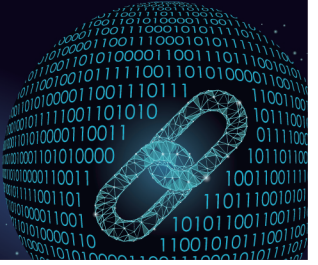




中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 受控自回归系统的带惩罚项梯度辨识算法

孙焕琪, 熊伟丽, 丁锋

引用本文:

孙焕琪,熊伟丽,丁锋. 受控自回归系统的带惩罚项梯度辨识算法[J]. 控制与决策, 2024, 39(8): 2719–2727.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0205>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于KRLS的非均匀采样非线性系统辨识

Identification of non-uniformly sampled nonlinear systems based on KRLS

控制与决策. 2021, 36(12): 3049–3055 <https://doi.org/10.13195/j.kzyjc.2020.0962>

#### 基于高斯混合模型聚类的非均匀采样系统的多模型切换辨识

Multi-model switching identification for non-uniformly sampled systems based on Gaussian mixture model clustering

控制与决策. 2021, 36(12): 2946–2954 <https://doi.org/10.13195/j.kzyjc.2020.0671>

#### 一种基于MOEA/D的组合权重方法

A combination weight method based on MOEA/D

控制与决策. 2021, 36(12): 3056–3062 <https://doi.org/10.13195/j.kzyjc.2020.0592>

#### 双重驱动的果蝇优化算法及其在PID控制器中的应用

Double drive fruit fly optimization algorithm and its application in PID controller

控制与决策. 2021, 36(9): 2225–2233 <https://doi.org/10.13195/j.kzyjc.2020.0046>

#### 工序加工时间不确定的可重入调度问题的求解

Solution of reentrant scheduling problem with uncertain processing time

控制与决策. 2021, 36(5): 1127–1136 <https://doi.org/10.13195/j.kzyjc.2019.1036>

# 受控自回归系统的带惩罚项梯度辨识算法

孙焕琪, 熊伟丽, 丁 锋<sup>†</sup>

(江南大学 物联网工程学院, 江苏 无锡 214122)

**摘 要:** 基于带惩罚项准则函数, 研究受控自回归系统的辨识问题. 首先, 通过负梯度搜索, 极小化带惩罚项的准则函数, 得到计算参数估计的递推关系, 并利用一维线搜索确定最佳步长, 推导带惩罚项投影梯度辨识算法和带惩罚项随机梯度辨识算法; 然后, 为了提高带惩罚项随机梯度算法的收敛速度, 使用多新息辨识理论, 推导带惩罚项多新息随机梯度辨识算法; 最后, 通过仿真实例验证所提出算法的有效性.

**关键词:** 参数估计; 随机梯度; 多新息辨识; 惩罚项; 准则函数

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0205

**引用格式:** 孙焕琪, 熊伟丽, 丁锋. 受控自回归系统的带惩罚项梯度辨识算法[J]. 控制与决策, 2024, 39(8): 2719-2727.

## Gradient identification algorithms based on the penalty criterion function for controlled autoregressive systems

SUN Huan-qi, XIONG Wei-li, DING Feng<sup>†</sup>

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

**Abstract:** Based on the criterion function with penalty, the identification problem of controlled autoregressive systems is studied. Through the gradient search, the recursive relationship of parameter estimation is obtained by minimizing the criterion function with penalty, and the optimal step-size is deducted to obtain the parameter estimation through the one-dimensional line search. The projection identification algorithm with penalty and the stochastic gradient identification algorithm with penalty are proposed. By using the multi-innovation theory, a multi-innovation stochastic gradient identification algorithm with penalty is proposed. The effectiveness of the proposed algorithm is verified by the simulation.

**Keywords:** parameter estimation; stochastic gradient; multi-innovation identification; penalty term; criterion function

## 0 引 言

系统辨识是用系统的输入输出数据建立的数学模型<sup>[1-3]</sup>. 系统辨识理论经过几十年的发展, 其理论和应用成果已广泛应用于电力系统、自适应控制系统、故障诊断、无线通信系统<sup>[4-6]</sup>. 辨识系统领域中, 迭代辨识算法<sup>[7]</sup>和递推辨识算法<sup>[8-10]</sup>是两类重要的辨识方法. 迭代辨识算法使用批量数据, 随着迭代次数的增加更新参数估计, 算法中的数据长度是不变的且迭代变量与时间无关, 多适用于离线辨识; 递推辨识算法随着时间的推移, 边采集数据, 边实时更新参数估计, 因此适用于在线辨识. 递推算法的思想是当前时刻的参数估计等于前一时刻的参数估计加上修正

项<sup>[11]</sup>. 文献[12]针对带外加输入的自回归模型, 提出了加权递阶随机梯度算法, 该算法的修正项在递推过程中以加权形式出现; 文献[13]在陀螺去噪方面, 提出了一种随机加权的递推最小二乘方法, 利用随机加权实现了对偏差的估计.

多新息辨识<sup>[14]</sup>作为系统辨识领域的一个重要分支, 受到了国内外学者的广泛关注. 其基本思想是将标量新息扩展为新息向量或将新息向量扩展为新息矩阵<sup>[15-17]</sup>. 根据系统参数与系统输出响应间的函数关系, 引入多新息辨识理论解决传递函数的辨识问题, 推导出多新息随机梯度算法<sup>[18]</sup>.

在系统辨识领域, 准则函数可用来衡量所建立的

收稿日期: 2023-02-23; 录用日期: 2023-07-17.

基金项目: 国家自然科学基金项目(62273167).

责任编委: 柴利.

<sup>†</sup>通讯作者. E-mail: fding@jiangnan.edu.cn.

数学模型与待辨识系统的接近程度. 对于同一系统的不同辨识算法而言, 准则函数的值越小, 对应的模型与实际系统的接近程度越高. 准则函数通常由模型输出与实际系统输出的误差的平方和构成. 本文针对受控自回归系统, 基于带惩罚项的准则函数, 首先, 推导带惩罚项准则函数的投影梯度算法和随机梯度算法; 然后, 结合多新息辨识理论, 提出带惩罚项准则函数的多新息随机梯度辨识算法; 最后, 通过数值仿真从算法性能和模型验证两个方面验证所提出算法的有效性.

## 1 系统描述与辨识模型

定义有关符号如下: “ $X := A$ ”或“ $A := X$ ”表示  $A$  被定义为  $X$ ; 矩阵范数表示为  $\|X\|^2 := \text{tr}[X^T X]$ ; 上标  $T$  为矩阵的转置;  $I_n$  为  $n$  维的单位阵, 其对角线元素全为 1, 其余元素全为 0;  $\mathbf{1}_n$  表示所有元素全为 1 的  $n$  维列向量. 考虑如下描述的受控自回归 (controlled autoregressive, CAR) 系统:

$$A(z)y(t) = B(z)u(t) + v(t). \quad (1)$$

其中:  $y(t) \in \mathbb{R}$  和  $u(t) \in \mathbb{R}$  分别为系统的输出序列和输入序列;  $v(t) \in \mathbb{R}$  表示均值为 0, 方差为  $\sigma^2$  的随机白噪声序列;  $A(z)$  和  $B(z)$  表示单位后移算子  $z^{-1}$  的多项式, 定义为

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a},$$

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

设阶次  $n_a$  和  $n_b$  是已知的, 记  $n := n_a + n_b$ . 参数向量  $\theta$  和  $\varphi(t)$  定义如下所示:

$$\theta := [a_1, a_2, \dots, a_{n_a}, b_1, b_2, \dots, b_{n_b}]^T \in \mathbb{R}^n,$$

$$\varphi(t) := [-y(t-1), \dots, -y(t-n_a),$$

$$u(t-1), \dots, u(t-n_b)]^T \in \mathbb{R}^n.$$

将  $A(z)$  和  $B(z)$  的表达式代入式 (1), 得到

$$y(t) = [1 - A(z)]y(t) + B(z)u(t) + v(t) = \varphi^T(t)\theta + v(t). \quad (2)$$

## 2 带惩罚项的投影梯度辨识算法

梯度辨识方法是一种重要的系统辨识方法, 其主要思路是利用梯度搜索原理, 沿着准则函数关于系统参数的负梯度方向进行搜索, 求解极小化准则函数的参数估计值. 相较于最小二乘辨识方法, 梯度辨识方法的计算量较小.

在系统辨识中, 当数据乘积矩阵是非奇异的时, 可得到最小二乘解, 这就要求输入  $u(t)$  为满足强持续激励条件的至少  $n$  阶持续激励信号, 为了解决不满足

持续激励条件时的数据乘积矩阵奇异带来的矩阵不可逆问题, 本文提出并推导了准则函数加上惩罚项的梯度辨识算法.

由辨识模型 (2), 构造带惩罚项的准则函数为

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

其中  $\lambda$  为正则化因子. 准则函数  $J_1(\theta)$  对参数向量  $\theta$  的梯度如下所示:

$$\begin{aligned} \text{grad}[J_1(\theta)] &:= \frac{\partial J_1(\theta)}{\partial \theta} = \\ &= \varphi(t)[y(t) - \varphi^T(t)\theta] + \lambda\theta. \end{aligned}$$

设  $\hat{\theta}(t)$  为参数向量  $\theta$  在时刻  $t$  的估计,  $\mu(t)$  为步长, 使用负梯度搜索, 极小化  $J_1(\theta)$ , 得到如下递推关系:

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) - \mu(t)\text{grad}[J(\hat{\theta}(t-1))] = \\ &= \hat{\theta}(t-1) + \mu(t)[\varphi(t)e(t) - \lambda\hat{\theta}(t-1)], \quad (3) \end{aligned}$$

其中  $e(t) := y(t) - \varphi^T(t)\hat{\theta}(t-1) \in \mathbb{R}$  为新息. 运用一维线搜索确定最佳步长  $\mu(t)$ , 将式 (3) 代入  $J_1(\hat{\theta}(t))$ , 得到

$$\begin{aligned} g(\mu(t)) &:= J_1(\hat{\theta}(t)) = \\ &= \frac{1}{2}[y(t) - \varphi^T(t)\hat{\theta}(t)]^2 + \frac{\lambda}{2}\hat{\theta}^T(t)\hat{\theta}(t) = \\ &= \frac{1}{2}\{y(t) - \varphi^T(t)[\hat{\theta}(t-1) + \mu(t)\varphi(t)e(t) - \\ &\quad \lambda\mu(t)\hat{\theta}(t-1)]\}^2 + \frac{\lambda}{2}[\hat{\theta}(t-1) + \\ &\quad \mu(t)\varphi(t)e(t) - \lambda\mu(t)\hat{\theta}(t-1)]^T[\hat{\theta}(t-1) + \\ &\quad \mu(t)\varphi(t)e(t) - \lambda\mu(t)\hat{\theta}(t-1)]. \end{aligned}$$

极小化  $g(\mu(t))$ , 令其导数为 0, 即

$$\begin{aligned} \frac{dg(\mu(t))}{d\mu(t)} &= -[\|\varphi(t)\|^2 e(t) - \lambda\varphi^T(t)\hat{\theta}(t-1)] \times \\ &\quad [e(t) - \mu(t)\|\varphi(t)\|^2 e(t) + \lambda\mu(t)\varphi^T(t) \times \\ &\quad \hat{\theta}(t-1)] + \lambda[\varphi(t)e(t) - \lambda\hat{\theta}(t-1)]^T \times \\ &\quad [\hat{\theta}(t-1) + \mu(t)\varphi(t)e(t) - \\ &\quad \lambda\mu(t)\hat{\theta}(t-1)] = \\ &= -\gamma(t)[e(t) - \mu(t)\gamma(t)] + e(t)\omega(t) - \\ &\quad \lambda^2\|\hat{\theta}(t-1)\|^2 + \lambda\mu(t)[e(t)\|\varphi(t)\| - \\ &\quad \lambda\|\hat{\theta}(t-1)\|]^2 = \\ &= [-\gamma(t) + \omega(t)]e(t) - \lambda^2\|\hat{\theta}(t-1)\|^2 + \\ &\quad \mu(t)\gamma^2(t) + \lambda\mu(t)[e(t)\|\varphi(t)\| - \\ &\quad \lambda\|\hat{\theta}(t-1)\|]^2 = 0. \end{aligned}$$

由此可求得最佳步长为

$$\mu(t) = \frac{[\gamma(t) - \omega(t)]e(t) + \lambda^2\|\hat{\theta}(t-1)\|^2}{\lambda[e(t)\|\varphi(t)\| - \lambda\|\hat{\theta}(t-1)\|]^2 + \gamma^2(t)}. \quad (4)$$

其中:  $\gamma(t)$  和  $\omega(t)$  分别定义为

$$\gamma(t) := \|\varphi(t)\|^2 e(t) - \lambda \varphi^T(t) \hat{\theta}(t-1) \in \mathbb{R},$$

$$\omega(t) := \lambda \varphi^T(t) \hat{\theta}(t-1) \in \mathbb{R}.$$

联立式(3)和(4)得到估计参数向量  $\theta$  的带惩罚项投影辨识算法 (projection algorithm with penalty, P-PJ) 为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu(t) [\varphi(t)e(t) - \lambda \hat{\theta}(t-1)], \quad (5)$$

$$\mu(t) = \frac{[\gamma(t) - \omega(t)]e(t) + \lambda^2 \|\hat{\theta}(t-1)\|^2}{\lambda [e(t)\|\varphi(t)\| - \lambda \|\hat{\theta}(t-1)\|]^2 + \gamma^2(t)}, \quad (6)$$

$$e(t) = y(t) - \varphi^T(t) \hat{\theta}(t-1), \quad (7)$$

$$\gamma(t) = \|\varphi(t)\|^2 e(t) - \lambda \varphi^T(t) \hat{\theta}(t-1), \quad (8)$$

$$\omega(t) = \lambda \varphi^T(t) \hat{\theta}(t-1), \quad (9)$$

$$\varphi(t) = [-y(t-1), \dots, -y(t-n_a), \\ u(t-1), \dots, u(t-n_b)]^T. \quad (10)$$

带惩罚项的投影辨识算法(5)~(10)计算估计  $\hat{\theta}(t)$  的步骤如下.

step 1: 初始化: 令  $t=1$ , 给定数据长度  $L$  和正则化常数  $\lambda$ , 设置初值  $\hat{\theta}(0) = \mathbf{1}_n/p_0$ ,  $u(t-i)=0$ ,  $y(t-i)=0$  ( $i=1, 2, \dots, \max[n_a, n_b]$ ),  $p_0$  为一个很大的数, 如  $p_0=10^6$ .

step 2: 采集输入输出数据  $u(t)$  和  $y(t)$ , 由式(10)构造信息向量  $\varphi(t)$ .

step 3: 由式(8)计算  $\gamma(t)$ , 由式(9)计算  $\omega(t)$ .

step 4: 由式(6)计算步长  $\mu(t)$ , 由式(7)计算标量新息  $e(t)$ .

step 5: 由式(5)刷新参数估计向量  $\hat{\theta}(t)$ .

step 6: 若  $t < L$ , 则  $t$  增加 1, 转至 step 2; 否则输出参数估计  $\hat{\theta}(t)$ , 终止递推计算过程.

对于带惩罚项的投影辨识算法(5)~(9), 新息  $e(t)$  与残差  $\hat{v}(t) := y(t) - \varphi^T(t) \hat{\theta}(t) \in \mathbb{R}$  有以下关系:

$$\hat{v}(t) =$$

$$e(t) - \mu(t) \|\varphi(t)\|^2 e(t) + \lambda \mu(t) \varphi^T(t) \hat{\theta}(t-1),$$

$$e(t) = \frac{\hat{v}(t) - \lambda \mu(t) \varphi^T(t) \hat{\theta}(t-1)}{1 - \mu(t) \|\varphi(t)\|^2}.$$

将式(3)代入  $\hat{v}(t)$ , 得到

$$\hat{v}(t) = y(t) - \varphi^T(t) \hat{\theta}(t) =$$

$$y(t) - \varphi^T(t) [\hat{\theta}(t-1) + \mu(t) \varphi(t) e(t) -$$

$$\lambda \mu(t) \hat{\theta}(t-1)] =$$

$$e(t) - \mu(t) \|\varphi(t)\|^2 e(t) +$$

$$\lambda \mu(t) \varphi^T(t) \hat{\theta}(t-1).$$

### 3 带惩罚项的随机梯度辨识算法

通过上述投影算法可知, 对于时不变参数随机系统, 要获得一致参数估计, 步长  $\mu(t)$  必须随数据量增加而趋于 0. 取  $\lambda=0$  时, 步长为  $\mu(t) = \frac{1}{\|\varphi(t)\|^2}$ . 此时步长  $\mu(t)$  只与  $\|\varphi(t)\|$  的大小有关. 通过调节  $\lambda$ , 可控制参数估计的幅值, 使得参数估计的幅值变化趋于平缓, 避免因增益过小使得分母趋于 0, 或因增益过大步长减小得太快, 影响算法的收敛速度. 因此步长可选择为

$$\mu(t) = \frac{1}{r(t)},$$

$$r(t) =$$

$$r(t-1) + \frac{\lambda [e(t)\|\varphi(t)\| - \lambda \|\hat{\theta}(t-1)\|]^2 + \gamma^2(t)}{[\gamma(t) - \omega(t)]e(t) + \lambda^2 \|\hat{\theta}(t-1)\|^2}.$$

联立式(5)~(10)得到估计参数向量  $\theta$  的带惩罚项随机梯度算法 (stochastic gradient algorithm with penalty, P-SG) 为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{r(t)} [\varphi(t)e(t) - \lambda \hat{\theta}(t-1)], \quad (11)$$

$$r(t) =$$

$$r(t-1) + \frac{\lambda [e(t)\|\varphi(t)\| - \lambda \|\hat{\theta}(t-1)\|]^2 + \gamma^2(t)}{[\gamma(t) - \omega(t)]e(t) + \lambda^2 \|\hat{\theta}(t-1)\|^2}, \quad (12)$$

$$e(t) = y(t) - \varphi^T(t) \hat{\theta}(t-1), \quad (13)$$

$$\gamma(t) = \|\varphi(t)\|^2 e(t) - \lambda \varphi^T(t) \hat{\theta}(t-1), \quad (14)$$

$$\omega(t) = \lambda \varphi^T(t) \hat{\theta}(t-1), \quad (15)$$

$$\varphi(t) = [-y(t-1), \dots, -y(t-n_a),$$

$$u(t-1), \dots, u(t-n_b)]^T. \quad (16)$$

带惩罚项的随机梯度算法(11)~(16)计算估计  $\hat{\theta}(t)$  的步骤如下.

step 1: 初始化: 令  $t=1$ , 给定数据长度  $L$ , 设置初值  $\hat{\theta}(0) = \mathbf{1}_n/p_0$ ,  $r(0) = 1$ ,  $u(t-i)=0$ ,  $y(t-i)=0$  ( $i=1, 2, \dots, \max[n_a, n_b]$ ),  $p_0$  为一个很大的数, 如  $p_0=10^6$ . 设置正则化常数  $\lambda$ .

step 2: 采集观测的输入输出数据  $u(t)$  和  $y(t)$ , 由式(16)构造信息向量  $\varphi(t)$ .

step 3: 由式(14)计算  $\gamma(t)$ , 由式(15)计算  $\omega(t)$ .

step 4: 由式(12)计算  $r(t)$ , 由式(13)计算新息  $e(t)$ .

step 5: 由式(11)刷新参数估计  $\hat{\theta}(t)$ .

step 6: 若  $t < L$ , 则  $t$  增加 1, 转至 step 2; 否则输出参数估计  $\hat{\theta}(t)$ , 终止递推计算过程.

#### 4 带惩罚项的多新息随机梯度辨识算法

随机梯度算法的优点是大大减少了计算量,但是收敛速度较慢,为了提高参数估计的收敛速度,利用多新息辨识理论,推导了带惩罚项的多新息随机梯度算法,因为多新息随机梯度算法每次计算不仅使用当前数据和新息,还会使用有限的过去数据和新息,从而提高参数估计的精度. 用正整数  $p$  表征新息长度,考虑从  $j = t - p + 1$  到  $j = t$  的  $p$  个数据,并将堆积输出向量  $\mathbf{Y}(p, t)$  和堆积信息矩阵  $\Phi(p, t)$  定义为

$$\mathbf{Y}(p, t) := [y(t), y(t-1), y(t-2), \dots, y(t-p+1)]^T \in \mathbb{R}^p, \quad (17)$$

$$\Phi(p, t) := [\varphi(t), \varphi(t-1), \varphi(t-2), \dots, \varphi(t-p+1)] \in \mathbb{R}^{n \times p}. \quad (18)$$

将新息标量  $e(t) = y(t) - \varphi^T(t)\hat{\theta}(t-1)$  扩展为新息向量,有

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

定义梯度准则函数为

$$J_2(\theta) := \frac{1}{2} \|\mathbf{Y}(p, t) - \Phi^T(p, t)\theta\|^2 + \frac{\lambda}{2} \theta^T \theta.$$

准则函数  $J_2(\theta)$  相对于参数向量  $\theta$  的梯度如下所示:

$$\begin{aligned} \text{grad}[J_2(\theta)] &:= \frac{\partial J_2(\theta)}{\partial \theta} = \\ &= -\Phi(p, t)[\mathbf{Y}(p, t) - \Phi^T(p, t)\theta] + \lambda\theta. \end{aligned}$$

设  $\mu(t)$  为步长,使用负梯度搜索,极小化  $J_2(\theta)$ ,得到如下递推关系:

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) - \mu(t) \text{grad}[J_2(\hat{\theta}(t-1))] = \\ &= \hat{\theta}(t-1) + \mu(t) [\Phi(p, t)\mathbf{E}(p, t) - \\ &\quad \lambda\hat{\theta}(t-1)]. \end{aligned} \quad (19)$$

运用一维线搜索确定最佳步长  $\mu(t)$ ,将式(19)代入  $J_2(\hat{\theta}(t))$ ,得到

$$\begin{aligned} g(\mu(t)) &:= J_2(\hat{\theta}(t)) = \\ &= \frac{1}{2} \|\mathbf{Y}(p, t) - \Phi^T(p, t)\hat{\theta}(t)\|^2 + \frac{\lambda}{2} \hat{\theta}^T(t)\hat{\theta}(t) = \\ &= \frac{1}{2} \|\mathbf{Y}(p, t) - \Phi^T(p, t)[\hat{\theta}(t-1) + \\ &\quad \mu(t)\Phi(p, t)\mathbf{E}(p, t) - \lambda\mu(t)\hat{\theta}(t-1)]\|^2 + \\ &\quad \frac{\lambda}{2} [\hat{\theta}(t-1) + \mu(t)\Phi(p, t)\mathbf{E}(p, t) - \\ &\quad \lambda\mu(t)\hat{\theta}(t-1)]^T [\hat{\theta}(t-1) + \\ &\quad \mu(t)\Phi(p, t)\mathbf{E}(p, t) - \lambda\mu(t)\hat{\theta}(t-1)] = \\ &= \frac{1}{2} \|\mathbf{E}(p, t) - \mu(t)\Gamma(t)\|^2 + \frac{\lambda}{2} [\hat{\theta}(t-1) + \end{aligned}$$

$$\begin{aligned} &\mu(t)\Phi(p, t)\mathbf{E}(p, t) - \lambda\mu(t)\hat{\theta}(t-1)]^T \times \\ &[\hat{\theta}(t-1) + \mu(t)\Phi(p, t)\mathbf{E}(p, t) - \\ &\quad \lambda\mu(t)\hat{\theta}(t-1)]. \end{aligned}$$

极小化  $g(\mu(t))$ ,令其导数为0,即

$$\begin{aligned} \frac{dg(\mu(t))}{d\mu(t)} &= \\ &= -\Gamma^T(t)[\mathbf{E}(p, t) - \mu(t)\Gamma(t)] + \lambda[\Phi(p, t) \times \\ &\quad \mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)]^T \hat{\theta}(t-1) + \\ &\quad \lambda[\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)]^T \times \\ &\quad \mu(t)[\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)] = \\ &= -\Gamma^T(t)[\mathbf{E}(p, t) - \mu(t)\Gamma(t)] + \mathbf{E}^T(p, t) \times \\ &\quad \Omega(t) - \lambda^2 \|\hat{\theta}(t-1)\|^2 + \\ &\quad \lambda\mu(t) \|\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)\|^2 = \\ &= [-\Gamma(t) + \Omega(t)]^T \mathbf{E}(p, t) - \lambda^2 \|\hat{\theta}(t-1)\|^2 + \\ &\quad \mu(t) \|\Gamma(t)\|^2 + \lambda\mu(t) \|\Phi(p, t)\mathbf{E}(p, t) - \\ &\quad \lambda\hat{\theta}(t-1)\|^2 = 0. \end{aligned}$$

由此可求得最佳步长为

$$\mu(t) = \frac{[\Gamma(t) - \Omega(t)]^T \mathbf{E}(p, t) + \lambda^2 \|\hat{\theta}(t-1)\|^2}{\lambda \|\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)\|^2 + \|\Gamma(t)\|^2}. \quad (20)$$

其中:  $\Gamma(t)$  和  $\Omega(t)$  分别定义为

$$\Gamma(t) := \|\Phi(p, t)\|^2 \mathbf{E}(p, t) - \lambda \Phi^T(p, t)\hat{\theta}(t-1) \in \mathbb{R}^p,$$

$$\Omega(t) := \lambda \Phi^T(p, t)\hat{\theta}(t-1) \in \mathbb{R}^p.$$

联立式(17)~(20)得到估计参数向量  $\theta$  的带惩罚项多新息随机梯度算法 (multi-innovation stochastic gradient algorithm with penalty, P-MISG) 为

$$\begin{aligned} \hat{\theta}(t) &= \\ &= \hat{\theta}(t-1) + \frac{1}{r(t)} [\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)], \end{aligned} \quad (21)$$

$$\begin{aligned} r(t) &= r(t-1) + \\ &+ \frac{\lambda \|\Phi(p, t)\mathbf{E}(p, t) - \lambda\hat{\theta}(t-1)\|^2 + \|\Gamma(t)\|^2}{[\Gamma(t) - \Omega(t)]^T \mathbf{E}(p, t) + \lambda^2 \|\hat{\theta}(t-1)\|^2}, \end{aligned} \quad (22)$$

$$\mathbf{E}(p, t) = \mathbf{Y}(p, t) - \Phi^T(p, t)\hat{\theta}(t-1), \quad (23)$$

$$\Gamma(t) = \|\Phi(p, t)\|^2 \mathbf{E}(p, t) - \lambda \Phi^T(p, t)\hat{\theta}(t-1), \quad (24)$$

$$\Omega(t) = \lambda \Phi^T(p, t)\hat{\theta}(t-1), \quad (25)$$

$$\begin{aligned} \varphi(t) &= [-y(t-1), \dots, -y(t-n_a), \\ &\quad u(t-1), \dots, u(t-n_b)]^T. \end{aligned} \quad (26)$$

带惩罚项的多新息随机梯度算法(21)~(26)计算估计 $\hat{\theta}(t)$ 的步骤如下.

step 1: 设 $t=1$ , 给定数据长度 $L$ 和新息长度 $p$ , 设置正则化常数 $\lambda$ . 设置初值 $\hat{\theta}(0) = \mathbf{1}_n/p_0$ ,  $r(0) = 1$ ,  $u(t-i)=0$ ,  $y(t-i)=0$  ( $i=1, 2, \dots, \max[n_a, n_b]$ ),  $p_0$ 为一个很大的数, 如 $p_0=10^6$ .

step 2: 采集观测的输入输出数据 $u(t)$ 和 $y(t)$ , 由式(26)构造信息向量 $\varphi(t)$ .

step 3: 由式(17)构造堆积输出向量 $\mathbf{Y}(p, t)$ , 由式(18)构造堆积信息矩阵 $\Phi(p, t)$ .

step 4: 由式(24)计算 $\Gamma(t)$ , 由式(25)计算 $\Omega(t)$ .

step 5: 由式(22)计算步长 $r(t)$ , 由式(23)计算新息向量 $\mathbf{E}(p, t)$ .

step 6: 由式(21)刷新参数估计 $\hat{\theta}(t)$ .

step 7: 若 $t < L$ , 则 $t$ 增加1, 转至step 2; 否则输出参数估计 $\hat{\theta}(t)$ , 终止递推计算过程.

对于多新息随机梯度算法新息向量 $\mathbf{E}(p, t)$ 与残差向量 $\hat{\mathbf{V}}(p, t) := \mathbf{Y}(p, t) - \Phi^T(p, t)\hat{\theta}(t) \in \mathbb{R}^p$ 具有以下关系:

$$\begin{aligned}\hat{\mathbf{V}}(p, t) &= \mathbf{E}(p, t) - \mu(t)\|\Phi(p, t)\|^2\mathbf{E}(p, t) + \\ &\quad \lambda\mu(t)\Phi^T(p, t)\hat{\theta}(t-1), \\ \mathbf{E}(p, t) &= \frac{\hat{\mathbf{V}}(p, t) - \lambda\mu(t)\Phi^T(p, t)\hat{\theta}(t-1)}{1 - \mu(t)\|\Phi(p, t)\|^2}.\end{aligned}$$

由式(19), 得到

$$\begin{aligned}\hat{\mathbf{V}}(p, t) &= \mathbf{Y}(p, t) - \Phi^T(p, t)\hat{\theta}(t) = \\ &\quad \mathbf{Y}(p, t) - \Phi^T(p, t)[\hat{\theta}(t-1) + \\ &\quad \mu(t)\Phi(p, t)\mathbf{E}(p, t) - \lambda\mu(t)\hat{\theta}(t-1)] = \\ &\quad \mathbf{E}(p, t) - \mu(t)[\|\Phi(p, t)\|^2\mathbf{E}(p, t) - \\ &\quad \lambda\Phi^T(p, t)\hat{\theta}(t-1)].\end{aligned}$$

通过比较P-PJ、P-SG与P-MISG算法的增益可以发现: P-PJ算法的增益过大, 对噪声更加敏感, 参数估计精度易受到影响; P-MISG算法相较于P-SG算法, 虽然计算量随着新息长度的增加会逐渐增加, 但是能够提高数据的使用效率, 使得收敛速度更快. P-MISG算法是在同时保障计算量和收敛速度下的折中选择.

## 5 数值仿真

数值仿真是验证算法的有效性和准确性的重要方法. 考虑如下CAR模型仿真对象:

$$A(z)y(t) = B(z)u(t) + v(t),$$

$$\begin{aligned}A(z) &= 1 + a_1z^{-1} + a_2z^{-2} + a_3z^{-3} = \\ &\quad 1 - 0.33z^{-1} + 0.13z^{-2} + 0.50z^{-3},\end{aligned}$$

$$\begin{aligned}B(z) &= b_1z^{-1} + b_2z^{-2} + b_3z^{-3} = \\ &\quad 0.37z^{-1} + 0.30z^{-2} + 0.31z^{-3},\end{aligned}$$

$$\begin{aligned}\theta &= [a_1, a_2, a_3, b_1, b_2, b_3]^T = \\ &\quad [-0.33, 0.13, 0.50, 0.37, 0.30, 0.31]^T.\end{aligned}$$

仿真时, 系统输入 $\{u(t)\}$ 采用0均值且单位方差不相关的可测随机信号序列,  $\{v(t)\}$ 采用0均值方差为 $\sigma^2$ 的白噪声序列, 在给定模型参数下产生系统的输出序列 $\{y(t)\}$ . 算法在不同新息长度、不同噪声方差、不同正则化因子下的性能仿真和模型验证如下, 以此来验证所提出算法的有效性和准确性.

### 1) P-PJ算法的数值仿真.

取 $\sigma^2 = 0.10^2$ 时, 分别在不同正则化因子下和不同噪声方差条件下的P-PJ算法对系统进行辨识, 不同正则化因子和方差下参数估计误差随 $t$ 变化曲线如图1和图2所示.

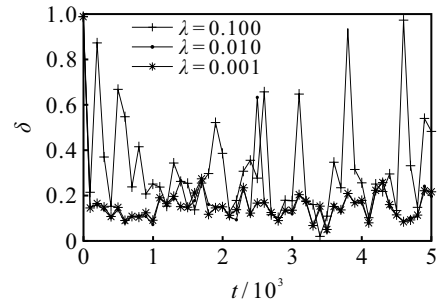


图1 P-PJ参数估计误差 $\delta$ 随 $t$ 变化的曲线( $\sigma^2 = 0.10^2$ )

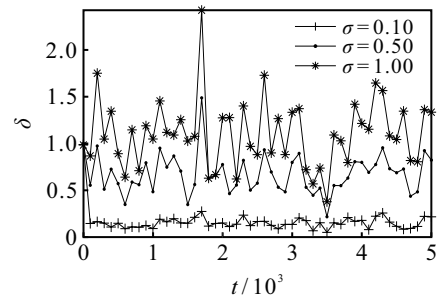


图2 P-PJ参数估计误差 $\delta$ 随 $t$ 变化的曲线( $\lambda = 0.001$ )

由图1和图2可见, 参数估计误差波动较大, 表明P-PJ算法对噪声敏感. 因此, 对P-SG和P-MISG算法在不同新息长度、不同噪声方差和不同正则化因子条件下进行仿真, 并比较它们的性能.

### 2) 不同新息长度下的性能仿真.

当新息长度 $p=1$ 时, P-MISG算法退化为P-SG算法. 取 $\sigma^2 = 0.10^2$ ,  $\lambda = 0.01$ , 新息长度分别为 $p=1$ 、 $p=2$ 、 $p=4$ 的P-MISG算法(21)~(26)估计这个系统的参数, 仿真结果如表1、表2和图3所示.

由表1和表2可见, 随着数据长度的增加, 参数估计误差逐渐减小. 从图3中的曲线可以看出, 当 $p=4$ 时参数估计精度高于 $p=1$ 和 $p=2$ 时, 这意味着大的

表1 P-SG参数估计和误差( $\sigma^2=0.10^2, \lambda=0.01$ )

| $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|-------|-----------|----------|----------|----------|----------|----------|-------------|
| 100   | -0.242 42 | 0.048 57 | 0.385 34 | 0.201 39 | 0.194 16 | 0.177 54 | 34.799 38   |
| 200   | -0.249 18 | 0.048 60 | 0.393 56 | 0.222 57 | 0.199 79 | 0.189 94 | 31.802 86   |
| 500   | -0.268 14 | 0.052 81 | 0.412 52 | 0.245 88 | 0.218 32 | 0.209 28 | 26.657 84   |
| 1 000 | -0.278 41 | 0.058 98 | 0.426 19 | 0.261 06 | 0.231 86 | 0.223 35 | 23.064 01   |
| 2 000 | -0.286 43 | 0.062 62 | 0.435 73 | 0.274 28 | 0.241 29 | 0.233 57 | 20.378 22   |
| 5 000 | -0.295 14 | 0.069 15 | 0.449 72 | 0.289 51 | 0.252 33 | 0.243 84 | 17.158 42   |
| 真值    | -0.330 00 | 0.130 00 | 0.500 00 | 0.370 00 | 0.300 00 | 0.310 00 |             |

表2 P-MISG参数估计和误差( $\sigma^2=0.10^2, \lambda=0.01$ )

| $p$ | $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|-----|-------|-----------|----------|----------|----------|----------|----------|-------------|
| 2   | 100   | -0.344 39 | 0.092 29 | 0.452 38 | 0.240 19 | 0.280 64 | 0.180 52 | 23.284 32   |
|     | 200   | -0.342 20 | 0.089 59 | 0.457 54 | 0.269 95 | 0.269 65 | 0.200 42 | 19.481 84   |
|     | 500   | -0.350 72 | 0.092 56 | 0.469 30 | 0.294 53 | 0.273 45 | 0.223 34 | 15.451 45   |
|     | 1 000 | -0.351 70 | 0.098 70 | 0.477 07 | 0.309 89 | 0.277 39 | 0.238 36 | 12.678 66   |
|     | 2 000 | -0.352 72 | 0.101 86 | 0.481 61 | 0.322 58 | 0.279 43 | 0.249 12 | 10.715 95   |
|     | 5 000 | -0.352 13 | 0.107 47 | 0.488 06 | 0.335 12 | 0.281 91 | 0.259 01 | 8.695 35    |
| 4   | 100   | -0.326 47 | 0.120 42 | 0.490 75 | 0.355 18 | 0.301 54 | 0.312 57 | 2.446 89    |
|     | 200   | -0.320 52 | 0.122 37 | 0.488 24 | 0.364 48 | 0.296 25 | 0.308 27 | 2.185 28    |
|     | 500   | -0.326 82 | 0.121 44 | 0.492 63 | 0.368 13 | 0.295 26 | 0.312 13 | 1.551 46    |
|     | 1 000 | -0.326 35 | 0.124 82 | 0.495 09 | 0.367 72 | 0.296 58 | 0.312 42 | 1.115 48    |
|     | 2 000 | -0.326 99 | 0.125 72 | 0.496 22 | 0.370 22 | 0.297 08 | 0.312 48 | 0.897 85    |
|     | 5 000 | -0.327 09 | 0.127 14 | 0.497 69 | 0.371 27 | 0.298 38 | 0.310 93 | 0.622 36    |
| 真值  |       | -0.330 00 | 0.130 00 | 0.500 00 | 0.370 00 | 0.300 00 | 0.310 00 |             |

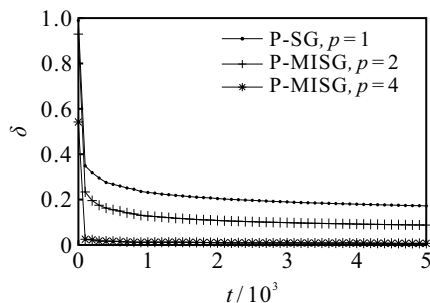


图3 P-SG和P-MISG参数估计误差 $\delta$ 随 $t$ 变化的曲线  
( $\sigma^2=0.10^2, \lambda=0.01$ )

新息长度可获得更高的估计精度,且P-MISG算法比P-SG算法收敛速度快.

3) 不同噪声方差下算法的性能仿真.

噪声方差可控制信噪比,为了比较不同噪声方差对算法估计精度的影响,取新息长度 $p=4$ ,正则化因子 $\lambda=0.01$ ,噪声方差分别为 $\sigma^2=1.00^2$ 、 $\sigma^2=0.50^2$ 和 $\sigma^2=0.10^2$ 的P-MISG算法对系统进行辨识,结果如表3和图4所示.

表3 P-MISG参数估计及误差( $p=4, \lambda=0.01$ )

| $\sigma^2$ | $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|------------|-------|-----------|----------|----------|----------|----------|----------|-------------|
| $1.00^2$   | 100   | -0.394 88 | 0.055 33 | 0.568 67 | 0.212 44 | 0.303 12 | 0.249 94 | 24.789 15   |
|            | 200   | -0.361 51 | 0.057 62 | 0.543 64 | 0.265 47 | 0.273 53 | 0.236 39 | 18.983 03   |
|            | 500   | -0.357 08 | 0.086 95 | 0.536 85 | 0.300 15 | 0.260 30 | 0.263 44 | 13.411 01   |
|            | 1 000 | -0.334 64 | 0.100 89 | 0.519 58 | 0.309 21 | 0.265 23 | 0.275 36 | 10.260 59   |
|            | 2 000 | -0.344 21 | 0.105 60 | 0.510 90 | 0.330 82 | 0.266 54 | 0.284 17 | 7.787 59    |
|            | 5 000 | -0.338 61 | 0.118 72 | 0.510 81 | 0.346 31 | 0.271 79 | 0.284 03 | 5.798 86    |
| $0.50^2$   | 100   | -0.363 53 | 0.082 62 | 0.546 24 | 0.289 17 | 0.313 83 | 0.289 65 | 13.452 03   |
|            | 200   | -0.339 47 | 0.083 39 | 0.527 78 | 0.329 67 | 0.288 60 | 0.276 72 | 9.186 66    |
|            | 500   | -0.348 16 | 0.088 52 | 0.526 89 | 0.350 70 | 0.278 44 | 0.295 31 | 7.399 89    |
|            | 1 000 | -0.334 70 | 0.100 71 | 0.519 93 | 0.353 44 | 0.282 21 | 0.300 17 | 5.302 00    |
|            | 2 000 | -0.336 92 | 0.104 25 | 0.515 20 | 0.366 82 | 0.283 39 | 0.303 56 | 4.262 41    |
|            | 5 000 | -0.332 20 | 0.112 23 | 0.513 90 | 0.373 61 | 0.287 73 | 0.300 25 | 3.325 12    |

表3 (续)

| $\sigma^2$ | $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|------------|-------|-----------|----------|----------|----------|----------|----------|-------------|
| $0.10^2$   | 100   | -0.326 47 | 0.120 42 | 0.490 75 | 0.355 18 | 0.301 54 | 0.312 57 | 2.446 89    |
|            | 200   | -0.320 52 | 0.122 37 | 0.488 24 | 0.364 48 | 0.296 25 | 0.308 27 | 2.185 28    |
|            | 500   | -0.326 82 | 0.121 44 | 0.492 63 | 0.368 13 | 0.295 26 | 0.312 13 | 1.551 46    |
|            | 1 000 | -0.326 35 | 0.124 82 | 0.495 09 | 0.367 72 | 0.296 58 | 0.312 42 | 1.115 48    |
|            | 2 000 | -0.326 99 | 0.125 72 | 0.496 22 | 0.370 22 | 0.297 08 | 0.312 48 | 0.897 85    |
|            | 5 000 | -0.327 09 | 0.127 14 | 0.497 69 | 0.371 27 | 0.298 38 | 0.310 93 | 0.622 36    |
| 真值         |       | -0.330 00 | 0.130 00 | 0.500 00 | 0.370 00 | 0.300 00 | 0.310 00 |             |

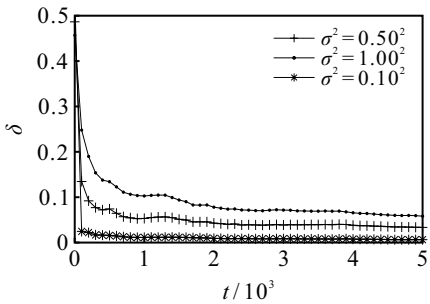


图4 P-MISG 参数估计误差 $\delta$ 随 $t$ 变化的曲线  
( $p=4, \lambda=0.01$ )

由表3和图4可见,噪声方差较小时比噪声方差较大时的参数估计精度高. 因此,噪声也是影响辨识方法估计精度的关键因素.

4) 不同正则化因子下算法的性能仿真.

考虑到正则化因子 $\lambda$ 对系统辨识结果的影响,取新息长度 $p=4$ ,噪声方差为 $\sigma^2=0.10^2$ ,正则化因子分别为 $\lambda=0.10$ 、 $\lambda=0.05$ 和 $\lambda=0.01$ 下的P-MISG算法进行辨识,结果如表4和图5所示.

表4 P-MISG 参数估计及其误差( $p=4, \sigma=0.10$ )

| $\lambda$ | $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|-----------|-------|-----------|----------|----------|----------|----------|----------|-------------|
| 0.10      | 100   | -0.323 80 | 0.119 63 | 0.484 01 | 0.346 05 | 0.298 98 | 0.306 15 | 3.765 50    |
|           | 200   | -0.316 72 | 0.120 49 | 0.479 63 | 0.355 63 | 0.291 90 | 0.301 89 | 3.820 02    |
|           | 500   | -0.322 76 | 0.119 19 | 0.482 82 | 0.359 30 | 0.290 52 | 0.305 61 | 3.138 63    |
|           | 1 000 | -0.322 30 | 0.122 64 | 0.484 94 | 0.359 07 | 0.291 87 | 0.306 38 | 2.776 93    |
|           | 2 000 | -0.322 75 | 0.123 36 | 0.485 29 | 0.361 54 | 0.292 16 | 0.306 46 | 2.562 19    |
|           | 5 000 | -0.322 98 | 0.124 99 | 0.486 79 | 0.362 78 | 0.293 40 | 0.304 96 | 2.301 41    |
| 0.05      | 100   | -0.325 26 | 0.120 00 | 0.487 70 | 0.351 09 | 0.300 29 | 0.309 60 | 3.006 13    |
|           | 200   | -0.318 84 | 0.121 46 | 0.484 37 | 0.360 47 | 0.294 25 | 0.305 34 | 2.899 22    |
|           | 500   | -0.325 00 | 0.120 38 | 0.488 22 | 0.364 12 | 0.293 10 | 0.309 14 | 2.203 21    |
|           | 1 000 | -0.324 54 | 0.123 79 | 0.490 53 | 0.363 81 | 0.294 45 | 0.309 66 | 1.803 51    |
|           | 2 000 | -0.325 09 | 0.124 62 | 0.491 31 | 0.366 29 | 0.294 86 | 0.309 73 | 1.554 54    |
|           | 5 000 | -0.325 25 | 0.126 14 | 0.492 79 | 0.367 43 | 0.296 13 | 0.308 22 | 1.278 09    |
| 0.01      | 100   | -0.326 47 | 0.120 42 | 0.490 75 | 0.355 18 | 0.301 54 | 0.312 57 | 2.446 89    |
|           | 200   | -0.320 52 | 0.122 37 | 0.488 24 | 0.364 48 | 0.296 25 | 0.308 27 | 2.185 28    |
|           | 500   | -0.326 82 | 0.121 44 | 0.492 63 | 0.368 13 | 0.295 26 | 0.312 13 | 1.551 46    |
|           | 1 000 | -0.326 35 | 0.124 82 | 0.495 09 | 0.367 72 | 0.296 58 | 0.312 42 | 1.115 48    |
|           | 2 000 | -0.326 99 | 0.125 72 | 0.496 22 | 0.370 22 | 0.297 08 | 0.312 48 | 0.897 85    |
|           | 5 000 | -0.327 09 | 0.127 14 | 0.497 69 | 0.371 27 | 0.298 38 | 0.310 93 | 0.622 36    |
| 真值        |       | -0.330 00 | 0.130 00 | 0.500 00 | 0.370 00 | 0.300 00 | 0.310 00 |             |

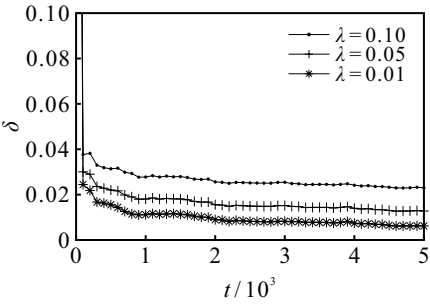


图5 P-MISG 参数估计误差 $\delta$ 随 $t$ 变化的曲线  
( $p=4, \sigma^2=0.10^2$ )

在带惩罚项准则函数中, $\lambda=0$ 时惩罚项消失,此时,带惩罚项多新息随机梯度算法退化为多新息随机梯度算法. 取新息长度 $p=4$ ,噪声方差为 $\sigma^2=0.10^2$ ,正则化因子分别为 $\lambda=0$ 和 $\lambda=0.01$ 的P-MISG算法进行辨识,结果如表5所示.

由表4和图5可见, $\lambda=0.01$ 时的参数估计精度高于 $\lambda=0.10$ 和 $\lambda=0.05$ 时,表明较小的正则化因子 $\lambda$ 可使得步长的变化更加平缓,提高算法的精度. 由表5可见,相较于多新息随机梯度算法,P-MISG算法

表 5 P-MISG 和 MISG 参数估计及其误差 ( $\sigma^2=0.10^2, p=4$ )

| $\lambda$ | $k$   | $a_1$     | $a_2$    | $a_3$    | $b_1$    | $b_2$    | $b_3$    | $\delta/\%$ |
|-----------|-------|-----------|----------|----------|----------|----------|----------|-------------|
| 0         | 100   | -0.326 81 | 0.120 57 | 0.491 56 | 0.356 21 | 0.301 93 | 0.313 39 | 2.318 46    |
|           | 200   | -0.320 96 | 0.122 65 | 0.489 25 | 0.365 51 | 0.296 81 | 0.309 06 | 2.010 56    |
|           | 500   | -0.327 29 | 0.121 75 | 0.493 76 | 0.369 15 | 0.295 84 | 0.312 93 | 1.420 30    |
|           | 1 000 | -0.326 81 | 0.125 12 | 0.496 25 | 0.368 72 | 0.297 15 | 0.313 16 | 0.984 97    |
|           | 2 000 | -0.327 47 | 0.126 03 | 0.497 47 | 0.371 22 | 0.297 66 | 0.313 20 | 0.809 31    |
|           | 5 000 | -0.327 55 | 0.127 42 | 0.498 93 | 0.372 24 | 0.298 96 | 0.311 64 | 0.568 92    |
| 0.01      | 100   | -0.326 47 | 0.120 42 | 0.490 75 | 0.355 18 | 0.301 54 | 0.312 57 | 2.446 89    |
|           | 200   | -0.320 52 | 0.122 37 | 0.488 24 | 0.364 48 | 0.296 25 | 0.308 27 | 2.185 28    |
|           | 500   | -0.326 82 | 0.121 44 | 0.492 63 | 0.368 13 | 0.295 26 | 0.312 13 | 1.551 46    |
|           | 1 000 | -0.326 35 | 0.124 82 | 0.495 09 | 0.367 72 | 0.296 58 | 0.312 42 | 1.115 48    |
|           | 2 000 | -0.326 99 | 0.125 72 | 0.496 22 | 0.370 22 | 0.297 08 | 0.312 48 | 0.897 85    |
|           | 5 000 | -0.327 09 | 0.127 14 | 0.497 69 | 0.371 27 | 0.298 38 | 0.310 93 | 0.622 36    |
| 真值        |       | -0.330 00 | 0.130 00 | 0.500 00 | 0.370 00 | 0.300 00 | 0.310 00 |             |

在保证精度的前提下,可解决  $u(t)$  不满足持续激励条件时的数据乘积矩阵奇异带来的矩阵不可逆问题,避免了病态矩阵无法求逆.

5) 模型验证.

使用数据长度从  $L=L_e+1=5\,001$  到  $L=L_e+L_r=6\,000$  的  $L_r=1\,000$  组输入输出数据进行模型验证. 利用噪声方差  $\sigma^2=0.01, \lambda=0.01, t=5\,000$  时 P-MISG 算法得到的参数估计作为最终的模型,其对应的模型输出为

$$\begin{aligned}\hat{y}(t) = & \varphi^T(t)\hat{\theta}(L) = \\ & -0.327\,09y(t-1) - 0.127\,14y(t-2) + \\ & 0.497\,69y(t-3) + 0.371\,27u(t-1) + \\ & 0.298\,38u(t-2) + 0.310\,93u(t-3).\end{aligned}$$

通过计算估计模型输出误差均方根(MSE)

$$\text{MSE}(\hat{y}(t)) = \left[ \frac{1}{L_r} \sum_{k=L+1}^{L+L_r} [\hat{y}(t) - y(t)]^2 \right]^{\frac{1}{2}},$$

评价估计模型质量,并得到  $\text{MSE}(\hat{y}(t))=0.102\,8$ . 图 6 给出了 MISG 算法系统输出和模型输出及其误差随  $t$  变化情况曲线.

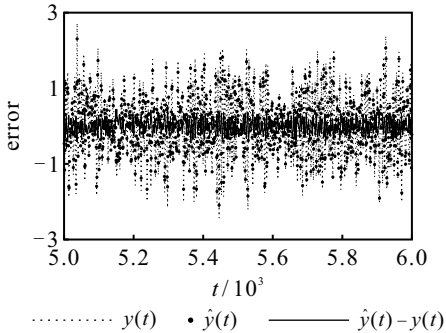


图 6 系统输出和模型输出及其误差随  $t$  变化情况  
( $p=4, \sigma^2=0.10^2$ )

由图 6 可知,模型预测的输出很接近系统的实际输出,表明模型输出能够跟踪系统输出,且两者间的误差与噪声的标准差  $\sigma=0.10$  十分接近,表明估计模型能够很好地捕捉系统动态.

6 结 论

本文以带惩罚项准则函数为基础,利用负梯度搜索原理研究了白噪声干扰下的受控自回归系统的辨识问题. 通过极小化带惩罚项的准则函数得到系统参数估计的递推关系,提出了带惩罚项的投影梯度辨识算法和带惩罚项的随机梯度算法,并利用多新息辨识思想,进一步推导了基于带惩罚项准则函数的多新息随机梯度辨识方法. 仿真结果验证了所提出算法的性能. 所提出参数辨识算法可结合其他估计算法,研究线性 and 非线性系统的新估计方法,并可推广到有噪声干扰的随机系统中.

参考文献(References)

[1] 丁锋. 系统辨识新论[M]. 北京: 科学出版社, 2013: 14-31.  
(Ding F. System identification—New theory and methods[M]. Beijing: Science Press, 2013: 14-31.)  
[2] 丁锋. 现代控制理论[M]. 北京: 清华大学出版社, 2018: 345-349.  
(Ding F. Modern control theory[M]. Beijing: Tsinghua University Press, 2018: 345-349.)  
[3] 张霄, 丁锋. 双线性状态空间系统的状态观测器设计[J]. 控制与决策, 2023, 38(1): 274-280.  
(Zhang X, Ding F. State observers for bilinear state-space systems[J]. Control and Decision, 2023, 38(1): 274-280.)  
[4] 王立夫, 钟昊男, 郭戈. 基于拥堵系数的道路交通网络关键路段辨识[J]. 控制与决策, 2023, 38(3): 843-849.

- (Wang L F, Zhong H N, Guo G. Identification of key road sections of road traffic network based on congestion coefficient[J]. Control and Decision, 2023, 38(3): 843-849.)
- [5] 余明裕, 李仲昆, 王泊桦. 基于固定时间扰动观测器的水面无人艇精确编队控制[J]. 控制与决策, 2023, 38(2): 379-387.  
(Yu M Y, Li Z K, Wang B H. Fixed-time disturbance observer-based accurate formation control of unmanned surface vehicles[J]. Control and Decision, 2023, 38(2): 379-387.)
- [6] 付华, 刘昊. 多策略融合的改进麻雀搜索算法及其应用[J]. 控制与决策, 2022, 37(1): 87-96.  
(Fu H, Liu H. Improved sparrow search algorithm with multi-strategy integration and its application[J]. Control and Decision, 2022, 37(1): 87-96.)
- [7] Fan Y M, Liu X M. Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity[J]. International Journal of Robust and Nonlinear Control, 2020, 30(14): 5492-5509.
- [8] Ji Y, Kang Z, Zhang C. Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering[J]. International Journal of Control, Automation and Systems, 2021, 19(8): 2706-2715.
- [9] Xu L, Song G L. A recursive parameter estimation algorithm for modeling signals with multi-frequencies[J]. Circuits, Systems, and Signal Processing, 2020, 39(8): 4198-4224.
- [10] 李峰, 罗印升, 李博, 等. 基于组合式信号源的 Hammerstein-Wiener 模型辨识方法[J]. 控制与决策, 2022, 37(11): 2959-2967.  
(Li F, Luo Y S, Li B, et al. Identification method of the Hammerstein-Wiener model based on combined signal sources[J]. Control and Decision, 2022, 37(11): 2959-2967.)
- [11] 丁锋, 郑嘉芸, 张霄, 等. 基于数据滤波的随机梯度辨识方法[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2022.1867.  
(Ding F, Zheng J Y, Zhang X, et al. Filtering-based stochastic gradient identification methods[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2022.1867.)
- [12] Dong R Q, Zhang Y, Wu A G. Weighted hierarchical stochastic gradient identification algorithms for ARX models[J]. International Journal of Systems Science, 2020, 52(2): 363-373.
- [13] 陈光武, 李少远, 李文元, 等. 基于递推最小二乘与互补滤波的姿态估计[J]. 控制理论与应用, 2019, 36(7): 1096-1103.  
(Chen G W, Li S Y, Li W Y, et al. Attitude estimation based on recursive least square and complementary filtering[J]. Control Theory & Applications, 2019, 36(7): 1096-1103.)
- [14] 丁锋. 系统辨识——多新息辨识理论与方法[M]. 北京: 科学出版社, 2016: 59-68.  
(Ding F. System identification—Multi-innovation identification theory and methods[M]. Beijing: Science Press, 2016: 59-68.)
- [15] Xu L. Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window[J]. Circuits, Systems, and Signal Processing, 2022, 41(2): 805-830.
- [16] Fan Y M, Liu X M. Auxiliary model-based multi-innovation recursive identification algorithms for an input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity[J]. International Journal of Adaptive Control and Signal Processing, 2022, 36(3): 521-540.
- [17] Pan J, Jiang X, Wan X K, et al. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems[J]. International Journal of Control, Automation and Systems, 2017, 15(3): 1189-1197.
- [18] 徐玲. 基于移动数据窗的传递函数多新息随机梯度辨识方法[J]. 控制与决策, 2017, 32(6): 1091-1096.  
(Xu L. Moving data window based multi-innovation identification stochastic gradient parameter identification method for transfer functions[J]. Control and Decision, 2017, 32(6): 1091-1096.)

## 作者简介

孙焕琪(1995—), 女, 博士生, 从事系统辨识、工业过程检测等研究, E-mail: hqsun11@163.com;

熊伟丽(1978—), 女, 教授, 博士生导师, 从事智能软测量技术及应用、基于数据的工业过程检测、故障诊断等研究, E-mail: weili\_xiong@jiangnan.edu.cn;

丁锋(1963—), 男, 教授, 博士生导师, 从事系统辨识、过程系统、自适应控制等研究, E-mail: fding@jiangnan.edu.cn.