

耦合多新息随机梯度辨识方法性能分析

鹿振宇, 黄攀峰

(西北工业大学 a. 航天学院, b. 航天飞行动力学技术重点实验室, 西安 710072)

摘要: 针对一类耦合参数多变量系统, 提出一种耦合多新息随机梯度方法. 通过该方法进行参数辨识并对该方法进行性能分析. 该方法的基本思路在于利用历史新息中包含的信息, 将耦合随机梯度算法中的新息项扩展为多新息向量, 从而提升耦合随机梯度算法中单个子系统的辨识效果. 仿真结果表明, 通过增加新息长度可以提升辨识结果的收敛速度和精度.

关键词: 参数估计; 多变量系统; 耦合随机梯度; 多新息向量; 收敛性分析

中图分类号: TP273

文献标志码: A

Performance analysis of coupled multi-innovation stochastic gradient identification method

LU Zhen-yu, HUANG Pan-feng

(a. School of Astronautics, b. National Key Laboratory of Aerospace Flight Dynamics, Northwestern Polytechnical University, Xi'an 710072, China. Correspondent: HUANG Pan-feng, E-mail: pfhuang@nwpu.edu.cn)

Abstract: For a kind of coupled parameters multivariable system, the coupled multi-innovation stochastic gradient (C-MISG) identification method is proposed to estimate the parameters, and its performance analysis is made. The basic idea of this method is utilizing the historical information to extend the scalar innovation item to an innovation vector to enhance the identified effect of each subsystem. Simulation results show that increasing the innovation length can enhance the convergence rate and accuracy of the identified results.

Keywords: parameter estimation; multivariable system; coupled stochastic gradient; multi-innovation vector; convergence analysis

0 引言

近年来, 针对多变量系统的建模与辨识的研究受到了越来越多学者的关注, 相应的新方法也层出不穷, 例如随机梯度 (SG) 法^[1-2]、递阶辨识法^[3-4]、多新息辨识^[5-7]、辅助模型辨识^[8-9]、递推最小二乘 (RLS) 辨识^[10-11] 和子空间辨识法^[12]等. SG 算法和 RLS 算法是两种典型的多变量系统辨识算法, 相比而言, SG 算法的计算复杂度较低, 而 RLS 算法的收敛速度比 SG 算法快^[3].

多新息随机梯度算法 (MISG) 是一种将多新息辨识与 SG 算法相结合而得到的辨识方法, 该算法的收敛速度比 SG 算法快, 计算复杂度低于 RLS 算法. 但是, MISG 算法只针对单系统辨识, 并不能解决由多个子系统组成的耦合参数多变量系统的参数辨识问题.

针对可分解为耦合参数子系统的多变量系统的参数辨识问题, 丁锋等提出了部分耦合随机 (P-CSG) 梯度算法^[2]和耦合最小二乘 (CLS) 算法^[11]. P-CSG 算法旨在解决耦合多变量系统非均匀采样的问题, 而 CLS 算法则将多变量系统分解为若干个存在耦合参数的子系统, 并对子系统参数依次辨识, 从而避免 RLS 算法中矩阵的求逆运算. 通过证明可知, CLS 算法与 RLS 算法的辨识结果相同.

比较 CSG 与 MICG 算法, 不难发现两种算法的相似点: 利用当前最优的辨识结果来替代其他子系统或本系统的历史最优辨识结果. CSG 算法如前文所述, 是用当前最优结果替代相邻子系统的前一时刻的最优辨识结果; 而 MISG 算法是将当前最优的辨识结果代入历史信息向量, 并替代历史辨识结果, 从

收稿日期: 2014-04-14; 修回日期: 2014-10-11.

基金项目: 国家自然科学基金项目(11272256, 61005062).

作者简介: 鹿振宇(1989-), 男, 博士生, 从事空间遥操作、空间机器人参数辨识的研究; 黄攀峰(1974-), 男, 教授, 博士生导师, 从事空间机器人、空间遥操作等研究.

而构成新息向量. 但是, 这两种替代效果并不重复, 进而可以将两种算法结合得到耦合多新息随机梯度 (C-MISG) 算法.

首先, 本文对耦合参数多变量系统进行建模; 然后, 介绍 C-MISG 算法, 并对算法性能进行分析; 最后, 通过仿真算例验证 C-MISG 算法的有效性和可行性, 并对新息长度对收敛效果的影响进行分析.

1 问题描述

考虑如下多变量线性系统:

$$\mathbf{Y}(t) = \Phi^T(t)\theta + \mathbf{v}(t). \quad (1)$$

其中: $\mathbf{Y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]^T \in R^m$ 为输出向量; $\Phi^T(t) = [\varphi_1^T(t), \varphi_2^T(t), \dots, \varphi_m^T(t)] \in R^{m \times n}$ 为信息向量矩阵, $\varphi_i^T(t) \in R^n$ ($i = 1, 2, \dots, m$) 表示子系统 i 中包含输入数据和输出数据的信息向量; $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_m(t)] \in R^m$ 为系统噪声向量; $\theta \in R^n$ 为待辨识参数. 根据输入输出变量个数, 系统 (1) 可分解为 m 个具有相同待辨识耦合参数的子系统, 即

$$\begin{cases} y_1(t) = \varphi_1^T(t)\theta + v_1(t), \\ y_2(t) = \varphi_2^T(t)\theta + v_2(t), \\ \vdots \\ y_m(t) = \varphi_m^T(t)\theta + v_m(t). \end{cases} \quad (2)$$

虽然式 (2) 中的每个子系统的参数都可以单独辨识, 但可能都不适合作为整个系统的辨识结果, 所以辨识的结果应当是各子系统的协调结果, 即系统 (1) 整体的辨识参数.

2 C-MISG 算法

定义多新息误差向量

$$\mathbf{E}_i(p, t) = [e_i(t), e_i(t-1), \dots, e_i(t-p+1)]^T \in R^p.$$

其中: p 为新息的长度; $e_i(t-k)$ ($k = 1, 2, \dots, p-1$) 的定义为

$$e_i(t-k) = y_i(t-k) - \varphi_i^T(t-k)\hat{\theta}_{i-1}(t), \quad (3)$$

其中 $\hat{\theta}_{i-1}(t)$ 为子系统 i .

定义多新息向量的信息矩阵 $\Gamma_i(p, t)$ 和输出向量 $\mathbf{Y}_i(p, t)$ 分别为

$$\begin{cases} \Gamma_i(p, t) = [\varphi_i(t), \dots, \varphi_i(t-p+1)] \in R^{n \times p}, \\ \mathbf{Y}_i(p, t) = [y_i(t), \dots, y_i(t-p+1)]^T \in R^p. \end{cases}$$

由式 (3) 可知, 多新息向量可以表示为

$$\mathbf{E}_i(p, t) = \begin{cases} \mathbf{Y}_1(p, t) - \Gamma_1^T(p, t)\hat{\theta}_m(t-1), & i = 1; \\ \mathbf{Y}_i(p, t) - \Gamma_i^T(p, t)\hat{\theta}_{i-1}(t). \end{cases} \quad (4)$$

由文献 [13] 直接给出系统 (2) 中各子系统中向量的耦合多新息随机梯度 (C-MISG) 辨识算法的步骤,

即有

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + \frac{\Gamma_i(p, t)}{R_i(t)} \mathbf{E}_i(p, t), \quad i = 2, 3, \dots, m; \quad (5)$$

$$\hat{\theta}_1(t) = \hat{\theta}_m(t-1) + \frac{\Gamma_1(p, t)}{R_1(t)} \mathbf{E}_1(p, t); \quad (6)$$

$$R_i(t) = R_i(t-1) + \|\Gamma_i(t)\|^2, \quad r_i(0) = I. \quad (7)$$

3 C-MISG 算法性能分析

为了分析 C-MISG 算法的收敛性能, 首先引入如下两条假设和两条引理.

假设 1 各子系统的信息向量 $\Phi_i(t)$ 均为持续激励, 且存在常数 $0 < \alpha \leq \beta < \infty$ 和正整数 $N \geq n$, 使得如下持续激励条件成立:

$$\alpha I \leq \frac{1}{N} \sum_{i=0}^{N-1} \varphi_i(k+i)\varphi_i(k+i)^T \leq \beta I.$$

假设 2 各子系统的噪声均为零均值白噪声, 即满足

$$\begin{aligned} E(v_i(t)) &= 0; \\ E\{v_i(t_1)v_i(t_2)\} &= \begin{cases} 0, & t_1 \neq t_2; \\ \sigma^2, & t_1 = t_2. \end{cases} \end{aligned}$$

引理 1^[14] 若 $\{x(k)\}$ 、 $\{a_k\}$ 和 $\{b_k\}$ 为非负实序列, 且满足 $x(k+1) \leq (1-a_k)x(k) + b_k$, $k \geq 0$, 其中 $a_k \in [0, 1)$ 且 $x(0) < \infty$, 则有 $\lim_{t \rightarrow \infty} x(t) \leq \lim_{t \rightarrow \infty} (b_k/a_k)$.

引理 2 根据假设 1, 式 (7) 中的 $R_i(t)$ ($i = 1, 2, \dots, m$) 满足 $nm\alpha(t-N) + 1 \leq R_i(t) \leq nm\beta(t+N-1) + 1$.

证明 由假设 1 不难得到

$$mnN\alpha \leq \sum_{j=1}^m \sum_{i=1}^{N-1} \|\varphi_j(t+i)\|^2 \leq mnN\beta.$$

令 $[a]$ 表示不超过 a 的最大整数, 则有

$$R_m(t) = R_m(t-1) + \sum_{j=1}^m \|\varphi_j(t)\|^2 =$$

$$\sum_{j=1}^m \sum_{i=1}^T \|\varphi_j(i)\|^2 + R(0) \leq$$

$$\sum_{j=1}^m \sum_{k=0}^{\lfloor (t-1)/N \rfloor} \sum_{i=1}^N \|\varphi_j(kN+i)\|^2 + R(0) \leq$$

$$nm\beta(t+N-1) + 1,$$

$$R_1(t) > R_m(t-1) = R_m(t-2) + \sum_{j=1}^m \|\varphi_j(t)\|^2 =$$

$$\sum_{j=1}^m \sum_{i=1}^{t-1} \|\varphi_j(i)\|^2 + R(0) \geq$$

$$\sum_{j=1}^m \sum_{k=0}^{\lfloor (t-1)/N \rfloor - 1} \sum_{i=1}^N \|\varphi_j(kN+i)\|^2 + R(0) \geq$$

$$nm\alpha(t-N) + 1.$$

由 $R_i(t)$ 定义可知 $R_1(t) \leq R_i(t) \leq R_m(t)$, 故存在 $nm\alpha(t - N) + 1 \leq R_i(t) \leq nm\beta(t + N - 1) + 1$.

详细证明过程可参见文献 [15] 中的引理 2 以及文献 [16]. \square

定理 1 对于系统 (1) 和多新息耦合随机梯度算法 (5)~(7), 当满足假设条件时, 系统的参数辨识结果满足 $\lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_i(t)\|^2\} = 0, i = 1, 2, \dots, m$.

证明 定义 $\tilde{\theta}_i(t) = \hat{\theta}_i(t) - \theta$, 则式 (6) 两边分别减去 θ , 可得

$$\begin{aligned} \tilde{\theta}_m(t) = & \tilde{\theta}_{m-1}(t) + \frac{\Gamma_m(p, t)}{R_m(t)}(-\Gamma_m^T(p, t)\tilde{\theta}_{m-1}(t) + v_m(p, t)) = \\ & \left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right)\tilde{\theta}_{m-1}(t) + \frac{\Gamma_m(p, t)v_m(p, t)}{R_m(t)}, \end{aligned}$$

两边取范数, 可得

$$\begin{aligned} \|\tilde{\theta}_m(t)\|^2 \leq & \left\| \left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right)\tilde{\theta}_{m-1}(t) \right\|^2 + \\ & 2\tilde{\theta}_{m-1}^T(t) \left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right) + \\ & \left\| \frac{\Gamma_m(p, t)v_m(p, t)}{R_m(t)} \right\|^2 \leq \\ & \lambda_{\max} \left[\left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right) \right] \|\tilde{\theta}_{m-1}(t)\|^2 + \\ & 2\tilde{\theta}_{m-1}^T(t) \left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right) + \\ & \frac{\|\Gamma_m(p, t)v_m(p, t)\|^2}{R_m(t)}, \end{aligned} \tag{8}$$

取 $p = N$, 由引理 2 和假设 2 可得

$$\begin{aligned} I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)} & \leq \left(1 - \frac{N\alpha}{nm\beta(t + N - 1)}\right)I, \\ E\{\|\Gamma_m(p, t)v_m(p, t)\|^2\} & \leq \\ E\{\lambda_{\max}[\Gamma_m(p, t)\Gamma_m^T(p, t)]\|v_m(p, t)\|^2\} & \leq \\ p\beta E\{\|v_m(p, t)\|^2\} & = p^2\beta\sigma^2 = N^2\beta\sigma^2. \end{aligned}$$

对式 (8) 两边取期望, 可得

$$\begin{aligned} E\{\|\tilde{\theta}_m(t)\|^2\} \leq & \lambda_{\max} \left[\left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right) \right] \{E\|\tilde{\theta}_{m-1}(t)\|^2\} + \\ 2E\left\{\tilde{\theta}_{m-1}^T(t) \left(I - \frac{\Gamma_m(p, t)\Gamma_m^T(p, t)}{R_m(t)}\right) \times \right. & \\ \left. \frac{\Gamma_m(p, t)v_m(p, t)}{R_m(t)}\right\} + \frac{E\{\|\Gamma_m(p, t)v_m(p, t)\|^2\}}{R_m(t)} & \leq \\ \left(1 - \frac{N\alpha}{nm\beta(t + N - 1)}\right)E\{\|\tilde{\theta}_{m-1}(t)\|^2\} + & \\ \frac{N^2\beta\sigma^2}{[nm\alpha(t - N) + 1]^2}. & \end{aligned} \tag{9}$$

其中 $\lambda_{\max}[A]$ 为对称矩阵 A 的最大特征值, 令

$$f(\alpha, \beta) = 1 - \frac{N\alpha}{nm\beta(t + N - 1)},$$

$$g(\alpha, \beta) = \frac{N^2\beta\sigma^2}{[nm\alpha(t - N) + 1]^2},$$

则式 (9) 可化为

$$E\{\|\tilde{\theta}_m(t)\|^2\} \leq f(\alpha, \beta)E\{\|\tilde{\theta}_{m-1}(t)\|^2\} + g(\alpha, \beta).$$

不难推知

$$\begin{aligned} E\{\|\tilde{\theta}_m(t)\|^2\} \leq & f(\alpha, \beta)^m E\{\|\tilde{\theta}_m(t-1)\|^2\} + \sum_{i=0}^{m-1} f(\alpha, \beta)^i g(\alpha, \beta). \end{aligned}$$

由引理 1 可得

$$\begin{aligned} \lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_m(t)\|^2\} = & \lim_{t \rightarrow \infty} \frac{\sum_{i=0}^{m-1} f(\alpha, \beta)^i g(\alpha, \beta)}{1 - f(\alpha, \beta)^m} = \lim_{t \rightarrow \infty} \frac{g(\alpha, \beta)}{1 - f(\alpha, \beta)} = \\ & \lim_{t \rightarrow \infty} \frac{N^2\beta\sigma^2}{[nm\alpha(t - N) + 1]^2} \frac{nm\beta(t + N - 1) + 1}{N\alpha} = \\ & \lim_{t \rightarrow \infty} \frac{N\beta\sigma^2[n\beta(t + N - 1) + 1]}{\alpha m[n\alpha(t - N) + 1]^2} \sim \lim_{t \rightarrow \infty} \frac{N\beta\sigma^2}{nm\alpha^3} \times \frac{1}{t}. \end{aligned}$$

显然, $\frac{N\beta\sigma^2}{nm\alpha^3}$ 为有界变量, 且满足 $\lim_{t \rightarrow \infty} \frac{1}{t} = 0$, 可得 $\lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_m(t)\|^2\} = 0$, 即子系统参数 $\theta_m(t)$ 收敛.

根据参数 $\theta_m(t)$ 的收敛性分析其他子系统参数的收敛性能. 以 $\tilde{\theta}_1(t + 1)$ 为例, 存在如下关系:

$$\begin{aligned} \lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_1(t + 1)\|^2\} \leq & \lim_{t \rightarrow \infty} \left(1 - \frac{N\alpha}{nm\beta(t + N)}\right) E\{\|\tilde{\theta}_m(t)\|^2\} + \\ & \left[\frac{N}{t - N + 1}\right]^2 \frac{\beta\sigma^2}{nm\alpha^2}. \end{aligned}$$

显然, 当 $t \rightarrow \infty$ 时, 易得 $\lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_1(t + 1)\|^2\} = 0$.

以此类推, 可得 $\lim_{t \rightarrow \infty} E\{\|\tilde{\theta}_i(t)\|^2\} = 0, i = 1, 2, \dots, m$. \square

4 仿真算例

考虑一个可分解为 4 个子系统的多变量系统, 各子系统的表达式为

$$\begin{aligned} A(z)y_i(t) & = B(z)u_i(t) + v(t), \\ A(z) & = 1 + a_1z^{-1} + a_2z^{-2} = 1 + 1.65z^{-1} - 0.85z^{-2}, \\ B(z) & = 1 + b_1z^{-1} + b_2z^{-2} = 1 - 0.31z^{-1} + 0.53z^{-2}, \\ \theta & = [-1.65 \quad 0.85 \quad -0.31 \quad 0.53]. \end{aligned}$$

其中: 输入信号 $\{u(t)\}$ 为零均值单位方差且不相关激励信号, $\{v(t)\}$ 为白噪声信号且方差为 $\sigma^2 = 0.2^2$, 子系统的数目 $m = 4$, 辨识结果误差值 $\delta = \|\hat{\theta}(t) - \theta\|/\theta$. 分别用不同新息长度下的 C-MISG 算法对系统进行参数辨识, 辨识步数均为 2000, 最终辨识结果及其误差值分别如表 1 和图 1 所示.

表 1 C-MISG 算法

p	a_1	a_2	b_1	b_2	$\delta/\%$
1	-1.127	0.356	-0.306	0.333	37.1
2	-1.428	0.643	-0.311	0.450	15.3
3	-1.531	0.740	-0.314	0.490	8.4
4	-1.579	0.786	-0.315	0.510	5.5
8	-1.586	0.792	-0.316	0.512	4.5
真值	-1.65	0.85	-0.31	0.53	

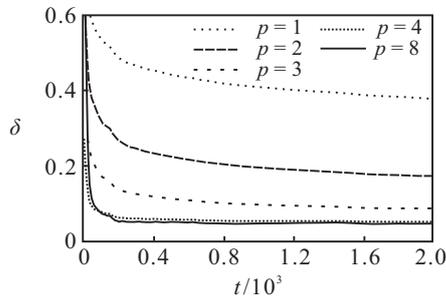


图 1 C-MISG 算法辨识

从以上图表可看出,不同新息长度下的收敛速度不同,当 $p=1$ 时,C-MISG 算法退化为 CSG 算法,此时收敛速度较慢,收敛误差较大.但随着 p 值增加,C-MISG 算法的收敛速度加快,收敛误差也降低了很多,当 $p>4$ 时,通过增加新息长度已不能对 C-MISG 算法的收敛速度和收敛质量产生较为明显的效果.

5 结 论

本文提出了 C-MISG 算法,并对 C-MISG 算法的性能进行了分析.相比于 CSG 算法,C-MISG 算法的收敛速度和收敛效果有较大的提高.该方法可以推广至包含有色噪声干扰的耦合参数多变量系统的参数辨识等问题.

参考文献(References)

- [1] Ding F, Shi Y, Chen T. Gradient-based identification methods for Hammerstein nonlinear ARMAX models[J]. *Nonlinear Dynamics*, 2006, 45(1/2): 31-43.
- [2] Ding F, Liu G, Liu X P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems[J]. *IEEE Trans on Automatic Control*, 2010, 55(8): 1976-1981.
- [3] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems[J]. *IEEE Trans on Automatic Control*, 2005, 50(3): 397-402.
- [4] Zhang Z, Ding F, Liu X. Hierarchical gradient based iterative parameter estimation algorithm for multivariable output error moving average systems[J]. *Computers and Mathematics with Applications*, 2011, 61(3): 672-682.
- [5] Ding F, Chen H, 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.

- [6] Liu Y, Xiao Y, Zhao X. 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.
- [7] Ding F. Several multi-innovation identification methods[J]. *Digital Signal Processing*, 2010, 20(4): 1027-1039.
- [8] Ding F, Liu P X, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises[J]. *Signal Processing*, 2009, 89(10): 1883-1890.
- [9] Liu Y, Xiao Y, Zhao X. 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.
- [10] Ding F, Chen T, Qiu L. Bias compensation based recursive least-squares identification algorithm for MISO systems[J]. *IEEE Trans on Circuits and Systems II: Express Briefs*, 2006, 53(5): 349-353.
- [11] Ding F. Coupled least squares identification for multi-variable systems[J]. *IET Control Theory and Applications*, 2013, 7(1): 68-79.
- [12] Oku H, Kimura H. Recursive 4SID algorithms using gradient type subspace tracking[J]. *Automatica*, 2002, 38(6): 1035-1043.
- [13] 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 F, Wang F F. Coupled multi-innovation stochastic gradient type identification methods for multivariate systems[J]. *J of Nanjing University of Information Science and Technology: Natural Science Edition*, 2014, 6(1): 1-16.)
- [14] 郭雷.时变随机系统——稳定性、估计与控制[M].长春:吉林科学技术出版社,1993: 14-16.
(Guo L. Time-varying stochastic systems—Stability, estimation and control[M]. Changchun: Jilin Science and Technology Press, 1993: 14-16.)
- [15] 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 J, Xie L, Ding F. Performance analysis of multi-innovation stochastic gradient identification for non-uniformly sampled systems[J]. *Control and Decision*, 2011, 26(9): 1338-1342.)
- [16] 于丽,丁锋,张佳波.多新息随机梯度辨识方法的收敛性研究[J].*科学技术与工程*,2007,7(21): 5474-5478.
(Yu L, Ding F, Zhang J B. Convergence of multi-innovation stochastic gradient identification methods[J]. *Science Technology and Engineering*, 2007, 7(21): 5474-5478.)