Feng.Computational efficiency of the identification methods.Part A:Recursive algorithms[J].Journal of Nan-

jing University of Information Science and Technology: Natural Science Edition,2012,4(4):289-300

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

[13] 丁锋.辨识方法的计算效率(3):信息向量耦合算法[J].南京信息工程大学学报:自然科学版,2012,4(6):481-495
DING Feng.Computational efficiency of the identification methods. Part C: Coupled information vector algorithms [J].Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2012, 4 (6): 481-495.

[14] Ding F, Liu G, Liu X P.Partially coupled stochastic gradient identification methods for non-uniformly sampled systems[J]. IEEE Transactions on Automatic Control, 2010,55(8):1976-1981

[15] Ding F.Coupled least squares identification for multivariable systems[J].IET Control Theory and Applications, 2013,7(1):68-79.

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

[17] Ding F, Chen T.Hierarchical gradient-based identification of multivariable discrete-time systems[J]. Automatica, 2005,41(2):315-325

[18] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems[J].IEEE Transactions on Automatic Control,2005,50(3):397-402

[19] Ding F, Chen T.Hierarchical identification of lifted state-space models for general dual-rate systems[J]. IEEE Transactions on Circuits and Systems.I:Regular Papers, 2005,52(6):1179-1187

[20] 丁锋,萧德云.多变量系统状态空间模型的递阶辨识[J].控制与决策,2005,20(8):848-853+859
DING Feng, XIAO Deyun. Hierarchical identification of state space models for multivariable systems[J].Control and Decision,2005,20(8):848-853+859

[21] 丁锋,杨家本.大系统的递阶辨识[J].自动化学报,1999,25(5):647-654
DING Feng, YANG Jiaben. Hierarchical identification of large scale systems[J].Acta Automatica Sinica,1999,25(5):647-654

[22] 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

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

[24] 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

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

[26] 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

[27] 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

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

[29] 丁锋,丁韬,杨家本,等.时变参数遗忘梯度估计算法的收敛性[J].自动化学报,2002,28(6):962-968
DING Feng, DING Tao, YANG Jiaben, et al.Convergence of forgetting gradient estimation algorithm for time-varying parameters[J]. Acta Automatica Sinica, 2002, 28 (6): 962-968

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

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

[32] 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

[33] 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

编者按:本文作者丁锋教授从 2011 年第 1 期到 2012 年第 6 期已在本刊发表了关于系统辨识方面的连载论文 11 篇,其中前 8 篇由本刊结集编印成《系统辨识论文连载文集》,以此蓝本为基础,丁锋教授进行补充,便形成了《系统辨识新论》一书,已由科学出版社出版[1].从 2014 年第 1 期开始,丁锋教授又将陆续连载系统辨识论文,内容涉及辨识新思想、新理论、新原理、新概念与辨识新方法.敬请期待!

Coupled multi-innovation stochastic gradient type identification methods for multivariate systems

DING Feng[1,2,3] WANG Feifei[1]
1 School of Internet of Things Engineering,Jiangnan University,Wuxi 214122
2 Control Science and Engineering Research Center,Jiangnan University,Wuxi 214122
3 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),Jiangnan University,Wuxi 214122

Abstract For multivariate linear regression systems,using the coupling identification concept and the multi-innovation identification theory,this paper discusses a multivariate stochastic gradient algorithm,a multivariate multi-innovation stochastic gradient algorithm,and an interval-varying multivariate multi-innovation gradient algorithm,decomposes a multivariate system into several subsystems,and presents a coupled subsystem stochastic gradient algorithm,a coupled stochastic gradient algorithm,a coupled subsystems multi-innovation stochastic gradient algorithm and a coupled multi-innovation stochastic gradient algorithm.These methods are extended to multivariate pseudo-linear moving average systems and multivariate pseudo-linear autoregressive moving average systems.Finally,this paper gives the steps and diagrams for computing the parameter estimates using several typical coupled stochastic gradient algorithms and coupled multi-innovation stochastic gradient algorithms.

Key words parameter estimation;recursive identification;gradient search;least squares;auxiliary model identification idea;multi-innovation identification theory;hierarchical identification principle;coupling identification concept;multivariate system