

多变量系统的耦合梯度辨识算法与性能分析

刘艳君, 丁 锋

(江南大学 a. 轻工过程先进控制教育部重点实验室, b. 物联网工程学院, 江苏 无锡 214122)

摘要: 针对多变量系统维数大、参数多、一般的辨识算法计算量大的问题, 基于耦合辨识概念, 推导多变量系统的耦合随机梯度算法, 利用鞅收敛定理分析算法的收敛性能. 算法的主要思想是将系统模型分解为多个单输出子系统, 在子系统的递推辨识过程中, 将每个子系统的参数估计值耦合起来. 所提出算法与最小二乘算法和耦合最小二乘算法相比, 具有较少的计算量, 收敛速度可以通过引入遗忘因子得到改善. 性能分析表明了所提出算法收敛, 仿真实例验证了算法的有效性.

关键词: 多变量系统; 参数估计; 随机梯度; 性能分析; 耦合辨识概念

中图分类号: TP273

文献标志码: A

Coupled stochastic gradient algorithm and performance analysis for multivariable systems

LIU Yan-jun, DING Feng

(a. Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, b. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China. Correspondent: LIU Yan-jun, Email: yanjunliu_1983@126.com)

Abstract: It is an issue that multivariable systems with high dimensions have many parameters, resulting in heavy computational costs in identification methods. Therefore, a coupled stochastic gradient algorithm is derived for multivariable systems based on the coupling identification concept. The identification model is decomposed into several single-output systems, and the parameter estimates are coupled during the subsystem identification by using the gradient search. The convergence properties are analyzed by using the martingale convergence theorem. Compared with the recursive least squares algorithm and the coupled least squares algorithm, the proposed algorithm has less computational load. The convergence rate can be improved by introducing a forgetting factor. Performance analysis verifies that the proposed algorithm converges. The simulation results show the effectiveness of the proposed algorithm.

Keywords: multivariable systems; parameter estimation; stochastic gradient; performance analysis; coupling identification concept

0 引 言

多变量系统的参数估计是辨识领域的一项重要内容, 20年前开始受到重视^[1-4]. 多变量系统常用的辨识方法为子系统辨识方法, 在这类子系统辨识方法中, 每个子系统的参数辨识是相互独立的, 是以每个子系统的损失函数各自达到最小为目的得到的, 并不是以系统的整体损失函数达到最小^[5]. 由于多变量系统的变量较多, 系统参数化后模型的维数较大, 导致一般的辨识方法计算量很大. 近年来, 人们开始致力于寻求计算量小的辨识方法, 如递阶辨识类方法^[6]. 这类辨

识方法根据辨识模型的特点将高维的辨识模型分解为两个或多个维数较少的子系统辨识模型进行辨识, 有效地减少计算量. 多变量系统的耦合辨识方法是另一种能有效减少辨识计算量的方法^[7], 其基本思想是, 将多变量系统辨识模型分解为子系统进行辨识, 并在子系统辨识中将每个子系统的辨识结果耦合起来, 使得每个子系统的辨识结果相互关联. 多变量系统的耦合最小二乘算法和整体辨识模型的最小二乘算法等价, 都是以最小化整体损失函数得到的, 但由于在辨识过程中避免了大矩阵的求逆运算, 有效提高了计算

收稿日期: 2015-05-06; **修回日期:** 2015-11-26.

基金项目: 国家自然科学基金项目(61304138, 61473136, 61203111); 江苏省自然科学基金项目(BK20130163).

作者简介: 刘艳君(1983—), 女, 讲师, 博士, 从事系统辨识、自适应控制等研究; 丁锋(1963—), 男, 教授, 博士生导师, 从事系统辨识、自适应控制等研究.

效率^[8-9]. 随机梯度算法的计算量远小于最小二乘法, 文献 [10] 提出了多变量系统 ARX-like 模型的半耦合随机梯度算法, 文献 [11] 探讨了多元系统耦合多新息随机梯度类辨识方法.

本文基于耦合辨识概念, 结合梯度搜索原理, 推导出多变量系统的耦合梯度辨识方法, 并将其计算量与现有的最小二乘算法和耦合最小二乘算法进行比较, 体现其计算量的优越性. 利用鞅收敛定理对算法的收敛性能进行分析, 并将该算法用于一非线性系统和一类多变量系统进行仿真分析, 结果验证了算法的有效性.

1 耦合随机梯度辨识算法

多变量线性系统可利用如下多元回归模型表示^[12-13]:

$$\mathbf{y}(t) = \Phi(t)\boldsymbol{\vartheta} + \mathbf{v}(t). \quad (1)$$

其中 $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]^T \in \mathbf{R}^m$ 为输出向量, $\Phi(t) \in \mathbf{R}^{m \times n}$ 为输入输出数据构成的回归信息矩阵, $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_m(t)]^T \in \mathbf{R}^m$ 为随机白噪声向量, $\boldsymbol{\vartheta} \in \mathbf{R}^n$ 为需要辨识的参数向量. 极小化总体损失函数 $J(\boldsymbol{\vartheta}) := \|\mathbf{y}(t) - \Phi(t)\boldsymbol{\vartheta}\|^2$ 不难得到估计参数向量 $\boldsymbol{\vartheta}$ 的随机梯度算法为

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}(t) &= \\ \hat{\boldsymbol{\vartheta}}(t-1) - \frac{\mu(t)}{2} \text{grad}[J(\hat{\boldsymbol{\vartheta}}(t-1))] &= \\ \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\Phi^T(t)}{r(t)} [\mathbf{y}(t) - \Phi(t)\hat{\boldsymbol{\vartheta}}(t-1)], & \quad (2) \\ r(t) &= r(t-1) + \|\Phi(t)\|^2, \\ r(0) &= 1. \end{aligned} \quad (3)$$

如果令 $\Phi_i(t) \in \mathbf{R}^{1 \times n_0}$ 为信息矩阵 $\Phi(t)$ 的第 i 行, 则模型 (1) 可以分解为 m 个子系统辨识模型, 即

$$y_i(t) = \Phi_i(t)\boldsymbol{\vartheta} + v_i(t), \quad i = 1, 2, \dots, m.$$

每个子系统的参数向量都是 $\boldsymbol{\vartheta}$, 且对于每个子系统都有如下随机梯度算法成立:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_i(t) &= \\ \hat{\boldsymbol{\vartheta}}_i(t-1) + \frac{\Phi_i^T(t)}{r_i(t)} [y_i(t) - \Phi_i(t)\hat{\boldsymbol{\vartheta}}_i(t-1)], & \quad (4) \\ r_i(t) &= r_i(t-1) + \|\Phi_i(t)\|^2, \\ r_i(0) &= 1. \end{aligned} \quad (5)$$

其中 $\hat{\boldsymbol{\vartheta}}_i(t)$ 为第 i 个子系统辨识得到的 $\boldsymbol{\vartheta}$ 的参数估计值. 由式 (4) 可见, $\hat{\boldsymbol{\vartheta}}_i(t)$ 只依赖于各个子系统的输出变量 $y_i(t)$ 和信息向量 $\Phi_i(t)$, 因此每个 $\hat{\boldsymbol{\vartheta}}_i(t)$ 是独立的. 注意到, 信息向量 $\Phi(t)$ 持续激励并不能保证每个子系统的信息向量 $\Phi_i(t)$ 持续激励, 式 (4) 和 (5) 不一定能将 $\boldsymbol{\vartheta}$ 中每个参数都辨识出来, 因此不能简单地通过

求 $\hat{\boldsymbol{\vartheta}}_i(t)$ ($i = 1, 2, \dots, m$) 的平均值得到 $\hat{\boldsymbol{\vartheta}}(t)$.

根据耦合辨识概念^[7], 将式 (4) 中第 i ($i = 2, 3, \dots, m$) 个子系统 ($t-1$) 时刻的参数估计值 $\hat{\boldsymbol{\vartheta}}_i(t-1)$ 用前一个子系统 t 时刻的估计值 $\hat{\boldsymbol{\vartheta}}_{i-1}(t)$ 代替, 第 1 个子系统 ($t-1$) 时刻的估计值用第 m 个子系统 ($t-1$) 时刻的估计值 $\hat{\boldsymbol{\vartheta}}_m(t-1)$ 代替, 可得下列耦合随机梯度辨识 (C-SG) 算法:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_i(t) &= \\ \hat{\boldsymbol{\vartheta}}_{i-1}(t) + \frac{\Phi_i^T(t)}{R_i(t)} [y_i(t) - \Phi_i(t)\hat{\boldsymbol{\vartheta}}_{i-1}(t)]; & \quad (6) \end{aligned}$$

$$\begin{aligned} R_i(t) &= \\ R_i(t-1) + \|\Phi_i(t)\|^2, \quad i = 2, 3, \dots, m; & \quad (7) \\ \hat{\boldsymbol{\vartheta}}_1(t) &= \end{aligned}$$

$$\hat{\boldsymbol{\vartheta}}_m(t-1) + \frac{\Phi_1^T(t)}{R_1(t)} [y_1(t) - \Phi_1(t)\hat{\boldsymbol{\vartheta}}_m(t-1)]; \quad (8)$$

$$R_1(t) = R_1(t-1) + \|\Phi_1(t)\|^2. \quad (9)$$

最后一个子系统的参数估计值 $\hat{\boldsymbol{\vartheta}}_m(t)$ 为 t 时刻 $\boldsymbol{\vartheta}$ 的参数估计值.

耦合随机梯度 (C-SG) 辨识算法 (6)~(9) 递推计算不需要计算协方差矩阵, 更无需计算大维协方差矩阵的逆, 因此具有较小的计算量. 递推算法的计算量可用每步计算所需的乘法次数与加法次数来衡量. 表 1 给出了多变量系统模型 (1) 递推最小二乘 (RLS) 算法、耦合递推最小二乘 (C-RLS) 算法和 C-SG 算法的计算量. 大括号中的数据为 $m = 2, n = 10$ 时每次递推计算中所需的计算量, 表 1 中的数据忽略了 RLS 算法中矩阵求逆的计算量, 可以看出, C-SG 算法的计算量最小.

表 1 算法计算量比较

算法	加法次数	乘法次数
RLS	$2mn^2 + 2m^2n + 2mn$	$2mn^2 + 2m^2n$
	{520}	{480}
C-RLS	$2mn^2 + 3mn$	$2mn^2 + 3mn + m$
	{460}	{462}
C-SG	$3mn + m$	$3mn + 2m$
	{62}	{64}

对于 C-SG 算法, 由式 (8), 第 2 个子系统的参数估计值可写为

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_2(t) &= \\ \hat{\boldsymbol{\vartheta}}_1(t) + \frac{\Phi_2^T(t)}{R_2(t)} [y_2(t) - \Phi_2(t)\hat{\boldsymbol{\vartheta}}_1(t)] &= \\ \hat{\boldsymbol{\vartheta}}_m(t-1) + \left[\mathbf{I} - \frac{\Phi_2^T(t)\Phi_2(t)}{R_2(t)} \right] \times & \\ \frac{\Phi_1^T(t)}{R_1(t)} [y_1(t) - \Phi_1(t)\hat{\boldsymbol{\vartheta}}_m(t-1)] + & \end{aligned}$$

$$\frac{\Phi_2^T(t)}{R_2(t)}[y_2(t) - \Phi_2(t)\hat{\vartheta}_m(t-1)].$$

利用式 (6) 依次推算可获得第 m 个子系统的参数估计值 $\hat{\vartheta}_m(t)$ 与 $\hat{\vartheta}_m(t-1)$ 的递推关系为

$$\begin{aligned} \hat{\vartheta}_m(t) = & \hat{\vartheta}_m(t-1) + \prod_{i=2}^m \left[\mathbf{I} - \frac{\Phi_i^T(t)\Phi_i(t)}{R_i(t)} \right] \times \\ & \frac{\Phi_1^T(t)}{R_1(t)}[y_1(t) - \Phi_1(t)\hat{\vartheta}_m(t-1)] + \\ & \prod_{i=3}^m \left[\mathbf{I} - \frac{\Phi_i^T(t)\Phi_i(t)}{R_i(t)} \right] \frac{\Phi_2^T(t)}{R_2(t)}[y_2(t) - \\ & \Phi_2(t)\hat{\vartheta}_m(t-1)] + \dots + \left[\mathbf{I} - \frac{\Phi_m^T(t)\Phi_m(t)}{R_m(t)} \right] \times \\ & \frac{\Phi_{m-1}^T(t)}{R_{m-1}(t)}[y_{m-1}(t) - \Phi_{m-1}(t)\hat{\vartheta}_m(t-1)] + \\ & \frac{\Phi_m^T(t)}{R_m(t)}[y_m(t) - \Phi_m(t)\hat{\vartheta}_m(t-1)]. \end{aligned} \quad (10)$$

由式 (10) 可知, C-SG 算法和 SG 算法并没有像耦合递推最小二乘和递推最小二乘算法那样的等价关系. 为提高算法的收敛速度, 可以在 C-SG 算法中引入遗忘因子 λ , 即将式 (7) 和 (9) 改写为

$$\begin{aligned} R_1(t) &= \lambda R_1(t-1) + \|\Phi_1(t)\|^2, \quad 0 < \lambda \leq 1; \\ R_i(t) &= \lambda R_i(t-1) + \|\Phi_i(t)\|^2, \quad i = 2, 3, \dots, m. \end{aligned}$$

2 性能分析

引理 1 对于耦合随机梯度算法 (6)~(9), 下列不等式成立:

$$\sum_{i=1}^m \sum_{t=1}^{\infty} \frac{\|\Phi_i(t)\|^2}{R_i(t)} < \infty.$$

证明 根据 $R_i(t)$ 的定义可以推导出

$$\begin{aligned} \sum_{i=1}^m \sum_{t=1}^{\infty} \frac{\|\Phi_i(t)\|^2}{R_i^2(t)} &\leq \\ \sum_{i=1}^m \sum_{t=1}^{\infty} \frac{R_i(t) - R_i(t-1)}{R_i(t)R_i(t-1)} &= \\ \sum_{i=1}^m \sum_{t=1}^{\infty} \left[\frac{1}{R_i(t-1)} - \frac{1}{R_i(t)} \right] &= \\ \sum_{i=1}^m \left[\frac{1}{R_i(0)} - \frac{1}{R_i(\infty)} \right] &< \infty. \quad \square \end{aligned}$$

定理 1 对于辨识模型 (1) 和耦合随机梯度算法 (6)~(9), 假设 $\{v_i(t), F_t\}$ ($i = 1, 2, \dots, m$) 为定义在概率空间 $\{\Omega, F, P\}$ 上的随机鞅差序列, $\{F_t\}$ 是由初始时刻直到 t 时刻的观测生成的递增 σ 代数序列, 若子系统噪声满足: 1) $E[v_i(t)|F_{t-1}] = 0$, a.s.; 2) $E[\|v_i(t)\|^2|F_{t-1}] \leq \sigma^2 < \infty$; 3) 数据乘积矩阵 $\mathbf{Q}(t) := \sum_{i=1}^m \Phi^T(i)\Phi(i)$ 和 $R(t) := \|\Phi(t)\|^2$ 满足 $R(t)$

$= O(\lambda_{\min}[\mathbf{Q}(t)])$, a.s., 则参数估计 $\hat{\vartheta}_m$ 一致收敛于真参数 ϑ .

证明 定义子系统参数估计误差

$$\tilde{\vartheta}_i(t) := \hat{\vartheta}_i(t) - \vartheta, \quad i = 1, 2, \dots, m.$$

由式 (6) 和 (8), 可得

$$\begin{aligned} \tilde{\vartheta}_1(t) = & \tilde{\vartheta}_1(t-1) + \frac{\Phi_1^T(t)}{R_1(t)}[y_1(t) - \Phi_1(t)\hat{\vartheta}_m(t-1)] =: \\ & \tilde{\vartheta}_m(t-1) + \frac{\Phi_1^T(t)}{R_1(t)}[-\tilde{y}_1(t) + v_1(t)], \end{aligned} \quad (11)$$

其中

$$\tilde{y}_1(t) := \Phi_1(t)\tilde{\vartheta}_m(t-1).$$

同样可知

$$\tilde{\vartheta}_i(t) = \tilde{\vartheta}_{i-1}(t) + \frac{\Phi_i^T(t)}{R_i(t)}[-\tilde{y}_i(t) + v_i(t)], \quad (12)$$

$$\tilde{y}_i(t) := \Phi_i(t)\tilde{\vartheta}_{i-1}(t), \quad i = 2, 3, \dots, m. \quad (13)$$

由式 (11) 和 (12), 可得

$$\begin{aligned} \tilde{\vartheta}_m(t) = & \tilde{\vartheta}_{m-1}(t) + \frac{\Phi_m^T(t)}{R_m(t)}[-\Phi_m(t)\tilde{\vartheta}_{m-1}(t) + v_m(t)] = \\ & \tilde{\vartheta}_m(t-1) + \frac{\Phi_1^T(t)}{R_1(t)}[-\Phi_1(t)\tilde{\vartheta}_m(t-1) + v_1(t)] + \\ & \sum_{i=2}^m \frac{\Phi_i^T(t)}{R_i(t)}[-\Phi_i(t)\tilde{\vartheta}_{i-1}(t) + v_i(t)]. \end{aligned} \quad (14)$$

对式 (11) 和 (12) 两边取范数, 可得

$$\begin{aligned} \|\tilde{\vartheta}_1(t)\|^2 = & \left\| \tilde{\vartheta}_m(t-1) + \frac{\Phi_1^T(t)}{R_1(t)}[-\tilde{y}_1(t) + v_1(t)] \right\|^2 = \\ & \|\tilde{\vartheta}_m(t-1)\|^2 - \frac{1}{R_1(t)}\tilde{y}_1^2(t) - \frac{R_m(t-1)}{R_1^2(t)}\tilde{y}_1^2(t) + \\ & \frac{2R_m(t-1)}{R_1^2(t)}\tilde{y}_1(t)v_1(t) + \frac{\|\Phi_1(t)\|^2}{R_1^2(t)}v_1^2(t) \leq \\ & \|\tilde{\vartheta}_m(t-1)\|^2 - \frac{1}{R_1(t)}\tilde{y}_1^2(t) + \\ & \frac{R_m(t-1)}{R_1^2(t)}\tilde{y}_1(t)v_1(t) + \frac{\|\Phi_1(t)\|^2}{R_1(t)}v_1^2(t), \end{aligned} \quad (15)$$

$$\begin{aligned} \|\tilde{\vartheta}_i(t)\|^2 = & \|\tilde{\vartheta}_{i-1}(t)\|^2 + \frac{2\tilde{y}_i(t)}{R_i(t)}[-\tilde{y}_i(t) + v_i(t)] + \\ & \frac{\|\Phi_i(t)\|^2}{R_i^2(t)}[\tilde{y}_i^2(t) - 2\tilde{y}_i(t)v_i(t) + v_i^2(t)] = \\ & \|\tilde{\vartheta}_{i-1}(t)\|^2 - \frac{R_i(t) + R_{i-1}(t)}{R_i^2(t)}\tilde{y}_i^2(t) + \\ & \frac{2R_{i-1}(t)}{R_i^2(t)}\tilde{y}_i(t)v_i(t) + \frac{\|\Phi_i(t)\|^2}{R_i^2(t)}v_i^2(t) \leq \\ & \|\tilde{\vartheta}_{i-1}(t)\|^2 - \frac{1}{R_i(t)}\tilde{y}_i^2(t) + \frac{2R_{i-1}(t)}{R_i^2(t)}\tilde{y}_i(t)v_i(t) + \end{aligned}$$

$$\frac{\|\Phi_i(t)\|^2}{R_i^2(t)} v_i^2(t), \quad i = 2, 3 \dots, m. \quad (16)$$

由式(15)和(16),可知

$$\begin{aligned} & \|\tilde{\vartheta}_m(t)\|^2 \leq \\ & \|\tilde{\vartheta}_{m-1}(t)\|^2 - \frac{1}{R_m(t)} \tilde{y}_m^2(t) + \\ & \frac{2R_{m-1}(t)}{R_m^2(t)} \tilde{y}_m(t)v_m(t) + \frac{\|\Phi_m(t)\|^2}{R_m^2(t)} v_m^2(t) \leq \\ & \|\tilde{\vartheta}_1(t)\|^2 + \sum_{i=2}^m \frac{2R_{i-1}(t)}{R_i^2(t)} \tilde{y}_i(t)v_i(t) - \\ & \sum_{i=2}^m \frac{1}{R_i(t)} \tilde{y}_i^2(t) + \sum_{i=2}^m \frac{\|\Phi_i(t)\|^2}{R_i^2(t)} v_i^2(t) \leq \\ & \|\tilde{\vartheta}_m(t-1)\|^2 - \sum_{i=1}^m \frac{1}{R_i(t)} \tilde{y}_i^2(t) + \\ & \sum_{i=1}^m \frac{2R_{i-1}(t)}{R_i^2(t)} \tilde{y}_i(t)v_i(t) + \sum_{i=1}^m \frac{\|\Phi_i(t)\|^2}{R_i^2(t)} v_i^2(t). \quad (17) \end{aligned}$$

由于 $\tilde{\vartheta}_m(t-1)$ 、 $\tilde{y}_i(t)$ 和 $\Phi_i(t)$ 与 $v_i(t)$ 均不相关, 且为 F_{t-1} 可测, 式(17)两边对 F_{t-1} 取条件期望并利用条件 1) 和 2) 可得

$$\begin{aligned} & E[\|\tilde{\vartheta}_m(t)\|^2 | F_{t-1}] \leq \\ & \|\tilde{\vartheta}_m(t-1)\|^2 - \sum_{i=1}^m \frac{1}{R_i(t)} \tilde{y}_i^2(t) + \sum_{i=1}^m \frac{\|\Phi_i(t)\|^2}{R_i^2(t)} \sigma^2. \quad (18) \end{aligned}$$

由引理 1 可知, 式(18)右边最后一项满足鞅收敛定理中的不等式^[13], 对式(18)应用鞅收敛定理可得 $\|\tilde{\vartheta}_m(t)\|^2$ 几乎必然收敛于一有限变量 C_1 , 即

$$\lim_{t \rightarrow \infty} \|\tilde{\vartheta}_m(t)\|^2 = C_1 < \infty, \text{ a.s.}$$

且

$$\sum_{t=1}^{\infty} \frac{\|\tilde{y}(t)\|^2}{R(t)} < \infty, \text{ a.s.} \quad (19)$$

由于 $R_i(t) \rightarrow \infty$, 利用 Kronecker 引理^[13]可知

$$\lim_{t \rightarrow \infty} \frac{1}{R(t)} \sum_{i=1}^t \|\tilde{y}(i)\|^2 = 0, \text{ a.s.} \quad (20)$$

由 $R(t)$ 的定义可知 $R(t) \leq R_i(t)$, 故

$$\lim_{t \rightarrow \infty} \frac{1}{R(t)} \sum_{j=1}^t \|\tilde{y}(j)\|^2 = 0, \text{ a.s.} \quad (21)$$

定义

$$\begin{aligned} \varepsilon_i(t) &= \frac{\Phi_i^T(t)}{R_i(t)} [-\tilde{y}_i(t) + v_i(t)] = \\ & \frac{\Phi_i^T(t)}{R_i(t)} [y_i(t) - \Phi_i(t)\theta], \end{aligned}$$

由式(11)~(13), 可得

$$\tilde{\vartheta}_i(t) = \tilde{\vartheta}_m(t-1) + \sum_{j=1}^i \varepsilon_j(t), \quad (22)$$

$$\tilde{y}_i(t) = \Phi_i(t) \left[\tilde{\vartheta}_m(t-1) + \sum_{j=1}^{i-1} \varepsilon_j(t) \right], \quad (23)$$

因此有

$$\tilde{y}(t) = \Phi(t)\tilde{\vartheta}_m(t-1) + \Xi(t), \quad (24)$$

其中

$$\Xi(t) = \begin{bmatrix} 0 \\ \Phi_2(t)\varepsilon_1(t) \\ \Phi_3(t)[\varepsilon_1(t) + \varepsilon_2(t)] \\ \vdots \\ \Phi_m(t) \sum_{i=1}^{m-1} \varepsilon_i(t) \end{bmatrix} \in \mathbf{R}^m.$$

由式(22)可知

$$\tilde{\vartheta}_m(t) = \tilde{\vartheta}_m(t-j-1) + \sum_{r=0}^j \sum_{i=1}^m \varepsilon_i(t-r). \quad (25)$$

将式(24)中的 t 用 $t-j$ 代替, 得到

$$\tilde{y}(t-j) = \Phi(t-j)\tilde{\vartheta}_m(t-j-1) + \Xi(t-j).$$

由式(25)可知

$$\begin{aligned} & \Phi(t-j)\tilde{\vartheta}_m(t) = \\ & \tilde{y}(t-j) + \Phi(t-j) \sum_{r=0}^j \sum_{i=1}^m \varepsilon_i(t-r) - \Xi(t-j), \end{aligned}$$

两边取范数, 并利用关系式

$$(a+b+c)^2 \leq 2(a^2+b^2+c^2)$$

可得

$$\begin{aligned} & \|\Phi(t-j)\tilde{\vartheta}_m(t)\|^2 = \\ & \left\| \tilde{y}(t-j) + \Phi(t-j) \sum_{r=0}^j \sum_{i=1}^m \varepsilon_i(t-r) - \Xi(t-j) \right\|^2 \leq \\ & 2\|\tilde{y}(t-j)\|^2 + 2\left\| \Phi(t-j) \sum_{r=0}^j \sum_{i=1}^m \varepsilon_i(t-r) \right\|^2 + \\ & 2\|\Xi(t-j)\|^2. \quad (26) \end{aligned}$$

由 $Q(t)$ 的定义可知

$$\lambda_{\min}[Q(t)]\|\tilde{\theta}(t)\|^2 \leq \tilde{\theta}^T(t)Q(t)\tilde{\theta}(t),$$

将式(26)对 j 从 0 到 $t-1$ 求和并除以 $R(t)$ 可得

$$\begin{aligned} & \frac{\lambda_{\min}[Q(t)]\|\tilde{\theta}(t)\|^2}{R(t)} \leq \\ & \frac{\tilde{\vartheta}_m^T(t) \sum_{j=1}^t \Phi^T(j)\Phi(j)\tilde{\vartheta}_m(t)}{R(t)} \leq \\ & \frac{2}{R(t)} \sum_{j=1}^t \|\tilde{y}(j)\|^2 + \frac{2}{R(t)} \sum_{j=1}^t \|\Xi(j)\|^2 + \\ & \frac{2}{R(t)} \sum_{j=0}^{t-1} \left\| \Phi(t-j) \sum_{r=0}^j \sum_{i=1}^m \varepsilon_i(t-r) \right\|^2 \leq \\ & \frac{2}{R(t)} \sum_{j=1}^t \|\tilde{y}(j)\|^2 + \end{aligned}$$

$$\begin{aligned} & \frac{2}{R(t)} \sum_{j=1}^t \|\Phi(j)\|^2 \left\| \sum_{r=j}^t \sum_{i=1}^m \varepsilon_i(r) \right\|^2 + \\ & \frac{2}{R(t)} \sum_{j=1}^t \sum_{r=2}^m \left[\Phi_r(j) \sum_{i=1}^{r-1} \varepsilon_i(j) \right]^2. \end{aligned} \quad (27)$$

假设 $\|\varepsilon_i(t)\|^2$ 有界, 参考文献 [3] 不难得出

$$\|\tilde{\theta}(t)\|^2 = o\left(\frac{R(t)}{\lambda_{\min}[\mathbf{P}(t)]}\right), \text{ a.s.}$$

利用条件 3) 可知 $\|\tilde{\theta}(t)\|^2 = o(1) \rightarrow 0$. \square

3 仿真实验

例 1 考虑两输入两输出多变量系统^[8]

$$\mathbf{y}(t) = \Phi(t)\boldsymbol{\theta} + \mathbf{v}(t).$$

其中: $\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \in \mathbf{R}^2$ 为输出向量; 信息矩阵

$\Phi(t)$ 为

$$\Phi(t) =$$

$$\begin{bmatrix} -y_1(t-1) & y_1(t-2) \sin(y_2(t-2)) & & & & & \\ -y_1(t-1) & y_1(t-2) \sin(t/\pi) & & & & & \\ & y_2(t-1) & y_2(t-2)u_1(t-1) & u_1(t-1) & & & \\ & y_2(t-1) & y_1(t-2)u_2(t-2) & u_1^2(t-1) & & & \\ & u_1(t-2)u_2(t-2) & u_2(t-1) \cos(t) & & & & \\ & \sin(u_2(t-2)) & u_1(t-1)+u_2(t-2) & & & & \end{bmatrix} \in \mathbf{R}^{2 \times 7},$$

$u_1(t)$ 和 $u_2(t)$ 为两个通道输入信号, 仿真时采用零均值单位方差的随机高斯信号. 参数向量真值为

$$\boldsymbol{\theta} =$$

$$[0.67, 0.45, 0.21, -0.78, 0.33, -0.58, 1.26]^T.$$

$\mathbf{v}(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \in \mathbf{R}^2$ 为零均值白噪声向量, 噪声方差为 $\sigma_1^2 = \sigma_2^2 = 0.10^2$.

令相对参数估计误差为

$$\delta := \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|,$$

采用 C-SG 算法估计该系统参数, 当遗忘因子取不同值时, 相对参数估计误差 δ 随数据长度 t 变化的曲线如图 1 所示. 取遗忘因子为 $\lambda = 0.90$, 利用 RLS、SG 和 C-SG 算法分别估计该系统参数, 参数估计误差随数据长度变化曲线如图 2 所示.

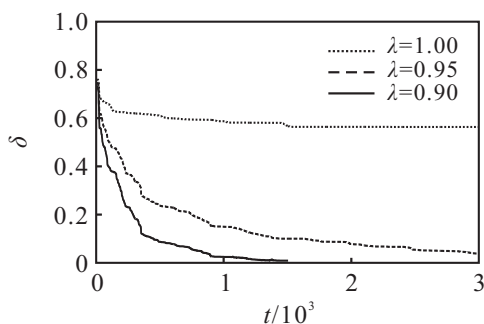


图 1 δ 随 t 的变化曲线 (λ 取不同值)

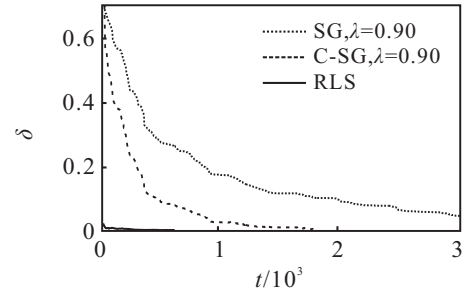


图 2 δ 随 t 的变化曲线 (例 1)

由图 1 和图 2 可见, C-SG 算法虽然收敛速度慢, 但只要选取合适的收敛因子便可以大大改善收敛速度, 并获得良好的参数估计. 在相同条件下, RLS 算法的收敛速度比 SG 和 C-SG 算法要快. 由表 1 可知, RLS 算法的计算量较大, C-SG 算法的收敛速度比 SG 算法快.

例 2 考虑两输入两输出系统

$$\alpha(z)\mathbf{y}(t) = \mathbf{Q}(z)\mathbf{u}(t) + \mathbf{v}(t),$$

$$\alpha(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} = 1 + 0.85z^{-1} + 0.6z^{-2},$$

$$\mathbf{Q}(z) = \mathbf{Q}_1 z^{-1} + \mathbf{Q}_2 z^{-2} =$$

$$\begin{bmatrix} 2.00 & 0.80 \\ 1.00 & 0.70 \end{bmatrix} z^{-1} + \begin{bmatrix} 0.60 & 1.00 \\ 0.80 & 2.00 \end{bmatrix} z^{-2}.$$

其中: $\mathbf{y}(t) \in \mathbf{R}^2$ 为输出向量, $\mathbf{u}(t) \in \mathbf{R}^2$ 为输入向量, $\mathbf{v}(t) \in \mathbf{R}^2$ 为零均值白噪声向量. 令

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2], \boldsymbol{\theta}^T = [\mathbf{Q}_1, \mathbf{Q}_2],$$

系统模型可以写成如下多元回归模型:

$$\mathbf{y}(t) = \Phi(t)\boldsymbol{\vartheta} + \mathbf{v}(t), \boldsymbol{\vartheta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \text{col}[\boldsymbol{\theta}] \end{bmatrix},$$

$$\Phi(t) = [-y(t-1), -y(t-2), \mathbf{I}_2 \otimes \boldsymbol{\varphi}^T(t)] \in \mathbf{R}^{2 \times 7},$$

$$\boldsymbol{\varphi}(t) = [\mathbf{u}^T(t-1), \mathbf{u}^T(t-2)]^T \in \mathbf{R}^4.$$

其中: $\text{col}[\boldsymbol{\theta}]$ 为由 $\boldsymbol{\theta}$ 各列组成的列向量; \otimes 为 Kronecker 积, 如果 $\mathbf{A} = [a_{ij}] \in \mathbf{R}^{m \times n}$, $\mathbf{B} = [b_{ij}] \in \mathbf{R}^{p \times q}$, 则 $\mathbf{A} \otimes \mathbf{B} = [a_{ij} \otimes \mathbf{B}] \in \mathbf{R}^{mp \times nq}$. 仿真条件与例 1 相同, 分别采用 SG 算法和 C-SG 算法对该系统参数进行辨识, 不同遗忘因子下的仿真结果如图 3 和表 2 所示. 仿真结果表明, 当引入遗忘因子时, C-SG 算法有较快的收敛速度和良好的收敛性能, 且在相同条件下比 SG 算法收敛得快.

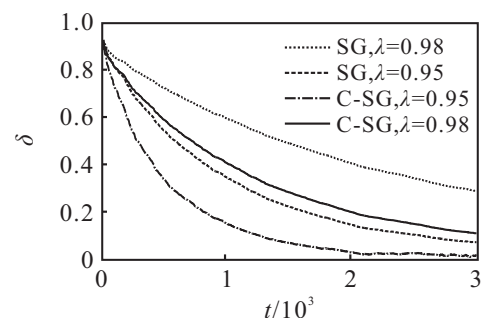


图 3 δ 随 t 的变化曲线 (例 2)

表 2 当 $\lambda = 0.95$, $t = 3000$ 时 SG 和 C-SG 算法的参数估计和相对误差

算法	α_1	α_2	$Q(1, 1)$	$Q(2, 1)$	$Q(1, 2)$	$Q(2, 2)$	$Q(1, 3)$	$Q(2, 3)$	$Q(1, 4)$	$Q(2, 4)$	$\delta/\%$
SG	0.787 03	0.588 91	1.945 27	0.558 74	0.984 27	0.806 32	0.609 31	0.895 46	0.616 89	1.900 06	7.446 29
C-SG	0.813 44	0.609 40	1.991 87	0.566 45	1.015 19	0.826 78	0.765 06	0.970 02	0.698 35	1.999 94	2.074 75
真值	0.850 00	0.600 00	2.000 00	0.600 00	1.000 00	0.800 00	0.800 00	1.000 00	0.700 00	2.000 00	0.000 00

4 结 论

本文推导了多变量线性系统的耦合随机梯度算法, 该算法与最小二乘算法相比具有较小的计算量. 理论分析表明, 当系统噪声满足零均值且方差有界, 并满足一弱持续收敛条件时, 参数估计值收敛到真值. 仿真例子验证了所提出算法具有良好的收敛性. 通过引入适当的遗忘因子可以达到较快的收敛速率.

参考文献(References)

- [1] El-Sherief H. Multivariable system structure and parameter identification using the correlation method[J]. Automatica, 1981, 17(3): 541-544.
- [2] Zheng W X. On a least squares based algorithm for identification of stochastic linear systems[J]. IEEE Trans on Signal Processing, 1998, 46(6): 1631-1638.
- [3] Liu Y J, Sheng J, Ding R F. Convergence of stochastic gradient estimation algorithm for multivariable ARX-like systems[J]. Computers & Mathematics with Applications, 2010, 59(8): 2615-2627.
- [4] 靳其兵, 窦俊, 李斯年, 等. 基于加速收敛与递阶迭代的多变量系统辨识算法[J]. 清华大学学报: 自然科学版, 2012, 52(9): 1194-1198.
(Jin Q B, Dou J, Li S N, et al. Identification method based on accelerated convergence and hierarchical iterations for multivariable systems[J]. J of Tsinghua University: Science and Technology, 2012, 52(9): 1194-1198.)
- [5] 方崇智, 萧德云. 过程辨识[M]. 北京: 清华大学出版社, 1988: 424-430.
(Fang C Z, Xiao D Y. Process identification[M]. Beijing: Tingshua University Press, 1988: 424-430.)
- [6] Ding F, Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems[J]. Automatica, 2005, 41(2): 315-325.
- [7] 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 F. System identification, Part H: Coupled identification concept and methods[J]. J of Nanjing University of Information Science and Technology: Natural Science Edition, 2012, 4(3): 193-212.)
- [8] Ding F. Coupled-least-squares idnetification for multi-variable systems[J]. IET Control Theory and Applications, 2013, 7(1): 68-79.
- [9] Liu Y J, Ding F, Shi Y. A new parameter estimation algorithm for non-uniformly multirate sampled data systems[C]. American Control Conf. San Francisco, 2011: 84-89.
- [10] Ding F, Liu G, Liu X. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems[J]. IEEE Trans on Automatic Control, 2010, 55(8): 1976-1981.
- [11] 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.)
- [12] 丁锋. 系统辨识新论[M]. 北京: 科学出版社, 2013: 349-358.
(Ding F. System Identification—New Theory and Methods[M]. Beijing: Science Press, 2013: 349-358.)
- [13] Goodwin G, Sin K. Adaptive filtering prediction and control[M]. Englewood Cliffs: Prentice-Hall, 1984: 94-98.

(责任编辑: 郑晓蕾)