

## 具有量测误差的二值输出FIR系统参数迭代辨识

郭健<sup>1</sup>, 薛文超<sup>1†</sup>, 王婷<sup>2</sup>, 张纪峰<sup>1</sup>

(1. 中国科学院数学与系统科学研究院 系统科学研究所, 北京 100080;

2. 北京科技大学 智能科学与技术学院, 北京 100080)

**摘要:** 本文考虑了具有量测误差的二值输出有限脉冲响应 (FIR, finite impulse response) 的系统参数辨识问题, 其中量测误差使得二值型量测值有一定概率得到相反的取值. 首先, 对所考虑的 FIR 系统, 给出了参数的极大似然估计 (MLE), 在噪声满足一定正则条件下证明了 MLE 的强收敛性和渐近正态性. 此外, 通过分析似然函数的性质, 给出了一种基于 EM (Expectation-maximization) 算法的 MLE 迭代求解算法. 为适应更一般的量测误差情形, 给出了带投影的迭代求解算法, 保证了迭代估计序列的有界性. 进一步, 在给定数量的观测下, 给出了似然函数具有唯一最大值点的必要和充分条件. 在持续激励输入条件下, 证明了迭代估计误差以指数速度收敛到零. 最后, 基于典型系统的数值模拟结果显示了所提出算法的有效性.

**关键词:** 二值观测; 极大似然估计; 系统辨识; 强收敛性; 渐近正态性; 指数收敛速度

**引用格式:** 郭健, 薛文超, 王婷, 张纪峰. 控制理论与应用, xxxx, xx(x): xxx – xxx

DOI: 10.7641/CTA.201x.xxxxx

## Iterative Parameter Identification of Binary Output FIR Systems with Measurement Errors

GUO Jian<sup>1</sup>, XUE Wenchao<sup>1†</sup>, WANG Ting<sup>2</sup>, ZHANG Ji-Feng<sup>1</sup>

(1. Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, China;

2. School of Intelligence Science and Technology, University of Science and Technology Beijing 100080, China )

**Abstract:** In this paper, we consider the problem of parameter identification for a binary output finite impulse response (FIR) system with measurement error, where the measurement error makes the binary measurement values take opposite values with a certain probability. First, for the considered FIR system, the maximum likelihood estimation (MLE) of the parameters is given, and the strong convergence and asymptotic normality of the MLE are proved under certain regularity conditions of the noise. In addition, to solve MLE, by analyzing the properties of the likelihood function, an iterative algorithm based on the Expectation-Maximization algorithm is given. In order to adapt to more general measurement error cases, an iterative solution algorithm with projection is given to ensure the boundedness of the iterative estimation sequence. Further, a necessary and sufficient condition for the likelihood function to have a unique maximum point is given. It is shown that the iterative estimation error converges to zero at an exponential rate under persistently excitation input conditions. Finally, numerical simulation results based on a typical system show the effectiveness of the proposed algorithm.

**Key words:** Binary-valued observation; maximum likelihood estimate; system identification; strong convergence; asymptotic normality; exponential rate

**Citation:** Guo Jian, Xue Wen, Wang Ting, Zhang Ji-Feng. Iterative Parameter Identification of Binary Output FIR Systems with Measurement Errors. *Control Theory & Applications*, xxxx, xx(x): xxx – xxx

### 1 引言

随着信息技术的发展, 集值输出系统 [20] 越来越多地出现在各类实际场景中, 其中集值输出指系统的

输出不能被准确测量, 而可以测量得到的信息仅说明输出是否属于某个集合. 这类特殊的集值输出系统在工业生产 [1]、生物制药技术 [2]和信息产业 [13]等

收稿日期: xxxx-xx-xx; 录用日期: xxxx-xx-xx.

†通信作者. E-mail: wenchaoxue@amss.ac.cn.

本文责任编辑:

国家重点研发计划(2018YFA0703800), 国家自然科学基金项目(T2293770, 12226305), 中国科学院青年创新促进会资助.

Supported by National Key R&D Program of China under Grant 2018YFA0703800, National Natural Science Foundation of China (T2293770 and 12226305) and Youth Innovation Promotion Association CAS.

谨以此文献给著名控制学者、控制界老前辈秦化淑研究员九十华诞.

领域的实际场景中普遍存在并具有重要的研究意义,因此受到了广泛的关注.一类具有代表性的集值输出系统是二值输出系统,其输出数据通常由二进制的传感器获得.例如,在异步传输模式(Asynchronous transmission mode)网络等通信系统中,测量比特率、队列长度等流量信息的传感器均为二值形式 [20];在汽车系统中,废气氧开关传感器等也为二值形式 [11].在这些实际场景中,由于只有有限的信息可用,集值传感器引入了实质性的困难,特别是在二值观测情形下,系统的辨识问题更加难以解决.

近年来,关于集值输出系统辨识的研究已经产生了大量的文献([4, 7-8, 12, 18]).参考文献 [20] 提出了一种经验测量方法来研究具有二值信息的线性系统的识别误差.文献 [16] 提出了一种用于有限脉冲响应(FIR, finite impulse response)系统的非截断经验测量方法,并证明了其在Cramér-Rao下界意义上的渐近有效性.文献 [9] 将期望值最大化(EM, expectation-maximization)算法引入到 FIR 系统中,并利用该方法辨识量化输出的系统.该文章中提供了一些仿真结果验证了该方法的有效性.文献 [21] 提出了一种针对 FIR 系统的EM型算法,该算法能在一定数量观测的基础上,迭代求解得到极大似然估计.文中进一步证明该算法在迭代求解出极大似然估计意义上具有指数级收敛速度.在固定阈值的量化器下, [10] 提出了一种用于 FIR 模型的递归投影算法,该算法被证明是可收敛的.文献 [15] 将该算法扩展到矩阵输入和矢量输出的情况,并获得了更快的收敛率,可以达到  $O(1/k)$ .文献 [17] 进一步引入了一个统一的方法,该方法可以同时解决随机框架和确定性框架下的二值输出系统的参数辨识问题.其不同之处在于构造了一个随时间变化的阈值量化器,对于无噪声情况,证明了对时间步长  $k$  的平方收敛率接近于  $O(1/k^2)$ ;对于具有独立同分布随机噪声的情况,估计误差的均方收敛率为  $O(1/k)$ .此外,文献 [19] 证明了Cramér-Rao 下界可以在适当的权重系数下渐近地达到,这表明了该文章中的算法是渐近有效的.

上述的工作主要都是在准确的二值观测的基础上进行的.然而,在一些实际问题中,量测通常会由传感器野值、数据包丢失、通信错误等问题引起误差,这使得观测值可能会以一定的概率得到相反的情况.例如,在一个通信系统中,0-1比特信息可能会被错误地传输或接收.此外,集值辨识算法的一个重要应用是目标分类的问题 [14].然而,训练数据的标签在获取时有一定的概率被错误分类,这将影响集值辨识算法的应用,导致最终输出一个错误的分类器.因此考虑到现实存在的误差现象,有两个问题需要被进一步研究:第一个问题是当系统二值输出存在量测误差时,如何设计算法辨识参数?另一个问题是设计的辨识算

法是否具有一致性.本文考虑含有一定量测误差的二值输出系统,在文献 [21] 的启发下,将二值输出有一定概率相反的系统辨识问题转化为相应的极大似然(ML, maximum likelihood)估计问题,给出了极大似然估计(MLE, maximum likelihood estimation),证明了MLE的强收敛性和渐近正态性,同时给出了MLE的迭代求解算法,证明了该算法的指数收敛性.具体贡献包括:

i) 对所考虑的带量测误差的二值输出 FIR 系统,给出了参数的MLE并证明了一致性.基于ML准则,给出了MLE,在噪声满足一定正则条件下证明了MLE的强收敛性,即随着观测数量的增加,MLE存在且几乎处处收敛到参数真值.同时,给出了MLE的渐近正态性;

ii) 在给定数量的观测下,提出了求解MLE的迭代算法.通过分析似然函数的性质,给出了一种基于EM算法的MLE迭代求解算法.进一步,为适应更一般的量测误差情形,给出了带投影的迭代求解算法,保证了迭代估计序列的有界性.

iii) 给出了MLE存在唯一性的充要条件并证明了迭代估计算法对于MLE的一致性.在给定数量的观测下,给出了似然函数具有唯一最大值点的必要和充分条件以及相应条件的验证准则.其次,在持续激励输入条件下,证明了迭代估计误差以指数速度收敛到零.最后,基于典型系统的数值模拟结果显示了所提算法的有效性.

本文的其余部分组织如下:第2节介绍了辨识问题及其相应的ML准则,给出了关于噪声分布函数的正则性条件,并在该正则性条件下证明了MLE的强收敛性和渐近正态性.第3节考虑了在给定数量的观测下,MLE的求解问题.基于EM算法提出了一种迭代求解算法;具体分析了似然函数的性质,给出了似然函数最大值点存在和唯一的必要和充分条件;证明了迭代算法在一定条件下的收敛性,其中收敛速度可以达到指数速率.第4节通过数值模拟验证了在带有量测误差的二值输出模型下,所提出的MLE迭代求解算法的收敛性;并与不考虑量测误差的方法进行了比较.第5节总结了整个论文,讨论了相关的未来工作.

## 2 参数辨识问题描述及极大似然估计分析

在这一节中,我们将首先介绍具有二值观测误差的辨识问题,随后介绍解决该问题的方法.

### 2.1 问题描述

考虑以下关于随机 FIR 系统的参数辨识问题:

$$\begin{cases} y_k = \phi_k^T \theta_0 + e_k, \\ \tilde{s}_k = I_{[y_k \leq C]}, \quad 1 \leq k \leq N, \end{cases} \quad (1)$$

其中  $\theta_0$  是未知但时不变的  $p$ -维参数向量,其属于某个

参数空间  $\Theta \subseteq \mathbb{R}^p$  且维数  $p$  已知;  $\phi_k \in \mathbb{R}^p$  是包含当前和过去输入信息的回归量;  $y_k \in \mathbb{R}$  和  $e_k \in \mathbb{R}$  分别是系统虚拟输出和系统噪声;  $\tilde{s}_k \in \{0, 1\}$  是系统的二值观测输出, 具体是通过比较虚拟系统输出和给定的传感器阈值  $C \in \mathbb{R}$  给出, 若超过阈值取 0, 反之取 1;  $I$  是示性函数;  $N$  代表当前获取到的数据长度.

在实际中, 由于传感器野值、数据包丢失、通信错误等导致了量测误差, 系统真实的二值观测可能以一定的概率被观测到相反的取值. 因此, 在这种情形下, 该系统真实的二值输出  $\tilde{s}_k$  并不能准确地获得, 反之, 有一定的概率获得与真实二值输出相反的结果, 这将极大的影响后续系统辨识步骤的实施, 若不考虑这种情形, 甚至会得到与真实模型相差甚远的辨识结果. 这种带有量测误差的二值观测可以在数学上被表示为如下的结果:

$$s_k = \begin{cases} \tilde{s}_k, & \text{依概率 } p_k, \\ 1 - \tilde{s}_k, & \text{依概率 } 1 - p_k, \end{cases} \quad (2)$$

其中  $p_k \in (0, 1)$ , 其值用以表征获得正确二值输出观测的概率.  $p_k$  的取值或模型取决于导致错误的方式. 例如若为通信错误,  $p_k$  可取常值, 若由 FIR 过程导致, 则  $p_k$  可以是一个  $\theta_0, C$ , 和  $\phi_k$  的函数.

本文的目标包括两个, 第一是在二值观测带有量测误差的模型下, 利用输入数据集  $J_N = \{\phi_1, \phi_2, \dots, \phi_N\}$  和有误差的二值量测数据集  $\mathcal{O}_N = \{s_1, s_2, \dots, s_N\}$ , 给出基于 ML 准则的参数估计, 即 MLE, 并证明 MLE 的强收敛性和渐近正态性; 第二是在给定数目的输入数据集  $J_N = \{\phi_1, \phi_2, \dots, \phi_N\}$  和有误差的二值量测数据集  $\mathcal{O}_N = \{s_1, s_2, \dots, s_N\}$  下, 给出 MLE 的迭代求解算法和未知参数  $\theta_0$  的估计和分析迭代算法的收敛性质.

在给出我们的主要结果前, 首先介绍 ML 准则, 并给出模型 (1)-(2) 下的似然函数与 MLE.

## 2.2 极大似然准则

在统计学中, MLE 是一种在给定若干观测值的情况下估计假设的概率分布的参数的方法, 目的是给出表征当前观测到的数据可能性的目标函数, 即似然函数, 并通过最大化似然函数来实现对观测数据总体分布参数的估计. 该参数估计结果对应的统计模型能使实际观测到的数据最有可能出现. ML 的逻辑既直观又灵活, 因此该方法已成为统计推断的主要手段. 似然函数被最大化的点称为 MLE. 现在我们考虑系统 (1) 和 (2), 给出带量测误差的二值输出模型下的极大似然函数.

对于任意  $1 \leq k \leq N$ , 给定输入数据  $\phi_k$  及参数  $\theta$ , 由全概率公式, 观测到二值输出为  $s_k = 1$  和  $s_k = 0$

的概率分别如下:

$$\begin{aligned} & P\{s_k = 1 \mid \phi_k, \theta\} \\ &= P\{s_k = \tilde{s}_k\}P\{s_k = 1 \mid s_k = \tilde{s}_k, \phi_k, \theta\} \\ & \quad + P\{s_k = 1 - \tilde{s}_k\}P\{s_k = 1 \mid s_k = 1 - \tilde{s}_k, \phi_k, \theta\} \\ &= P\{s_k = \tilde{s}_k\}P\{y_k \leq C \mid \phi_k, \theta\} \\ & \quad + P\{s_k = 1 - \tilde{s}_k\}P\{y_k > C \mid \phi_k, \theta\} \\ &= p_k F(C - \phi_k^T \theta) \\ & \quad + (1 - p_k)[1 - F(C - \phi_k^T \theta)] \\ &= [2p_k - 1]F(C - \phi_k^T \theta) + 1 - p_k, \end{aligned}$$

及

$$\begin{aligned} & P\{s_k = 0 \mid \phi_k, \theta\} = 1 - P\{s_k = 1 \mid \phi_k, \theta\} \\ &= [1 - 2p_k]F(C - \phi_k^T \theta) + p_k, \end{aligned}$$

其中  $F(x)$  是噪声  $e_k$  的分布函数. 令

$$G(C - \phi_k^T \theta) = [2p_k - 1]F(C - \phi_k^T \theta) + 1 - p_k, \quad (3)$$

使用上述条件概率, 我们可以在已知输入数据  $J_N$  的情况下得到以下关于参数  $\theta$  的似然函数

$$\begin{aligned} L_N(\theta) &= P\{\mathcal{O}_N \mid J_N, \theta\} \\ &= \prod_{k=1}^N P\{s_k \mid \phi_k, \theta\} \\ &= \prod_{\{k:s_k=1\}} P\{s_k = 1 \mid \phi_k, \theta\} \\ & \quad \cdot \prod_{\{k:s_k=0\}} P\{s_k = 0 \mid \phi_k, \theta\} \\ &= \prod_{\{k:s_k=1\}} G(C - \phi_k^T \theta) \\ & \quad \cdot \prod_{\{k:s_k=0\}} (1 - G(C - \phi_k^T \theta)). \end{aligned}$$

由于  $\log x$  是一个严格单调增的函数, 并且不失一般性,  $L_N(\theta)$  可以假定为正数, 最大化似然函数  $L_N(\theta)$  等价于最大化下述对数似然函数:

$$l_N(\theta) = \log L_N(\theta) \quad (4)$$

$$\begin{aligned} &= \sum_{k=1}^N [\log[G(C - \phi_k^T \theta)]I_{[s_k=1]} \\ & \quad + \log[1 - G(C - \phi_k^T \theta)]I_{[s_k=0]}]. \end{aligned} \quad (5)$$

通常情况下, 直接研究对数似然的性质会更加方便, 此时相对应的极大似然估计就是对数似然函数的最大值点:

$$\hat{\theta}_N = \arg \max_{\theta} l_N(\theta). \quad (6)$$

直观地说, 上式得到的 MLE 选择了使观察结果最可能出现的模型对应的参数值.

### 2.3 极大似然估计的强收敛性与渐近正态性

本节我们介绍在带有量测误差的二值输出模型(1)-(2)下,由式(6)给出的 MLE 的强收敛性与渐近正态性.注意到式(4)中的对数似然函数,则对数似然函数的梯度可以计算为:

$$s_N(\theta) = \nabla l_N(\theta) = \sum_{k=1}^N \left[ \left( \frac{-g(C - \phi_k^T \theta)}{G(C - \phi_k^T \theta)} I_{[s_k=1]} + \frac{g(C - \phi_k^T \theta)}{1 - G(C - \phi_k^T \theta)} I_{[s_k=0]} \right) \phi_k \right], \quad (7)$$

其中  $g(\cdot)$  是函数  $G(\cdot)$  的导数,即  $g(\cdot) = G'(\cdot)$ . 进一步,对数似然函数(4)的 Hessian 矩阵可以计算为:

$$\nabla^2 l_N(\theta) = - \sum_{k=1}^N \left[ \left( \frac{-g'(x)G(x) + g^2(x)}{G^2(x)} \cdot s_k + \frac{g'(x)(1-G(x)) + g^2(x)}{(1-G(x))^2} \cdot (1 - s_k) \right) \Big|_{x=C - \phi_k^T \theta} \cdot \phi_k \phi_k^T \right].$$

由此可以得到 Fisher 信息阵为:

$$F_N(\theta) = -E \nabla^2 l_N(\theta) = \sum_{k=1}^N \left[ \left( \frac{g^2(C - \phi_k^T \theta)}{G(C - \phi_k^T \theta)} + \frac{g^2(C - \phi_k^T \theta)}{1 - G(C - \phi_k^T \theta)} \right) \phi_k \phi_k^T \right]. \quad (8)$$

为表述简洁起见,后续将把在真实参数  $\theta_0$  处的函数值  $s_N(\theta_0)$  和  $F_N(\theta_0)$  分别简记为  $s_N$  与  $F_N$ ; 对于矩阵  $A$ ,  $A^T$  与  $A^{T/2}$  分别表示  $A$  和  $A^{1/2}$  的转置.同时,为分析收敛性,需给出如下关于噪声  $e_k$  和参数空间  $\Theta$  的正则性假设:对于任意  $\delta > 0$ , 定义

$$N_n(\delta) = \left\{ \theta : \left\| F_n^{T/2}(\theta - \theta_0) \right\| \leq \delta \right\}, \quad n = 1, 2, \dots$$

**假设 1** (a) 参数空间  $\Theta$  在  $\mathcal{R}^p$  中紧且凸;

(b) 信息矩阵  $F_N$  满足  $\lambda_{\min}(F_N) \rightarrow \infty$ ;

(c) 对于任意  $\delta > 0$ , 成立

$$\max_{\theta \in N_n(\delta)} \left\| \mathbf{V}_n(\theta) - \mathbf{I} \right\| \rightarrow 0,$$

其中  $\mathbf{V}_n(\theta) = F_n^{-1/2} F_n(\theta) F_n^{-T/2}$  是标准化的信息矩阵.

(d) 在参数空间  $\Theta$  中,存在  $\theta_0$  的邻域  $\Omega$  及常数  $c > 0$ ,  $\delta > 0$ ,  $n_1$ , 使得

$$\lambda_{\min}(F_n(\theta)) \geq c (\lambda_{\max}(F_n))^{1/2+\delta}, \quad \forall \theta \in \Omega, \forall n \geq n_1,$$

**注 1** 假设(a)是为了 MLE 的唯一性(如果 MLE 存在),假设(b), (d)保证了 MLE 的强收敛性,假设(b), (c)保证了 MLE 的渐近正态性.此外,假设(b)与经典系统辨识理论中的持续激励条件:  $\lambda_{\min} \left( \sum_{k=1}^N \phi_k \phi_k^T \right) \rightarrow \infty$  相关,其保证了  $F_N$  的正定性.

现在给出 MLE 的渐近结果.

**定理 1** 考虑系统(1)-(2)和式(6)中的 MLE,

在假设 1 中的条件 (a), (b) 和 (d) 下,存在随机变量序列  $\{\hat{\theta}_N\}$  及随机整数  $n_2$ , 使得

$$P(s_N(\hat{\theta}_N) = 0, \forall N \geq n_2) = 1,$$

并且  $\hat{\theta}_N \rightarrow \theta_0$ , a.s.

**证** 对给定的常数  $\varepsilon > 0$ , 定义邻域

$$K_\varepsilon(\theta_0) = \{\theta : \|\theta - \theta_0\| \leq \varepsilon\}.$$

令  $\varepsilon$  充分小,使得  $K_\varepsilon(\theta_0)$  被包含在条件(d)中的邻域  $\Omega$  中.为证明定理中的结果,只需要证明存在随机整数  $n_2$ ,使得下述事件成立的概率为 1:

$$l_N(\theta) - l_N(\theta_0) < 0, \quad \forall \|\theta - \theta_0\| = \varepsilon \quad (9)$$

以及  $\forall N \geq n_2$ .

下面证明这个结论.对对数似然函数(4)做泰勒展开可以得到:

$$l_N(\theta) - l_N(\theta_0) = (\theta - \theta_0)^T s_N - \frac{1}{2}(\theta - \theta_0)^T F_N(\tilde{\theta})(\theta - \theta_0), \quad (10)$$

其中  $\tilde{\theta}$  在  $\theta$  和  $\theta_0$  之间.记  $\alpha = (\theta - \theta_0)/\varepsilon$ , 由式(10)可以得到事件(9)等价于

$$\alpha^T s_N < \frac{\varepsilon}{2} \alpha^T F_N(\tilde{\theta}) \alpha, \quad \forall \alpha^T \alpha = 1, \quad N \geq n_2.$$

在上式两边同时除以  $\lambda_{\max}(F_N)^{1/2+\delta}$ , 则有

$$\frac{\alpha^T s_N}{\lambda_{\max}(F_N)^{1/2+\delta}} < \frac{\varepsilon}{2} \frac{\alpha^T F_N(\tilde{\theta}) \alpha}{\lambda_{\max}(F_N)^{1/2+\delta}}. \quad (11)$$

一方面,由于  $s_N$  任意元素的方差都不超过  $\lambda_{\max}(F_N)$ , 则由 Kolmogorov 强大数定律 [6] 可以得到

$$\frac{s_N}{\lambda_{\max}(F_N)^{1/2+\delta}} \rightarrow 0, \quad \text{a.s.}$$

注意到  $\alpha^T \alpha = 1$ , 则由 Cauchy-Schwarz 不等式得到

$$\frac{\alpha^T s_N}{\lambda_{\max}(F_N)^{1/2+\delta}} \rightarrow 0, \quad \text{a.s.}, \quad \forall \alpha \text{ 满足 } \alpha^T \alpha = 1.$$

另一方面,由条件(d)知,  $\forall N \geq n_1$ , 有

$$\frac{\varepsilon}{2} \frac{\alpha^T F_N(\tilde{\theta}) \alpha}{\lambda_{\max}(F_N)^{1/2+\delta}} \geq \frac{\varepsilon c \lambda_{\max}(F_N)^{1/2+\delta}}{2 \lambda_{\max}(F_N)^{1/2+\delta}} = \frac{c\varepsilon}{2}.$$

因此事件(11)可以以概率 1 成立,定理得证.

下面给出极大似然估计的渐近正态性.

**定理 2** 考虑系统(1)-(2)和式(6)中的 MLE, 在假设 1 下, MLE 是渐近正态的, 即

$$F_N^{T/2}(\hat{\theta}_N - \theta_0) \rightarrow_d N(0, \mathbf{I})$$

**证** 注意到  $s_N = s_N(\theta_0)$ , 由中值定理有

$$s_N = \left[ \int_0^1 F_N(\theta_0 + t(\hat{\theta}_N - \theta_0)) dt \right] (\hat{\theta}_N - \theta_0).$$

上式两边左乘  $F_N^{-1/2}$  并注意到

$$\mathbf{V}_N(\theta) = F_N^{-1/2} F_N(\theta) F_N^{-T/2},$$

则

$$F_N^{-1/2} s_N = \left[ \int_0^1 \mathbf{V}_N(\theta_0 + t(\hat{\theta}_N - \theta_0)) dt \right] F_N^{T/2} (\hat{\theta}_N - \theta_0). \quad (12)$$

由条件(c)及积分的性质, 对任意的  $\varepsilon > 0$ , 总存在  $\delta > 0$  以及充分大的  $N$  使得  $\hat{\theta}_N \in N_N(\delta)$ , 并且成立

$$\left\| \int_0^1 \mathbf{V}_N(\theta_0 + t(\hat{\theta}_N - \theta_0)) dt - \mathbf{I} \right\| \leq \int_0^1 \varepsilon dt = \varepsilon.$$

由定理 1, 这样的  $\delta$  总能取到, 因此我们有

$$\int_0^1 \mathbf{V}_N(\theta_0 + t(\hat{\theta}_N - \theta_0)) dt \rightarrow_p \mathbf{I}. \quad (13)$$

从式 (12) 和式 (13) 可以看出, 要想证明定理, 只需要证明

$$F_N^{-1/2} s_N \rightarrow_d N(0, \mathbf{I}).$$

对任意满足  $\alpha^T \alpha = 1$  的  $\alpha$ , 考虑  $\alpha^T F_N^{-1/2} s_N$  的矩母函数  $E \exp(t \alpha^T F_N^{-1/2} s_N)$ . 给定  $t$  和  $\alpha$ , 定义序列  $\theta_N = \theta_0 + t F_N^{-T/2} \alpha$ , 则  $\theta_N \in N_N(t)$ . 对对数似然函数做泰勒展开得到

$$l_N(\theta_N) = l_N(\theta_0) + (\theta_N - \theta_0)^T s_N - (\theta_N - \theta_0)^T F_N(\tilde{\theta}_N) (\theta_N - \theta_0) / 2,$$

其中  $\tilde{\theta}_N$  在  $\theta_N$  和  $\theta_0$  之间. 上式两边做自然对数的幂次得到

$$\begin{aligned} & \exp\left(\alpha^T \mathbf{V}_N(\tilde{\theta}_N) \lambda t^2 / 2\right) L_N(\theta_N) \\ &= \exp\left(t \alpha^T F_N^{-1/2} s_N\right) L_N(\theta_0). \end{aligned}$$

两边同时积分得到

$$\begin{aligned} & E_{\theta_N} \exp\left(\alpha^T \mathbf{V}_N(\tilde{\theta}_N) \lambda t^2 / 2\right) \\ &= E \exp\left(t \alpha^T F_N^{-1/2} s_N\right). \end{aligned} \quad (14)$$

注意到条件(c), 则对于任意  $\varepsilon > 0$ , 存在整数  $n_1$ , 使得对于任意  $N > n_1$ , 成立

$$\left| \exp\left(\alpha^T \mathbf{V}_N(\tilde{\theta}_N) \alpha t^2 / 2\right) - \exp(t^2 / 2) \right| \leq \varepsilon. \quad (15)$$

由式 (14) 和 (15),  $\alpha^T F_N^{-1/2} s_N$  的矩母函数  $E \exp(t \alpha^T F_N^{-1/2} s_N)$  收敛到  $\exp(t^2 / 2)$ , 此即是标准正态分布的矩母函数. 因此, 定理得证.

**注 2** 当噪声满足标准正态分布, 并且回归量  $\phi_k$  满足  $\lambda_{\max}\left(\sum_{k=1}^N \phi_k \phi_k^T\right) = O(N)$  及  $\lambda_{\min}\left(\sum_{k=1}^N \phi_k \phi_k^T\right) = O(N)$ , 注意到式 (8), 则信息矩阵  $F_N(\theta)$  在一定的参数空间  $\Theta$  内能够满足  $\lambda_{\max}(F_N(\theta)) = O(N)$  和  $\lambda_{\min}(F_N(\theta)) = O(N)$ .

因此, 假设 1 的条件 (b), (c) 和 (d) 均满足.

**注 3** 对于有限样本, MLE 不具有最优性质. 然而, 与其他估计方法一样, 随着样本量增加到无穷大, MLE 具有许多有吸引力的极限性质. 本节给出了在带有量测误差的二值输出模型下, 参数的 MLE 具有渐近性质需满足的正则性条件, 并证明了强收敛性和渐近正态性的结果, 表明了量测误差下 MLE 仍具有大样本性质, 只要噪声分布与回归量满足一定的条件, 即可利用 MLE 渐近地估计出真实参数.

但是 MLE 需要求解一个优化问题, 因此仍然需要考虑在固定样本量的情况下, 如何设计用于求解 MLE 的迭代算法, 并给出相应的收敛性质. 下一节开始, 我们通过研究极大似然函数的性质, 利用 EM 算法给出求解 MLE 的迭代算法. 由于此时关注的是给定数据集  $J_N$  和  $\mathcal{O}_N$ , 因此为了便于描述, 下文中, 在没有冲突的情况下使用符号  $l(\theta)$  和  $\hat{\theta}$  来简记  $l_N(\theta)$  和  $\hat{\theta}_N$ . 对于迭代算法, 设第  $t$  次迭代的估计为  $\hat{\theta}_{N,t}$ , 同样在没有冲突的情况下将其简记为  $\hat{\theta}_t$ .

### 3 极大似然估计的迭代求解算法

#### 3.1 算法设计

本节我们应用 EM 算法的思想来构造用于求解 (6) 的迭代算法. 在给出具体的算法和理论分析之前, 先给出如下的假设:

**假设 2** 对于任意  $k \leq N$ , 系统噪声  $\varepsilon_N = \{e_1, e_2, \dots, e_N\}$  相互独立, 且服从均值为 0, 方差为 1 的高斯分布.

**假设 3** 矩阵  $A \triangleq \sum_{k=1}^N \phi_k \phi_k^T$  正定.

**注 4** 假设 3 是批处理数据环境下持续激励条件的数学描述, 其是系统辨识研究中的一个常见假设. 假设 2 是关于噪声的假设. 一般情况下, 实际系统的误差分布可能是均值为 0, 方差为  $\sigma^2$  的高斯分布, 此时通过变换

$$y_k^* = y_k / \sigma, \theta^* = \theta / \sigma, e_k^* = e_k / \sigma, C^* = C / \sigma,$$

则新的关于参数  $\theta^*$  和输出  $y_k^*$  的系统, 其噪声  $e_k^*$  满足假设 2.

EM 算法是一种迭代方法, 用于找到模型中参数的 (局部) 最大似然. 给定第  $t$  次迭代时的估计值为  $\hat{\theta}_t$ , EM 算法目标是构造一个满足以下性质的函数:

(i)  $l(\theta | \hat{\theta}_t) \leq l(\theta)$  成立对于任意  $\theta \in \Theta$ ,

(ii)  $l(\hat{\theta}_t | \hat{\theta}_t) = l(\hat{\theta}_t)$ ,

然后计算  $\hat{\theta}_{t+1} = \arg \max_{\theta} l(\theta | \hat{\theta}_t)$ . 因此, 我们有

$$l(\hat{\theta}_{t+1}) \geq \max_{\theta} l(\theta | \hat{\theta}_t) \geq l(\hat{\theta}_t | \hat{\theta}_t) = l(\hat{\theta}_t).$$

如此设计得到的迭代序列确保了似然函数不会被缩减. 尽管 EM 迭代序列确实使观测数据的似然函数值逐步增大, 但通常不能保证该序列收敛到 MLE. 对于

带有量测误差的二值输出系统(1)和(2),在一定条件下,可以构造函数  $l(\theta|\hat{\theta}_t)$ , 并给出迭代求解算法,使得迭代序列收敛于 MLE. 函数  $l(\theta|\hat{\theta}_t)$  被定义为

$$l(\theta|\hat{\theta}_t) = -\frac{1}{2}\theta^T \left( \sum_{k=1}^N \phi_k \phi_k^T \right) \theta + \left[ \left( \sum_{k=1}^N \phi_k \phi_k^T \right) \hat{\theta}_t - \left( \sum_{k=1}^N \phi_k \left[ \frac{g(C-\phi_k^T \hat{\theta}_t) I_{[s_k=1]}}{G(C-\phi_k^T \hat{\theta}_t)} - \frac{g(C-\phi_k^T \hat{\theta}_t) I_{[s_k=0]}}{1-G(C-\phi_k^T \hat{\theta}_t)} \right] \right) \right]^T \theta + l_1(\hat{\theta}_t),$$

其中  $l_1(\hat{\theta}_t)$  是与  $\theta$  无关的部分, 函数  $G(\cdot)$  在(3)中定义以及  $g(\cdot) = G'(\cdot)$ . 为了保证  $\hat{\theta}_t$  的有界性, 我们先介绍下面的投影算子:

**定义 1.** 对一个给定的紧闭集  $\Theta \subseteq \mathbb{R}^p$  及一个正定矩阵  $A$ , 投影算子  $\Pi_{\Theta, A}(\cdot)$  被定义为

$$\begin{aligned} \Pi_{\Theta, A}(x) &= \arg \min_{\theta \in \Theta} \|x - \theta\|_A \\ &= \arg \min_{\theta \in \Theta} (x - \theta)^T A (x - \theta), \quad \forall x \in \mathbb{R}^p. \end{aligned}$$

在假设 2 下, 下述算法提供了利用带有量测误差的二值观测数据来求解 MLE (6)的迭代算法.

**算法 1:** 带有量测误差的二值输出系统的迭代估计算法

**输入与初始化:** 输入数据  $\mathcal{J}_N = \{\phi_1, \phi_2, \dots, \phi_N\}$ , 二值观测数据  $\mathcal{O}_N = \{s_1, s_2, \dots, s_N\}$ , 观测正确概率  $P_N = \{p_1, p_2, \dots, p_N\}$ , 阈值  $C$ , 初始值  $\hat{\theta}_1$ .

**迭代:** 在假设 3 下, 估计  $\hat{\theta}_{t+1}$  通过下式无投影迭代

$$\begin{aligned} \hat{\theta}_{t+1} &= \arg \max_{\theta} l(\theta|\hat{\theta}_t) \\ &= \hat{\theta}_t - \left( \sum_{k=1}^N \phi_k \phi_k^T \right)^{-1} \left( \sum_{k=1}^N \phi_k \cdot g \left( C - \phi_k^T \hat{\theta}_t \right) \left[ \frac{I_{[s_k=1]}}{G(C - \phi_k^T \hat{\theta}_t)} - \frac{I_{[s_k=0]}}{1 - G(C - \phi_k^T \hat{\theta}_t)} \right] \right). \end{aligned} \quad (16)$$

或下述带投影迭代

$$\hat{\theta}_{t+1} = \Pi_{\Theta, A} \{ \arg \max_{\theta} l(\theta|\hat{\theta}_t) \}. \quad (17)$$

对正确观测概率  $p_k$  施加某些条件, 可以证明算法 1 以指数速率收敛于式(6)中的 MLE. 在此之前, 我们先讨论 MLE (6)的存在唯一性.

### 3.2 极大似然估计的存在唯一性

在本节中, 我们通过分析对数似然函数(4)来研究 MLE 的性质. 在实际中, 通信过程和 FIR 过程都会引起量测误差, 从而导致二值观测有一定的概率被观测到相反的取值, 下面我们给出正确观测的概率  $p_k$  的函数形式为:

$$p_k = h(C, \theta_0, \phi_k) \quad (18)$$

其中函数  $h(C, \theta_0, \phi_k)$  的值域为  $(0, 1)$ .

**注 5** 当  $h(C, \theta_0, \phi_k) = p^*$ ,  $p^* \in (0, 1)$  时, 其可以描述通信过程产生错误的情形. 该通信过程与系统独立, 以固定的概率使得二值观测相反.

另外, 受限于传感器的灵敏度, 在 FIR 过程也可能导致错误观测. 注意到在实际中, 正确观测发生的概率应与真实值和阈值之间的误差有关. 函数  $h(\cdot)$  可以被约束为  $h(C, \theta_0, \phi_k) = K(C - \theta_0^T \phi_k)$ , 其可以描述 FIR 过程导致错误的情形. 进一步对  $K(\cdot)$  做如下的假设:

- (i)  $K(x)$  是偶函数, i.e.  $K(x) = K(-x)$ .
- (ii)  $K(x)$  在  $(0, \infty)$  上递增, 并且  $\lim_{x \rightarrow \infty} K(x) = 1$ .

条件(i)表明, 如果真值  $\theta_0^T \phi_k$  与阈值  $C$  的差值的绝对值相同, 则应当具有相同的获得相反观测值的概率. 条件(ii)表明, 如果该差值越大, 则观测错误的概率越小, 这是由于二值传感器在远离零值的地方的测量准确度应更高.

本节将针对这些产生量测误差的情形, 研究相应似然函数的凹性, 进而得到 MLE 的存在性和唯一性的充要条件.

在通信过程导致错误观测的情形下, 正确观测概率取值为  $p_k = p^*$ ,  $0 < p^* < 1$ . 令  $f(x) = F'(x)$  是标准正态分布的概率密度函数. 计算可以得到

$$\begin{aligned} G(x) &= 1 - F(x) - p^* + 2p^* F(x), \\ g(x) &= G'(x) = (2p^* - 1)f(x). \end{aligned} \quad (19)$$

此时, 对数似然函数(4)仅在  $\mathbb{R}^p$  的某有限子集上是凹的. 下面的引理阐述了这个结果.

**引理 1** 考虑系统(1)-(2)以及正确观测概率函数形式(18),  $h(C, \theta_0, \phi_k) = p^*$ , 并且假设 2-3 成立, 则对于任意  $k \leq N$ , 由(4)给出的对数似然函数不是全空间  $\mathbb{R}^p$  上的凹函数. 如果  $\Theta$  及  $p^*$  满足  $p^* \geq \inf_{|x| \leq C+LM} \frac{x F(x) + f(x) - x}{2x F(x) + 2f(x) - x}$ , 则对数似然函数(4)在集合  $\Theta$  上凹, 其中  $L = \sup_{\theta \in \Theta} \|\theta\|$  及  $M = \sup_k \|\phi_k\|$ .

**证** 对于任意  $k \leq N$ , 在假设 2 下计算对数似然函数的梯度向量和 Hessian 矩阵可以得到

$$\begin{aligned} \nabla l_N(\theta) &= \sum_{k=1}^N \left[ \left( \frac{-g(C - \phi_k^T \theta)}{G(C - \phi_k^T \theta)} I_{[s_k=1]} + \frac{g(\phi_k^T \theta - C)}{G(\phi_k^T \theta - C)} I_{[s_k=0]} \right) \phi_k \right]. \\ \nabla^2 l(\theta) &= - \sum_{k=1}^N \left[ \left( \frac{g(x)xG(x) + g^2(x)}{G^2(x)} \Big|_{x=C-\phi_k^T \theta} \cdot I_{[s_k=1]} + \frac{g(x)xG(x) + g^2(x)}{G^2(x)} \Big|_{x=\phi_k^T \theta - C} \cdot I_{[s_k=0]} \right) \phi_k \phi_k^T \right]. \end{aligned}$$

定义  $p(x) = \frac{g(x)xG(x) + g^2(x)}{G^2(x)}$ , 注意到在假设 2 下有  $G(x) = 1 - G(-x)$  及  $g(x) = g(-x)$ , 则 Hessian 矩

阵可以被改写为如下形式:

$$\begin{aligned} \nabla^2 l(\theta) = & - \sum_{k=1}^N [p(C - \phi_k^T \theta) I_{[s_k=1]} \\ & + p(\phi_k^T \theta - C) I_{[s_k=0]}] \phi_k \phi_k^T. \end{aligned}$$

因此, 对数似然函数的凹性等价于 Hessian 矩阵  $\nabla^2 l(\theta)$  的正定性. 因为  $A = \sum_{k=1}^N \phi_k \phi_k^T \geq 0$ , 如果对于任何  $\theta \in \mathbb{R}^p$ , 均成立  $\nabla^2 l(\theta) \leq 0$ , 则对于任何  $x \in \mathbb{R}$ , 我们有  $p(x) \geq 0$ . 由此可以得到, 对于任何  $x \in \mathbb{R}$ , 我们有  $g(x)xG(x) + g^2(x) \geq 0$ , 即, 对于任意  $x > 0$ , 都有

$$\begin{aligned} x(1-F(x)-p^*+2p^*F(x))+(2p^*-1)f(x) & \geq 0, \\ \Leftrightarrow p^*[2xF(x)+2f(x)-x] & \geq xF(x)+f(x)-x. \end{aligned} \quad (20)$$

因  $\forall x \geq 0, 2F(x)-1 \geq 0; \forall x \leq 0, 2F(x)-1 \leq 0$ , 则  $[2F(x)-1]x \geq 0$ . 注意到  $f(x) > 0, \forall x \in \mathbb{R}$ , 故  $2xF(x)+2f(x)-x = [2F(x)-1]x + 2f(x) > 0, \forall x \in \mathbb{R}$ . 因此, 如果式(20)对任意  $x \in \mathbb{R}$  均成立, 则  $p^*$  需满足

$$p^* \geq \inf_x \frac{x F(x) + f(x) - x}{2x F(x) + 2f(x) - x}. \quad (21)$$

然而, 由 L'Hôpital's rule [3], 得到

$$\lim_{x \rightarrow -\infty} \frac{x F(x) + f(x) - x}{2x F(x) + 2f(x) - x} = 1,$$

上式表明  $p^* \geq 1$ . 这一矛盾意味着只要  $p^* \neq 1$ , 对数似然函数不可能在全空间  $\mathbb{R}^p$  上凹. 引理的其余部分可以直接通过式(21)得到, 由此引理得证.

**注 6** 引理1表明, 在  $h(C, \theta_0, \phi_k) = p^*$  的情况下, 对数似然函数(4)在全空间  $\mathbb{R}^p$  上不可能是凹的. 然而, 如果限制在参数空间的有限集合上, Hessian 矩阵  $\nabla^2 l(\theta)$  仍然可能是负定的, 改用投影形式的迭代, 算法1仍适用. 第4节中的模拟也验证了算法1在情况下的有效性.

**注 7** 引理1中的概率值  $p^*$  需要已知. 在实际情形中, 若该概率值未知, 可以通过事先的大量试验对该信道错误概率进行测试与估计, 得到的正确传输的概率估计可以作为已知值.

当  $h(C, \theta_0, \phi_k) = K(C - \theta^T \phi_k)$  时, 式(3)定义的函数  $G(\cdot)$  及其导数  $g(\cdot)$  可以被计算为

$$\begin{aligned} G(x) &= (2K(x) - 1)F(x) + 1 - K(x), \\ g(x) &= G'(x) = 2K'(x)F(x) \\ &+ (2K(x) - 1)f(x) - K'(x). \end{aligned}$$

在这种情形下, 下面的定理给出了对数似然函数至多存在一个最大值点的条件.

**定理 3** 考虑系统(1)-(2)以及正确观测概率函数

形式(18), 其中  $h(C, \theta_0, \phi_k) = K(C - \theta^T \phi_k)$ , 并且假设2-3成立. 如果  $K(x)$  满足

$$\inf_{|x| \leq C+LM} \frac{g(x)xG(x) + g^2(x)}{G^2(x)} > 0,$$

其中  $L = \sup_{\theta \in \Theta} \|\theta\|$  及  $M = \sup_k \|\phi_k\|$ , 则由(4)给出的对数似然函数  $l(\theta)$  至多存在一个最大值点.

**证** 由于在这种情形下, Hessian 矩阵为

$$\begin{aligned} \nabla^2 l(\theta) = & - \sum_{k=1}^N [p(C - \phi_k^T \theta) I_{[s_k=1]} \\ & + p(\phi_k^T \theta - C) I_{[s_k=0]}] \phi_k \phi_k^T, \end{aligned}$$

其中

$$p(x) = \frac{g(x)xG(x) + g^2(x)}{G^2(x)},$$

则由假设  $\inf_{|x| \leq C+LM} p(x) > 0$ , 对于任意  $k \leq N$ , 存在  $\epsilon > 0$  使得在参数空间  $\Theta$  中, 均成立

$$\min_{k \leq N} (p(C - \phi_k^T \theta) I_{[s_k=1]} + p(\phi_k^T \theta - C) I_{[s_k=0]}) \geq \epsilon.$$

由此可以得到

$$\nabla^2 l(\theta) \leq -\epsilon \sum_{k=1}^N \phi_k \phi_k^T = -\epsilon A.$$

因此, 由假设3, 得到对数似然函数  $l(\theta)$  在集合  $\Theta$  上的强凹性. 由强凹性, 对数似然函数  $l(\theta)$  在集合  $\Theta$  上至多有一个最大值点.

**注 8** 定理3给出了在有限参数集合  $\Theta$  上对数似然函数的强凹性. 实际上, 对于一些保守的函数  $K(x)$ , 即错误概率随着真实输出与阈值之间误差绝对值的增大迅速减少的函数, 对数似然函数仍然可以在全空间上强凹, 由此保证了其在全空间上最大值点的存在唯一性. 例如, 如果正确观测概率  $K(x)$  建模为

$$K(x) = 1 - ae^{-bx^2}, \quad (22)$$

则存在正常数  $a^*$  和  $b^*$ , 使得对于任意  $0 < a < a^*$  及  $b > b^*$ , 在假设 2-3 下, 对数似然函数  $l(\theta)$  在全空间上至多存在一个最大值点.

为了给出判断 MLE 存在的条件, 根据文献 [21] 的结果, 我们使用了一些类似的条件.

**定义 2.** 定义

$$\begin{aligned} \Psi = & (\phi_1 (I_{[s_1=0]} - I_{[s_1=1]}), \\ & \dots, \phi_N (I_{[s_N=0]} - I_{[s_N=1]})) . \end{aligned}$$

给定输入数据  $\mathcal{J}_N$  和二值观测数据  $\mathcal{O}_N$ , 如果存在非零向量  $\gamma \in \mathbb{R}^n$  使得  $\Psi^T \gamma \geq 0$ , 则数据  $(\mathcal{J}_N, \mathcal{O}_N)$  被称作非有效的, 否则被称作有效的.

下面的定理给出了对数似然函数在参数集合  $\Theta$  上存在唯一最大值点的充分必要条件.

**定理 4** 考虑系统 (1)-(2) 以及正确观测概率函

数形式(18), 其中  $h(C, \theta_0, \phi_k) = K(C - \theta^T \phi_k)$ , 并且假设 2-3 成立. 如果  $K(x)$  满足定理 3 中的假设以及  $\forall |x| \leq C + LM$  成立

$$G'(x) = 2K'(x)F(x) + (2K(x) - 1)f(x) - K'(x) > 0,$$

其中  $L = \sup_{\theta \in \Theta} \|\theta\|$  及  $M = \sup_k \|\phi_k\|$ , 则对数似然函数(4)在  $\Theta$  上存在唯一最大值点当且仅当数据集  $(\mathcal{J}_N, \mathcal{O}_N)$  是有效的.

**证** 我们从两个方向证明这个定理.

充分性. 该证明与 [21]中定理2的证明相似, 因此在这里略去.

必要性. 如果  $(\mathcal{J}_N, \mathcal{O}_N)$  是非有效的, 则存在非零向量  $\gamma \in \mathbb{R}^n$  使得  $\Psi^T \gamma \geq 0$ . 由于  $\Psi \Psi^T = \sum_{k=1}^N \phi_k \phi_k^T$ , 则由假设3, 我们有  $\Psi^T \gamma > 0$ , 这意味着向量  $\Psi^T \gamma$  至少存在一个非零元素. 对于有限参数集合  $\Theta$  中任意的  $\theta$ , 定义标量函数  $h_{\theta, \gamma}(r)$ :

$$\begin{aligned} h_{\theta, \gamma}(r) &= l(\theta + r\gamma) \\ &= \sum_{k=1}^N [\log [G(-\phi_k^T \gamma r + C - \phi_k^T \theta)] I_{[s_k=1]} \\ &\quad + \log [G(\phi_k^T \gamma r + \phi_k^T \theta - C)] I_{[s_k=0]}]. \end{aligned}$$

注意到向量  $\Psi^T \gamma$  的第  $k$  个元素为

$$(\Psi^T \gamma)_k = -\phi_k^T \gamma I_{[s_k=1]} + \phi_k^T \gamma I_{[s_k=0]}$$

其与函数  $h_{\theta, \gamma}(r)$  的变量  $r$  的线性系数一致. 又由假设知,  $\forall |x| \leq C + LM$ ,

$$G'(x) = 2K'(x)F(x) + (2K(x) - 1)f(x) - K'(x) > 0,$$

并且注意到  $\log(\cdot)$  递增, 因此得到  $h_{\theta, \gamma}(r)$  是严格递增的函数. 假设存在函数  $l(\theta)$  的最大值点  $\theta^*$ , 这与  $h_{\theta^*, \gamma}(r) = l(\theta + r\gamma)$  严格递增矛盾. 所以, 对数似然函数  $l(\theta)$  在参数空间  $\Theta$  上没有最大值点. 定理得证.

**注 9** 当  $K(x)$  建模为  $K(x) = 1 - ae^{-bx^2}$ , 同样可以得到在全空间下存在唯一最大值点的充分必要条件. 具体为: 若  $a, b$  满足  $0 < a < \min\{a^*, 1/2\}$  及  $b > b^*$ , 则在假设2-3下, 对数似然函数  $l(\theta)$  在全空间上存在唯一最大值点等价于数据集  $(\mathcal{J}_N, \mathcal{O}_N)$  是有效的. 证明与定理4的证明类似, 只需要验证在这种情形下  $G'(x) > 0$ . 事实上,

$$G'(x) = 2(2F(x) - 1)xabe^{-bx^2} + (1 - 2ae^{-bx^2})f(x).$$

因为  $a < 1/2$ , 我们有  $1 - 2ae^{-bx^2} > 0$ . 因此, 再结合  $[2F(x) - 1]x \geq 0$ , 有  $G'(x) > 0$ .

定义2不容易验证数据的有效属性, [21]给出了一些例子, 并给出了下述验证有效性质的方法

**准则.** [21] 如果存在  $\rho \in \mathbb{R}^N > 0$  使得  $\Psi \rho = 0$ , 则数据  $(\mathcal{J}_N, \mathcal{O}_N)$  是有效的.

### 3.3 迭代估计误差的收敛性

在本节中, 我们将使用Lyapunov方法证明算法1在通信过程错误和 FIR 过程错误情况下的收敛性. 如定理4所示, 如果数据  $(\mathcal{J}_N, \mathcal{O}_N)$  无效, 则对数似然函数不存在任何有限最大值点, 因此, 以下假设是必要的.

**假设 4** 数据集  $(\mathcal{J}_N, \mathcal{O}_N)$  是有效的.

下面的定理证明了算法1的指数收敛性.

**定理 5** 考虑系统 (1)-(2), 正确观测概率函数形式 (18), 其中  $h(C, \theta_0, \phi_k) = K(C - \theta^T \phi_k)$  或  $h(C, \theta_0, \phi_k) = p^*$ , 及迭代求解算法1, 如果假设 2-4 成立, 则存在  $\epsilon \in (0, 1)$  使得由(16)或(17)迭代生成的序列  $\{\hat{\theta}_t\}$  满足

$$\|\hat{\theta}_t - \hat{\theta}\| \leq \sqrt{\frac{Q_1}{\lambda_{\min}(A)}} \cdot \frac{\sqrt{(1-\epsilon)^t}}{1 - \sqrt{(1-\epsilon)}},$$

其中  $\hat{\theta}$  是由 (6) 给出的 MLE;  $\|\cdot\|$  是 Euclidean 范数;  $A = \sum_{k=1}^N \phi_k \phi_k^T$ ,  $\lambda_{\min}(A)$  是  $A$  的最小特征值;  $Q_1 = (1 - \epsilon)^{-1}(\hat{\theta}_2 - \hat{\theta}_1)^T A(\hat{\theta}_2 - \hat{\theta}_1)$ .

在给出定理的证明之前, 先给出下面关于投影算子的性质.

**引理 2** [5] 定义1中给出的投影算子满足  $\forall x, y \in \mathbb{R}^p$ ,  $\|\Pi_{\Theta, A}(x) - \Pi_{\Theta, A}(y)\|_A \leq \|x - y\|_A$ .

**证** 首先, 由式(7), 我们将式(16)转换为

$$\begin{aligned} \hat{\theta}_{t+1} &= \Pi_{\Theta, A} \left( \hat{\theta}_t + \left( \sum_{k=1}^N \phi_k \phi_k^T \right)^{-1} \nabla l(\hat{\theta}_t) \right) \\ &= \Pi_{\Theta, A} \left( \hat{\theta}_t + A^{-1}(\nabla l(\hat{\theta}_t)) \right). \end{aligned}$$

迭代上述方程, 并由引理2, 可以得到

$$\begin{aligned} \|\hat{\theta}_{t+1} - \hat{\theta}_t\|_A & \\ \leq \|\hat{\theta}_t - \hat{\theta}_{t-1} + A^{-1}(\nabla l(\hat{\theta}_t) - \nabla l(\hat{\theta}_{t-1}))\|_A. & \quad (23) \end{aligned}$$

由中值定理, 我们有

$$\nabla l(\hat{\theta}_t) - \nabla l(\hat{\theta}_{t-1}) = \nabla^2 l(\check{\theta}_{t,t-1})(\hat{\theta}_t - \hat{\theta}_{t-1}). \quad (24)$$

其中  $\check{\theta}_{t,t-1}$  在  $\hat{\theta}_t$  和  $\hat{\theta}_{t-1}$  之间, 即存在常数  $0 < c < 1$  使得  $\check{\theta}_{t,t-1} = (1 - c)\hat{\theta}_t + c\hat{\theta}_{t-1}$ . 由(23)和(24), 我们有

$$\begin{aligned} \|\hat{\theta}_{t+1} - \hat{\theta}_t\|_A & \\ \leq \|(I + A^{-1}(\nabla^2 l(\check{\theta}_{t,t-1}))) (\hat{\theta}_t - \hat{\theta}_{t-1})\|_A. & \quad (25) \end{aligned}$$

由定理2, 存在  $\epsilon > 0$  使得对于任意  $t \geq 1$ ,

$$-A < \nabla^2 l(\check{\theta}_{t,t-1}) \leq -\epsilon \sum_{k=1}^N \phi_k \phi_k^T = -\epsilon A. \quad (26)$$

记  $B_t = -\nabla^2 l(\check{\theta}_{t,t-1})$ ,  $x_t = \hat{\theta}_{t+1} - \hat{\theta}_t$ . 则由(25)可得

$$\|x_t\|_A \leq \|(I - A^{-1}B_t)(x_{t-1})\|_A.$$



令Lyapunov 函数  $Q_t = (1 - \epsilon)^{-t} x_t^T A x_t$ , 则得到

$$Q_t \leq (1 - \epsilon)^{-t} x_{t-1}^T (I_n - A^{-1} B_t)^T \cdot A (I_n - A^{-1} B_t) x_{t-1}.$$

剩下的证明与 [21]中的定理3类似, 我们在这里略去.

该定理证明了迭代算法的收敛性, 同时得到了收敛速率, 这与无误差模型的迭代算法的收敛速度相同.

## 4 仿真分析

在本节中, 我们将通过一些模拟来说明主要结果.

### 4.1 对数似然曲线

为了直观地显示对数似然函数, 我们将模型维度限制为  $p = 1$ . 考虑系统  $y_k = \theta \phi_k + e_k$ , 其中真实参数  $\theta = 1$ .

**数据生成过程.** 固定样本量  $N = 20$ , 然后通过如下Matlab codes生成输入  $\phi_k$  和噪声  $e_k$ ,

$$\text{Phi} = \text{randn}(1, N); \text{E} = \text{randn}(1, N).$$

#### 对数似然函数曲线.

对于固定概率情形, 令  $p_k = 0.99$  和  $p_k = 1$ , 其中后者即为无观测错误情形. 对于  $p_k = K(C - \theta \phi)$ , 选择  $K(x) = 1 - \frac{1}{10} e^{-1/2x^2}$  的时变概率情况. 这三种情况的对数似然函数  $l(\theta)$  如图1所示, 其中  $\theta \in (-4, 4)$ . 由图可知, 在固定概率情况下, 只要存在微小的错误概率, 对数似然函数就不是全空间上的凹函数, 但是在真实参数的某个有限邻域集合内, 仍然是凹的. 对于选择的时变概率情况, 对数似然函数是全空间上的凹函数.

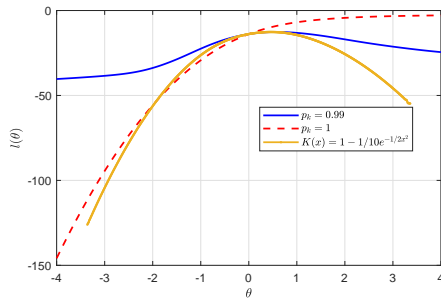


图 1 固定概率情况下的对数似然函数

### 4.2 迭代算法的收敛性

在本节中, 通过数值模拟证明了所提出迭代算法(17)的收敛性. 具体的仿真步骤如下:

**步骤1. 数据生成.** 固定数据长度  $N = 1000$ , 模型维数  $p = 3$ , 传感器阈值  $C = 1$ , 模型参数  $\theta = (-1, 0, 1)^T$ , 误差  $\epsilon_N$  和输入数据  $J_N$  根据标准正态分布生成. Matlab代码如下:

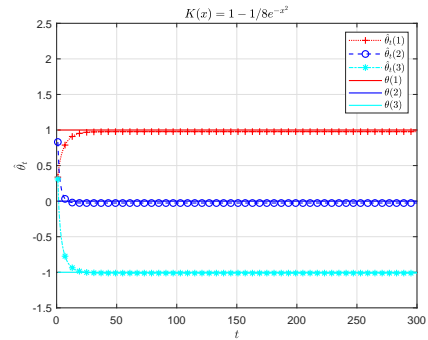
$$\text{error} = [\text{randn}(N, 1)]; \quad \text{phi} = [\text{randn}(N, 3)]$$

二值观测数据  $\mathcal{O}_N$  根据模型(1)和(2)生成.

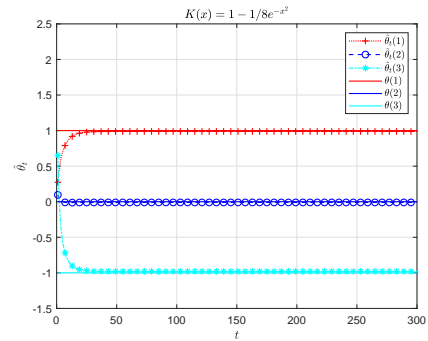
**步骤2. 初始向量选择.** 为了证明算法1可以收敛到具有相同有效数据  $\{J_N, \mathcal{O}_N\}$  的唯一 MLE, 我们采用随机向量作为迭代初始向量  $\hat{\theta}_1$ .  $\hat{\theta}_1$  由以下Matlab代码生成:  $\text{initial} = \text{randn}(3, 1) \times 3$ .

**步骤3. 参数估计.** 根据初始值  $\hat{\theta}_1$  和迭代过程(16)和(17), 我们可以生成迭代估计  $\{\hat{\theta}_t, t \geq 1\}$ .

迭代估计的仿真结果如图2, 和图3所示, 对应的错误概率函数分别取为  $K(x) = 1 - 1/8e^{-x^2}$ , 及  $p_k = 0.8$ . 对于不同的初始向量, 所有估计分量  $\{\hat{\theta}_t\}$  收敛到一个唯一的 MLE, 且接近于真实参数. 最后, 我们也使用传统的集值辨识算法 [21]进行了仿真, 但并未考虑可能存在的量测误差, 结果如图4所示, 对应的错误概率函数取为  $K(x) = 1 - 1/10e^{-1/2x^2}$ . 从图中可以看出, 在存在量测误差的情况下, 传统算法不会收敛到真实参数, 而算法1仍可以收敛到真实参数.



(a)



(b)

图 2  $K(x) = 1 - 1/8e^{-x^2}$  时不同初始值下的估计.

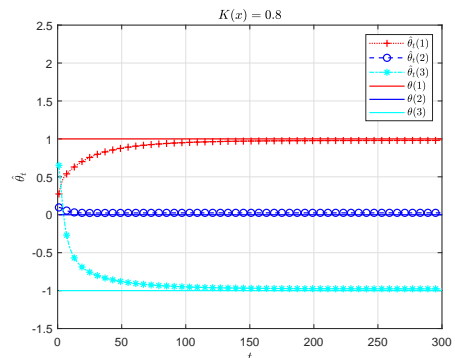


图 3 固定概率情况下的迭代估计结果

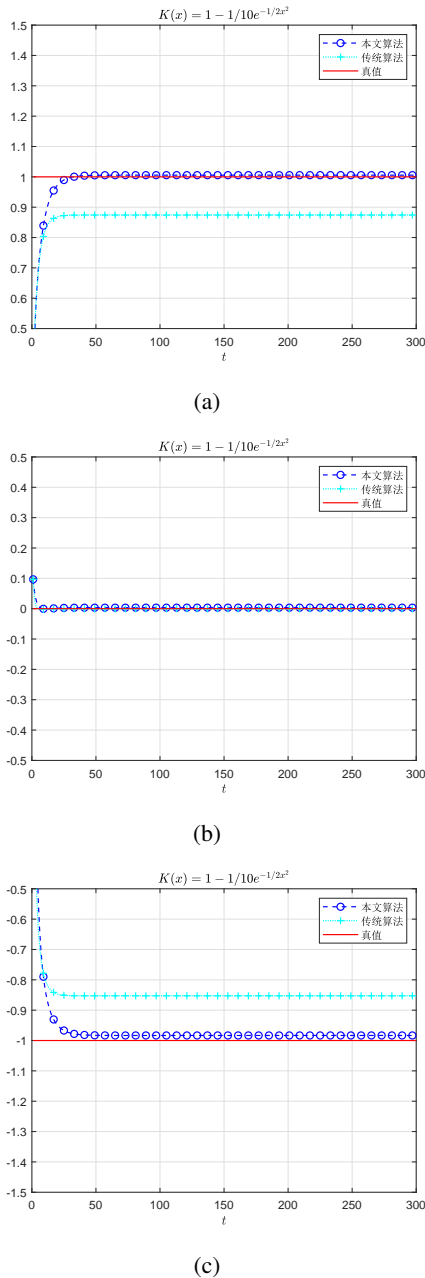


图 4 对比结果图.

## 5 结论与展望

本文考虑了具有量测误差的二值输出FIR系统的参数辨识问题, 其中量测误差使得二值型量测值有一定概率得到相反的取值. 首先, 计算了似然函数及Fisher信息阵并给出了MLE. 为分析收敛性, 给出了关于噪声和参数空间的正则性假设. 在此假设下证明了MLE的强收敛性和渐近正态性. 其次, 基于EM算法的思想, 通过分析似然函数的性质, 给出了MLE的迭代求解算法. 为保证迭代估计序列的有界性, 提出了带投影的迭代求解算法. 此外, 在给定数量的观测、噪声标准正态和回归量持续激励的假设下, 进一步分析了似然函数, 给出了MLE存在唯一性的必要和充分条件及相关条件的验证准则, 同时基于MLE的存在唯一性, 证明了迭代估计误差以指数速度收敛到

零. 最后, 典型系统的数值模拟结果展示了所提方法的有效性.

未来可以进一步考虑将算法应用于观察到相反二值输出的一般情况, 并设计更鲁棒的算法.

## 参考文献:

- [1] AGUERO J. C, GOODWIN G. C, and YUZ J. I. System identification using quantized data. In *2007 46th IEEE Conference on Decision and Control*, 2007: 4263–4268.
- [2] AKYILDIZ I. F, SU W, SANKARASUBRAMANIAM Y, and CAYIRCI E. Wireless sensor networks: a survey. *Computer networks*, 2002, 38(4):393–422.
- [3] BARTLE R. G, and SHERBERT D. R. . *Introduction to real analysis*. John Wiley & Sons, Inc., 2000.
- [4] CARBONE P, SCHOUKENS J. and MOSCHITTA A. Quick estimation of periodic signal parameters from 1-bit measurements. *IEEE Transactions on Instrumentation and Measurement*, 2019 69(2):339–353.
- [5] CHENEY E. W, CHENEY E, and CHENEY W. . *Analysis for applied mathematics*, volume 1. Springer, 2001.
- [6] CHOW Y. S, and TEICHER H. *Probability theory: independence, interchangeability, martingales*. Springer Science & Business Media, 2003.
- [7] COLINET E, and JUILLARD J. A weighted least-squares approach to parameter estimation problems based on binary measurements. *IEEE Transactions on Automatic Control*, 2009, 55(1):148–152.
- [8] DEPRAETERE B, STOEJ J, PINTE G, and SWEVERS J. Identification of linear systems with binary outputs using short independent experiments. *IFAC Proceedings Volumes*, 2012, 45(16):7–12.
- [9] GODOY B. I, GOODWIN G. C, AGÜERO J. C, MARELLI D, and WIGREN T. On identification of fir systems having quantized output data. *Automatica*, 2011, 47(9):1905–1915.
- [10] GUO J, and ZHAO Y. Recursive projection algorithm on fir system identification with binary-valued observations. *Automatica*, 2013, 49(11):3396–3401.
- [11] SAKAI G, MATSUNAGA N, SHIMANOE K, and YAMAZOE N. Theory of gas-diffusion controlled sensitivity for thin film semiconductor gas sensor. *Sensors and Actuators B: Chemical*, 2001, 80(2):125–131.
- [12] TAN S, GUO J, ZHAO Y, and ZHANG J. Adaptive control with saturation-constrained observations for drag-free satellites—a set-valued identification approach. *Science China Information Sciences*, 2021, 64:1–12.
- [13] WANG L. Y, YIN G. G, ZHANG J.-F, and ZHAO Y. *System identification with quantized observations*. Springer, 2010.
- [14] WANG T, BI W, ZHAO Y, and XUE W. Radar target recognition algorithm based on rcs observation sequence-set-valued identification method. *Journal of Systems Science and Complexity*, 2016, 29(3):573–588.
- [15] WANG T, HU M, and ZHAO Y. Adaptive tracking control of fir systems under binary-valued observations and recursive projection identification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019 51(9):5289–5299.
- [16] WANG T, TAN J, ZHAO Y. Asymptotically efficient non-truncated identification for fir systems with binary-valued outputs. *Sci. China Inf. Sci.*, 2018, 61(12):1–3.
- [17] WANG Y, ZHAO Y, ZHANG J.-F, and GUO J. A unified identification algorithm of fir systems based on binary observations with time-varying thresholds. *Automatica*, 2022, 135:109990.
- [18] YOU K. Recursive algorithms for parameter estimation with adaptive quantizer. *Automatica*, 2015, 52:192–201.

- [19] ZHANG H, WANG T, and ZHAO Y. Asymptotically efficient recursive identification of fir systems with binary-valued observations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019 51(5):2687–2700.
- [20] ZHANG J.-F, YIN G. G. System identification using binary sensors. *IEEE Transactions on Automatic Control*, 48(11):1892–1907, 2003.
- [21] ZHAO Y, BI W, and WANG T. Iterative parameter estimate with batched binary-valued observations. *Science China Information Sciences*, 2016, 59(5):1–18.

#### 作者简介:

**郭健** 博士研究生, 目前研究方向为稀疏系统辨识和模型参考自适应控制, E-mail: j.guo@amss.ac.cn;

**薛文超** 中科院数学与系统科学研究院副研究员、博士生导师, 研究领域主要为非线性不确定系统控制与估计等, E-mail: wenchaoxue@amss.ac.cn;

**王婷** 北京科技大学智能科学与技术学院副教授, 研究领域主要为多智能体趋同控制、集值系统的辨识与自适应控制, E-mail: wangting@ustb.edu.cn;

**张纪峰** 中科院数学与系统科学研究院研究员、博士生导师, IEEE Fellow, IFAC Fellow, CAA Fellow, CSIAM Fellow, 欧洲科学与艺术学院院士, 国际系统与科学院院士, 研究领域主要为系统建模与辨识、随机系统、适应控制、多个体系统、有限信息系统等, E-mail: jif@iss.ac.cn.