System Identification Using Binary Sensors

Le Yi Wang, *Senior Member, IEEE*, Ji-Feng Zhang, *Senior Member, IEEE*, and G. George Yin, *Fellow, IEEE*

***Abstract*—System identification is investigated for plants that are equipped with only binary-valued sensors. Optimal identification errors, time complexity, optimal input design, and impact of disturbances and unmodeled dynamics on identification accuracy and complexity are examined in both stochastic and deterministic information frameworks. It is revealed that binary sensors impose fundamental limitations on identification accuracy and time complexity, and carry distinct features beyond identification with regular sensors. Comparisons between the stochastic and deterministic frameworks indicate a complementary nature in their utility in binary-sensor identification.**

***Index Terms*—Binary sensors, estimation, system identification, time complexity.**

I. Introduction

BINARY-VALUED sensors are commonly employed in practical systems. Usually they are far more cost effective than regular sensors. In many applications they are the only ones available during real-time operations. There are numerous examples, such as switching sensors for exhaust gas oxygen, ABS, shift-by-wire, in automotive applications; industry sensors for brush-less dc motors, liquid levels, pressure switches; chemical process sensors for vacuum, pressure, and power levels; traffic condition indicators in the asynchronous transmission mode (ATM) networks; gas content sensors (CO, CO_2, H_2, etc.) in gas and oil industry. In medical applications, estimation and prediction of causal effects with dichotomous outcomes are closely related to binary-sensor systems. Before proceeding further, we present examples in three different application areas.[1]

1) **ATM ABR Traffic Control [28]**: An ATM network consists of sources, switches, and destinations. Due to variations in other higher priority network traffic, such as constant bit rate (CBR) and variable bit rate (VBR), an available bit rate (ABR) connection experiences significant uncertainty on the available bandwidth during its operation. A physical or logical buffer is used in a switch to accommodate bandwidth fluctuations. The actual amount of bandwidth an ABR connection receives is provided to the source using rate-based closed-loop feedback control. One typical technique for providing traffic information is relative rate marking, which uses two fields in the resource management (RM) cell—the no increase (NI) bit and the congestion indication (CI) bit. The NI bit is set when the queue reaches length C_1, and the CI bit is set when the queue length reaches C_2 ($C_2 > C_1$).

   In this system, the queue length is not directly available for traffic control. The NI and CI bits indicate merely that it takes values in one of the three uncertainty sets [0, C_1], (C_1, C_2] and (C_2, ∞). This can be represented by a typical case of tracking control with two binary sensors. It is noted that the desired queue length is usually a value between C_1 and C_2, rather than C_1 or C_2.
2) **LNT and Air-to-Fuel Ratio Control With an EGO Sensor [36], [37]**: In automotive and chemical process applications, oxygen sensors are widely used for evaluating gas oxygen contents. Inexpensive oxygen sensors are switching types that change their voltage outputs sharply when excess oxygen in the gas is detected. In particular, in automotive emission control, the exhaust gas oxygen sensor (EGO or HEGO) will switch its outputs when the air-to-fuel ratio in the exhaust gas crosses the stoichiometric value.

   To maintain conversion efficiency of the three-way catalyst or to optimize the performance of a lean NOx trap (LNT), it is essential to estimate the internally stored NOx and oxygen. In this case the switching point of the sensor has no direct bearing with the control target. The idea of using the switching sensor for identification purposes, rather than for control only, has resulted in a new emission control strategy [36], [37].
3) **Identification of Binary Perceptrons**: There is an interesting intersection between this study and statistical learning theory in neural networks. Consider an unknown binary perceptron that is used to represent a dynamic relationship:

$$y(t) = s(w_1 u(t) + w_2 u(t-1) + \cdots + w_n u(t-n+1) - C)$$

   where C is the known neuron firing threshold, $w_1, \ldots, w_n$ are the weightings to be learned, and $s(\cdot)$ is a binary-valued function switching at 0. This

Manuscript received October 10, 2002; revised May 5, 2003. Recommended by Associate Editor A. Garulli. The work of L. Y. Wang was supported in part by the National Science Foundation, the Ford Motor Company, and the Wayne State University Research Enhancement Program (REP). The work of J.-F. Zhang was supported in part by the National Natural Science Foundation of China, and the Ministry of Science and Technology of China. The work of G. G. Yin was supported in part by the National Science Foundation and the Wayne State University REP.

L. Y. Wang is with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202 USA (e-mail: lywang@ece.eng.wayne.edu).

J.-F. Zhang is with the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China (e-mail: jif@mail.iss.ac.cn).

G. G. Yin is with the Department of Mathematics, Wayne State University, Detroit, MI 48202 USA (e-mail: gyin@math.wayne.edu).

Digital Object Identifier 10.1109/TAC.2003.819073

[1]In all these examples, as well as many other applications, actual systems are discrete-time and involve signal quantization or data compression for computer or digital communication networks implementations. Quantization errors are usually negligibly small. This paper deals with discrete-time, analog-valued signals.

learning problem can be formulated as a special case of binary-sensor identification without disturbances or unmodeled dynamics. Traditional neural models, such as McCulloch–Pitts and Nagumo–Sato models, contain a neural firing threshold that introduces naturally a binary function [3], [13], [15], [23]. Fundamental stochastic neural learning theory studies the stochastic updating algorithms for neural parameters [32]–[34].

A. Problems

The use of binary sensors poses substantial difficulties since only very limited information is available for system modeling, identification and control. Since switching sensors are nonlinear components, studies of their roles and impact on systems are often carried out in nonlinear system frameworks, such as sliding mode control, describing function analysis, switching control, hybrid control, etc. In these control schemes, the switching points of the sensors are directly used to define a control target. However, their fundamental impact on system modeling and identification is largely unexplored. This paper intends to study the inherent consequences of using switching sensors in system identification and its potential in extending control capabilities.

The main scenario, which has motivated this work, is embodied in many applications in which modeling of such systems is of great importance in performing model predictive control, optimal control strategy development, control adaptation, etc. When inputs can be arbitrarily selected within certain bounds and outputs are measured by regular sensors, such identification problems have been extensively studied in the frameworks of either traditional stochastic system identification or worst-case identification. The issues of identification accuracy, convergence, model complexity, time complexity, input design, persistent excitation, identification algorithms, etc., have been pursued by many researchers. A vast literature is now available on this topic; see [19] and [22], among others.

Some fundamental issues emerge when the output sensor is limited to be binary-valued: How accurate can one identify the parameters of the system? How fast can one reduce uncertainty on model parameters? What are the optimal inputs for fast identification? What are the conditions for parameter convergence? What is the impact of unmodeled dynamics and disturbances on identification accuracy and time complexity? In contrast to classical system identification, answers to these familiar questions under switching sensors depart dramatically from the traditional setup.

It will be shown that binary sensors increase time complexity significantly; the optimal inputs differ from those in traditional identification; identification characteristics depart significantly between stochastic and deterministic noise representations; and unmodeled dynamics have fundamental influence on identification accuracy of the modeled part. Contrast to traditional system identification in which the individual merits of stochastic versus worst-case frameworks are still hotly debated, these two frameworks complement each other in binary-sensor identification problems.

B. Organization of the Paper

The paper is organized as follows. After a brief problem formulation in Section II, we start our investigation in Section III on system identification in a stochastic framework. Identification input design, convergence of the estimates, upper and lower bounds on identification errors, and time complexity are established. Section IV studies the identification problem when the disturbance is viewed unknown-but-bounded as in a worst-case framework. The results are significantly different from that of Section III. Identification time complexity and error lower bounds are established first, underscoring an inherent relationship between identification time complexity and the Kolmogorov ε-entropy. Identification input design and upper bounds on identification errors are then derived, demonstrating that Kolmogorov ε-entropy indeed defines the complexity rates. Section V presents a comparison between the stochastic and deterministic frameworks. Contrast to the common perception that these two are competing frameworks, we show that they complement each other in binary-sensor identification. Several examples are presented in Section VI to illustrate utilities of the approach. Finally, some potential future research directions are highlighted in Section VII. An Appendix containing the proofs of several technical results are included at the end of the paper.

C. Related Literature

This paper explores the issues arising in system identification with switching sensors. Traditional system identification using regular sensors is a relatively mature research area that bears a vast body of literature. There are numerous textbooks and monographs on the subject, such as [4], [18], and [19]. The focus of this paper is the impact of binary sensors on time complexity, identification accuracy, identifiability, and input design, which is a significant departure from early works of theoretical developments.

A key issue studied in this paper is time complexity. Complexity issues in identification have been pursued by many researchers. The concepts of ε-net and ε-dimension in the Kolmogorov sense [17] were first employed by Zames [43] in studies of model complexity and system identification. Time complexity in identification was studied in [30], [6], [25], [44], [38], [35], and [39]–[41]. A general and comprehensive framework of information-based complexity was developed in [29]. Milanese is one of the first researchers in recognizing the importance of worst-case identification. Milanese and Belforte [20] and Milanese and Vicino [22] introduced the problem of set membership identification and produced many interesting results on the subject. Many algorithms for worst-case identification have been reported; see [21], [22], and the references therein.

The idea of treating unmodeled dynamics and noise using mixed assumptions was explored in deterministic frameworks in [31]. A unified methodology which combines deterministic identification and probability framework was introduced in [39] and [40]. Many significant results have been obtained for identification and adaptive control involving random disturbances in the past decades [4], [14], [16], [18], [19].

The utility of binary sensors in this paper carries a flavor that is related to many branches of signal processing problems. One class of adaptive filtering problems that has recently drawn considerable attention uses "hard limiters" to reduce computational complexity. The idea, sometimes referred to as binary reinforcement [11], employs the sign operator in the error or/and the regressor, leading to a variety of sign-error, sign-regressor, and sign-sign algorithms. Some recent work in this direction can be found in [5], [7], and [8]. Emerging applications of this idea in wireless communications, e.g., code-division multiple-access implemented with direct-sequence (CDMA/DS), have been reported in [42].

Signal quantization and data compression is a typical A-D conversion process that has been studied extensively in the signal processing and computer science community. Studies of impact of quantization errors can be conducted in a worst-case or probabilistic framework, depending on how quantization errors are modeled. We refer the interested reader to [1], [12], and [26] for a comprehensive coverage of this topic. Quantized sensor information is fundamentally different from binary sensor information since binary sensors do not provide signal error bounds which are essential in quantization analysis.

Statistical learning theory [32], [33], especially its application to neural network models [3], [13], [15], [23], has led to some very interesting new development [34], in which dynamic system identification is studied in neural networks. The problem considered in this paper is motivated by entirely different applications. We study different problem aspects and move toward different directions from neural learning methods. Nevertheless, the intersection witnessed here due to model structure similarity makes potential applications of our results in neural learning theory and *vice versa*.

II. Problem Formulation

For a sequence of real numbers $x = \{x(t); t = 0, 1, \ldots\}$, $||x||_p, p \geq 1$, will be the standard l^p norm. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space (the set of n-tuples of real numbers). A ball of center $c \in \mathbb{R}^n$ and radius $r \geq 0$ (using l^p norm) in $\mathbb{R}^n$ will be denoted by $\mathrm{Ball}_p(c, r) = \{x \in \mathbb{R}^n : ||x - c||_p \leq r\}$. In this paper, the base-2 logarithm $\log_2$ will be simply written as $\log$.

Consider a single-input–single-output (SISO) linear time-invariant stable discrete-time system

$$y(t) = \sum_{i=0}^{\infty} a_i u(t-i) + d(t), \qquad t = t_0 + 1, \ldots$$

where $d(t)$ is the disturbance, $\{u(t)\}$ is the input with $u(t) = 0$, $t < 0$; and $a = \{a_i, i = 0, 1, \ldots\}$, satisfying $||a||_1 = \sum_{i=0}^{\infty} |a_i| < \infty$, is the vector-valued parameter. The input u is uniformly bounded $||u||_\infty \leq u_{\max}$, but can be designed otherwise. The output y is measured by a binary sensor with the known threshold C. Namely, the sensor indicates only whether $y \leq C$ or $y > C$. Without loss of generality, assume $C > 0$.[2] We will use the indicator function

$$s(t) = I_{\{y(t) \leq C\}} = \begin{cases} 1, & \text{if } y(t) \leq C \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

to represent the output of the sensor.

For a given model order n, the system parameters can be decomposed into the modeled part $\theta^T = [a_0, \ldots, a_{n-1}]$ and the unmodeled dynamics $\widetilde{\theta}^T = [a_n, a_{n+1}, \ldots]$. Then, the system input–output relationship becomes

$$y(t) = \phi^T(t)\theta + \widetilde{\phi}^T(t)\widetilde{\theta} + d(t), \qquad t = t_0 + 1, \ldots \tag{2}$$

where $\phi^T(t) = [u(t), u(t-1), \ldots, u(t-n+1)]$, and $\widetilde{\phi}^T(t) = [u(t-n), u(t-n-1), \ldots]$.

Under a selected input sequence u, the output $s(t)$ is measured for $t = t_0 + 1, \ldots, t_0 + N$. We would like to estimate θ on the basis of input/output observation on $u(t)$ and $s(t)$. The issues of identification accuracy, time complexity, and input design will be discussed in both stochastic and deterministic frameworks.

III. Stochastic Frameworks

When the disturbance d is modeled as a stochastic process, both y and s become stochastic processes. We assume the following prior information on the system uncertainty, including unmodeled dynamics and disturbance.

Assumption A1):

1) $\{d(t)\}$ is a sequence of independent and identically distributed (i.i.d.) zero-mean random variables with distribution function $F(x)$, which is a continuous function whose inverse F^{-1} exists and is continuous. The moment generating function of $d(t)$ exists.
2) $||\widetilde{\theta}||_1 \leq \eta$.

Remark 1: A typical example of the noise satisfying A1) is Gaussian random variables. The cases of random variables, whose distribution functions are only invertible in a finite interval $[-\kappa, \kappa]$, can be handled by applying the technique of dithers (see Section III-E) or combining stochastic and deterministic binary-sensor identification (see Section V). The assumption of the noise being a continuous random variable is not a restriction. When one deals with discrete random variables, suitable scaling and the central limit theorem lead to normal approximation.

The following formulation was introduced in [39] and [40]. It treats the disturbance as stochastic but unmodeled dynamics as unknown-but-bounded uncertainty. Consequently, a worst-case probability measure is used as a performance index. For a given set $\mathbb{L}(t_0, u, N)$ of admissible estimates $\widehat{\theta}$ of the true parameter θ, on the basis of N measurements on s starting at t_0 with input u, and an error tolerance level ε, we define

$$\lambda_N(\varepsilon) = \inf_{||u||_\infty \leq u_{\max}} \sup_{t_0} \inf_{\widehat{\theta} \in \mathbb{L}(t_0, u, N)} \sup_{||\widetilde{\theta}||_1 \leq \eta} P(||\widehat{\theta} - \theta||_1 \geq \varepsilon). \tag{3}$$

[2]Sensors with $C = 0$ can only detect the sign and usually do not provide adequate information for identification.

This is the optimal (over the input u and admissible estimates $\widehat{\theta}$) worst-case (over $\widetilde{\theta}$ and t_0) probability of errors larger than the given level ε. Then, for a given confidence level $0 \le \alpha < 1$

$$N_\alpha(\varepsilon) = \min\{N : \lambda_N(\varepsilon) \le \alpha\} \tag{4}$$

is the probabilistic time complexity. It is noted that if $\alpha = 0$, $N_\alpha(\varepsilon)$ is reduced to (module a set of probability measure 0) deterministic worst-case time complexity for achieving estimation accuracy ε.

A. Identification Algorithms and Convergence

For notational simplicity, assume $N = kn$ for some integer $k > 0$. As a result, we can group the N input–output equations into k blocks of size n

$$Y_l = \Phi_l\theta + \widetilde{\Phi}_l\widetilde{\theta} + D_l, \qquad l = 0, 1, \ldots, k-1$$

where $Y_l = [y(t_0+ln+1), \ldots, y(t_0+ln+n)]^T$, $\Phi_l = [\phi(t_0+ln+1), \ldots, \phi(t_0+ln+n)]^T$, $\widetilde{\Phi}_l = [\widetilde{\phi}(t_0+ln+1), \ldots, \widetilde{\phi}(t_0+ln+n)]^T$, $D_l = [d(t_0+ln+1), \ldots, d(t_0+ln+n)]^T$.

In particular, if the input is n-periodic, i.e., $u(t+n) = u(t)$, $\forall t$, we have $\Phi_l = \Phi_0$ and $\widetilde{\Phi}_l = \widetilde{\Phi}_0 = [\Phi_0, \Phi_0, \ldots]$, $l = 0, \ldots, k-1$. Moreover, the n-period input is said to be *full rank* if Φ_0 is invertible. In the following, a scalar function that is applied to a vector will mean component-wise operation of the function.

For each θ (fixed but unknown) and $\widetilde{\theta}$, define

$$\xi_k^i = \frac{1}{k}\sum_{l=0}^{k-1} s(t_0 + ln + i), \qquad i = 1, \ldots, n \tag{5}$$

$\xi_k = (\xi_k^1, \ldots, \xi_k^n)^T$. Note that the event $\{y(t_0 + ln + i) \le C\}$ is the same as the event $\{d(t_0 + ln + i) \le \widetilde{c}_i\}$, where $\widetilde{c}_i = C - \widetilde{C}_i$ and $\widetilde{C}_i$ is the i the component of $\Phi_l\theta + \widetilde{\Phi}_l\widetilde{\theta}$. Denote $\widetilde{c} = [\widetilde{c}_1, \ldots, \widetilde{c}_n]^T$. Then, ξ_k^i is precisely the value of the k-sample empirical distribution $\widehat{F}_k(x)$ of the noise d at $x = \widetilde{c}_i$. Denote $\xi_k = (\widehat{F}_k(\widetilde{c}_1), \ldots, \widehat{F}_k(\widetilde{c}_n))^T$. Let

$$\widehat{B}_k(x) = \sqrt{k}(\widehat{F}_k(x) - F(x)), \quad \text{for each } \ x \in \mathbb{R}.$$

Theorem 1: Under Assumption A1), the following assertions hold.

a) For any compact subset $S \subset \mathbb{R}$

$$\lim_{k\to\infty} \sup_{x\in S} |\widehat{F}_k(x) - F(x)| \to 0 \ \text{ w.p.1.} \tag{6}$$

b) $\widehat{B}_k(\cdot)$ converges weakly to $B(\cdot)$, a stretched Brownian bridge process such that the covariance of $B(\cdot)$ (for $x_1, x_2 \in \mathbb{R}$) is given by

$$EB(x_1)B(x_2) = \min(F(x_1), F(x_2)) - F(x_1)F(x_2). \tag{7}$$

Proof: By virtue of the well-known Glivenko–Cantelli Theorem [2, p. 103], $|\widehat{F}_k(x) - F(x)| \to 0$ w.p. 1, and the convergence is uniform on any compact subset $\mathbb{R}$. This yields a). Part b) follows from the convergence of a centered and scaled estimation error of empirical measures; see, for instance, [2, p. 105] and [24, p. 95]. □

Remark 2: The processes considered here is known as the empirical measure or sample distribution of the underlying sequence. Part a) above says that for large k, $\widehat{F}_k(x)$ should approximate the corresponding distribution function $F(x)$ uniformly on compact subsets. The Glivenko–Cantelli theorem is the best-known uniform strong law of large numbers in the literature.

We will find the limit distribution of a suitably scaled sequence of $\widehat{F}_k(x) - F(x)$ so that the convergence rate can be determined. The central limit theorem gives us hints on using a scaling factor $\sqrt{k}$. The limit turns out to be a Brownian bridge (see [2, p. 64] and [24, p. 95]). Note that a Brownian bridge is a function of a Brownian motion defined on [0, 1]. Loosely, it is a Brownian motion tied down at both endpoints of the interval [0, 1]. Between the two end points, the process evolves just as a Brownian motion. Now in the current case, since $\widetilde{c}_i$ can take real values outside [0, 1], the Brownian bridge becomes a stretched one. The terminology "stretched Brownian bridge" follows that of [24, p. 178].

Since F is invertible, we can define

$$\begin{aligned}\gamma_k^i =& F^{-1}(\widehat{F}_k(\widetilde{c}_i)) = F^{-1}(\xi_k^i), \ \ i = 1, \ldots, n, \text{ and}\\ \gamma_k =& [\gamma_k^1, \ldots, \gamma_k^n]^T = F^{-1}(\xi_k), \ L_k = C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \gamma_k. \end{aligned} \tag{8}$$

When the input u is n-periodic and full rank, Φ_0 is invertible and we define the estimate

$$\widehat{\theta}_k = \Phi_0^{-1}L_k. \tag{9}$$

Theorem 2: Under Assumption A1), if the input u is n-periodic and full rank, then $\widehat{\theta}_k$ converges to a constant $\widehat{\theta}$. That is, $\widehat{\theta}_k \to \widehat{\theta}$ w.p. 1 as $k \to \infty$. Furthermore, $||\widehat{\theta} - \theta||_1 \le \eta$, where θ is the true vector-valued parameter.

Remark 3: If $\eta = 0$, i.e., no unmodeled dynamics, then this estimate is unbiased.

Proof: By virtue of Theorem 1, as $k \to \infty$

$$\xi_k = (\xi_k^1, \ldots, \xi_k^n)^T \to (F(\widetilde{c}^1), \ldots, F(\widetilde{c}^n))^T \ \text{ w.p.1..}$$

Thus the continuity of $F^{-1}(\cdot)$ implies that $F^{-1}(\widehat{F}_k(\widetilde{c}_i)) \to \widetilde{c}_i$ w.p. 1. Hence

$$\gamma_k \to \widetilde{c} = \begin{bmatrix}\widetilde{c}_1\\ \vdots\\ \widetilde{c}_n\end{bmatrix} = C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - (\Phi_0\theta + \widetilde{\Phi}_0\widetilde{\theta}) \ \text{ w.p.1..}$$

Due to the periodicity of the input u, we have

$$\widehat{\theta}_k = \Phi_0^{-1}L_k = \Phi_0^{-1}\left(C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \gamma_k\right) \to \theta + \Phi_0^{-1}\widetilde{\Phi}_0\widetilde{\theta} =: \widehat{\theta}$$

w.p. 1. Note that $\Phi_0^{-1}\widetilde{\Phi}_0\widetilde{\theta} = [I, I, \ldots]\widetilde{\theta}$. Finally, by A1), $||\widehat{\theta} - \theta||_1 = ||[I, I, \ldots]\widetilde{\theta}||_1 \le ||\widetilde{\theta}||_1 \le \eta$. □

B. Upper Bounds on Estimation Errors and Time Complexity

Next, we shall establish bounds on identification errors and time complexity for a finite k. For a fixed $k > 0$

$$\begin{aligned}\widehat{\theta}_k - \widehat{\theta} =& \Phi_0^{-1}\left(C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \gamma_k\right) - (\theta + \Phi_0^{-1}\widetilde{\Phi}_0\widetilde{\theta})\\ =& \Phi_0^{-1}\left(C\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} - \gamma_k - (\Phi_0\theta + \widetilde{\Phi}_0\widetilde{\theta})\right)\\ =& \Phi_0^{-1}\left(\widetilde{c} - \gamma_k\right). \end{aligned} \tag{10}$$

Recall that $\widetilde{c} = [\widetilde{c}_1, \ldots, \widetilde{c}_n]^T = [C - \widetilde{C}_1, \ldots, C - \widetilde{C}_n]^T$. Since $||\widehat{\theta}_k - \widehat{\theta}||_1 \le ||\Phi_0^{-1}||_I ||\widetilde{c} - \gamma_k||_1$, where $||\cdot||_I$ is the l_1-induced operator norm, for any $\varepsilon_1 > 0$

$$\begin{aligned} P(||\widehat{\theta}_k - \widehat{\theta}||_1 \ge \varepsilon_1) \le & P\left(||\gamma_k - \widetilde{c}||_1 \ge \frac{\varepsilon_1}{||\Phi_0^{-1}||_I} \right) \\ \le & P\left(||\gamma_k - \widetilde{c}||_\infty \ge \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \\ \le & P\left(\cup_{i=1}^n \left\{ |\gamma_k^i - \widetilde{c}_i| \ge \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right\} \right) \\ \le & \sum_{i=1}^n P\left(|\gamma_k^i - \widetilde{c}_i| \ge \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right). \end{aligned}$$

The inequality $|\gamma_k^i - \widetilde{c}_i| \ge (\varepsilon_1/n||\Phi_0^{-1}||)_I$ is equivalent to

$$\gamma_k^i \ge \widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \text{ or } \gamma_k^i \le \widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I}. \quad (11)$$

Note that $\gamma_k^i = F^{-1}(\xi_k^i)$. Since $F^{-1}(\cdot)$ is monotone

$$\gamma_k^i \ge \widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \Leftrightarrow \xi_k^i \ge F\left(\widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \quad (12)$$

and

$$\gamma_k^i \le \widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \Leftrightarrow \xi_k^i \le F\left(\widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right). \quad (13)$$

It follows that

$$\begin{aligned} & P\left(\left|\left| \widehat{\theta}_k - \widehat{\theta} \right|\right|_1 \ge \varepsilon_1 \right) \\ \le & \sum_{i=1}^n P\left(|\gamma_k^i - \widetilde{c}_i| \ge \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \\ \le & \sum_{i=1}^n P\left(\xi_k^i \ge F\left(\widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right) \\ & + \sum_{i=1}^n P\left(\xi_k^i \le F\left(\widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right). \end{aligned} \quad (14)$$

For simplicity, use short-hand notation $s_l^i = s(t_0 + ln + i)$. Since $\{d(t)\}$ is i.i.d., for each $i = 0, 1, \ldots, n-1$, $\{s_l^i\}$ is also a sequence of i.i.d. random variables. Denote the moment generating function of s_0^i by $G_i(z) = E\exp(zs_0^i)$ with $z \in \mathbb{R}$. Let $g_i(t) = \inf_z E\exp(z(s_0^i - t)) = \inf_z \exp(-zt)G_i(z)$. By the definition of s_0^i, $Es_0^i = F(\widetilde{c}_i)$. By the mononicity of $F(\cdot)$, we have $F\left(\widetilde{c}_i + \varepsilon_1/n||\Phi_0^{-1}||_I\right) > F\left(\widetilde{c}_i\right)$ and $F\left(\widetilde{c}_i - \varepsilon_1/n||\Phi_0^{-1}||_I\right) \le F(\widetilde{c}_i)$. Consequently, an application of Chernoff's inequality [27, p. 326] yields

$$P\left(\xi_k^i \ge F\left(\widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right) \le \left(g_i\left(\widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right)^k \quad (15)$$

and

$$P\left(\xi_k^i \le F\left(\widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right) \le \left(g_i\left(\widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right)^k. \quad (16)$$

Combining (14)–(16), we obtain the following upper bounds.

Theorem 3: For any $\varepsilon_1 > 0$

$$\begin{aligned} & P\left(\left|\left| \widehat{\theta}_k - \widehat{\theta} \right|\right|_1 \ge \varepsilon_1 \right) \le H_{\varepsilon_1}(k): \\ = & \sum_{i=1}^n \left[\left(g_i\left(\widetilde{c}_i + \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right)^k + \left(g_i\left(\widetilde{c}_i - \frac{\varepsilon_1}{n||\Phi_0^{-1}||_I} \right) \right)^k \right]. \end{aligned} \quad (17)$$

Corollary 1: For any $\varepsilon > \eta > 0$, let $N = kn$. Then

a)

$$\lambda_N(\varepsilon) \le H_{\varepsilon-\eta}(k) \quad (18)$$

where $\lambda_N(\varepsilon)$ is defined in (3) and $H_{\varepsilon-\eta}(k)$ in Theorem 3;

b)

$$N_\alpha(\varepsilon) \le n \min\{k : H_{\varepsilon-\eta}(k) \le \alpha\}. \quad (19)$$

Proof:

a) By Theorem 3, the selected input and the estimate $\widehat{\theta}_k$ defined in (9) guarantee that

$$\begin{aligned} P(||\widehat{\theta}_k - \theta||_1 \ge \varepsilon) \le & P(||\widehat{\theta}_k - \widehat{\theta}||_1 + ||\widehat{\theta} - \theta||_1 \ge \varepsilon) \\ \le & P(||\widehat{\theta}_k - \widehat{\theta}||_1 \ge \varepsilon - \eta) \\ \le & H_{\varepsilon-\eta}(k). \end{aligned}$$

Since this is valid for all t_0 and $\widetilde{\theta}$, (18) follows.

b)

$$\begin{aligned} N_\alpha(\varepsilon) = & \min\{N : \lambda_N(\varepsilon) \le \alpha\} \\ \le & \min\{kn : H_{\varepsilon-\eta}(k) \le \alpha\} \\ \le & n \min\{k : H_{\varepsilon-\eta}(k) \le \alpha\} \end{aligned}$$

yields (19). □

Remark 4: Note that s_0^i has

$$\text{mean } F(\widetilde{c}_i) \text{ and variance } F(\widetilde{c}_i)(1 - F(\widetilde{c}_i)). \quad (20)$$

In the special case of Gaussian distribution of $d(t)$, s_0^i is also normally distributed with moment generating function

$$G_i(z) = \exp\left(F(\widetilde{c}_i)z + \frac{1}{2}F(\widetilde{c}_i)(1 - F(\widetilde{c}_i))z^2 \right)$$

Hence $g_i(t) = \exp(-(F(\widetilde{c}_i) - t)^2/(2F(\widetilde{c}_i)(1 - F(\widetilde{c}_i))))$. Using $g_i(t)$, one can then obtain more explicit bounds in Theorem 3.

C. Lower Bounds on Estimation Errors

To obtain lower bounds on the estimation error when the above full rank periodic input is used, we use a similar argument as that of the upper bound case.

From $\Phi_0(\widehat{\theta}_k - \widehat{\theta}) = \widetilde{c} - \gamma_k$, we have $||\widetilde{c} - \gamma_k||_1 \le ||\Phi_0||_I ||\widehat{\theta}_k - \widehat{\theta}||_1$. In view of (10), the independence of ξ_k^i for $i = 1, \ldots, n$ implies that for any $\varepsilon_1 > 0$

$$\begin{aligned} & P\left(\left|\left| \widehat{\theta}_k - \widehat{\theta} \right|\right|_1 \ge \varepsilon_1 \right) \\ \ge & P\left(||\widetilde{c} - \gamma_k||_1 \ge \varepsilon_1 ||\Phi_0||_I \right) \\ \ge & P\left(\cap_{i=1}^n |\widetilde{c}_i - \gamma_k^i| \ge \frac{\varepsilon_1 ||\Phi_0||_I}{n} \right) \\ \ge & \prod_{i=1}^n P\left(|\gamma_k^i - \widetilde{c}_i| \ge \frac{\varepsilon_1 ||\Phi_0||_I}{n} \right) \\ \ge & \prod_{i=1}^n P\left(\xi_k^i \ge F\left(\widetilde{c}_i + \frac{\varepsilon_1 ||\Phi_0||_I}{n} \right) \right) \\ & + \prod_{i=1}^n P\left(\xi_k^i \le F\left(\widetilde{c}_i - \frac{\varepsilon_1 ||\Phi_0||_I}{n} \right) \right). \end{aligned} \quad (21)$$

Our approach of obtaining the lower bounds involves two steps. First, if the random variables are normally distributed, the lower bounds can be obtained via the use of an inequality in [9] together with the properties of a normal distribution. The second step deals with the situation in which the noises are not normal, but are approximately normal via Barry-Esseen estimate.

Assume that $d(t)$ is normally distributed with mean 0 and variance σ^2. Suppose that $Z(x)$ is the distribution of the standard normal random variable, i.e., $Z(x) = \int_{-\infty}^{x} z(\zeta)d\zeta$, where $z(x) = 1/\sqrt{2\pi}\exp(-x^2/2)$, $-\infty < x < \infty$. It was shown in [9, Lemma 2, p. 175] that

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) z(x) < 1 - Z(x) < \frac{1}{x} z(x), \quad \text{for } x > 0. \tag{22}$$

Since $d(t)$ is normally distributed with mean zero and variance σ^2, ξ_k^i is also normally distributed with mean $F(\widetilde{c}_i)$ and variance $F(\widetilde{c}_i)(1 - F(\widetilde{c}_i))/k$. Therefore, $\sqrt{k}(\xi_k^i - F(\widetilde{c}_i))/\sqrt{F(\widetilde{c}_i)(1-F(\widetilde{c}_i))}$ is normally distributed with mean 0 and variance 1. As a result, to obtain the desired lower bounds via (21), for any $\varepsilon_1 > 0$, it suffices to consider $P(\xi_k^i \geq F(\widetilde{c}_i + \varepsilon_1\|\Phi_0\|_I/n))$ and $P(\xi_k^i \leq F(\widetilde{c}_i - \varepsilon_1\|\Phi_0\|_I/n))$. Denote

$$\widetilde{\alpha}_i^+ = \widetilde{\alpha}_i^+(\varepsilon_1) = \frac{\sqrt{k}\left(F\left(\widetilde{c}_i + \frac{\varepsilon_1\|\Phi_0\|_I}{n}\right) - F(\widetilde{c}_i)\right)}{\sqrt{F(\widetilde{c}_i)(1-F(\widetilde{c}_i))}}.$$

Then

$$P\left(\xi_k^i \geq F\left(\widetilde{c}_i + \frac{\varepsilon_1\|\Phi_0\|_I}{n}\right)\right) = P\left(\sqrt{k}\frac{(\xi_k^i - F(\widetilde{c}_i))}{\sqrt{F(\widetilde{c}_i)(1-F(\widetilde{c}_i))}} \geq \widetilde{\alpha}_i^+\right).$$

Therefore, by (22)

$$\begin{aligned} P\left(\xi_k^i \geq F\left(\widetilde{c}_i + \frac{\varepsilon_1\|\Phi_0\|_I}{n}\right)\right) &= 1 - Z(\widetilde{\alpha}_i^+) \\ &\geq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\widetilde{\alpha}_i^+} - \left(\frac{1}{\widetilde{\alpha}_i^+}\right)^3\right) \\ &\quad \exp\left(-\frac{(\widetilde{\alpha}_i^+)^2}{2}\right). \end{aligned} \tag{23}$$

Likewise, denote

$$\widetilde{\alpha}_i^- = \widetilde{\alpha}_i^-(\varepsilon_1) = \frac{\sqrt{k}\left(F\left(\widetilde{c}_i - \frac{\varepsilon_1\|\Phi_0\|_I}{n}\right) - F(\widetilde{c}_i)\right)}{\sqrt{F(\widetilde{c}_i)(1-F(\widetilde{c}_i))}}.$$

Note that $\widetilde{\alpha}_i^-(\varepsilon_1) < 0$. We obtain

$$\begin{aligned} P\left(\xi_k^i \leq F\left(\widetilde{c}_i - \frac{\varepsilon_1\|\Phi_0\|_I}{n}\right)\right) &= 1 - Z(-\widetilde{\alpha}_i^-) \\ \geq \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{\widetilde{\alpha}_i^-} + \left(\frac{1}{\alpha_i^-}\right)^3\right) &\exp\left(-\frac{(\widetilde{\alpha}_i^-)^2}{2}\right). \end{aligned} \tag{24}$$

Combining (23) and (24), we obtain the following lower bounds.

Theorem 4: For any $\varepsilon_1 > 0$

$$\begin{aligned} &P(\|\widehat{\theta}_k - \widehat{\theta}\|_1 \geq \varepsilon_1) \\ &\geq \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon_1)} - \left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon_1)}\right)^3\right) e^{-((\widetilde{\alpha}_i^+(\varepsilon_1))^2/2)} \\ &\quad + \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{\widetilde{\alpha}_i^-(\varepsilon_1)} + \left(\frac{1}{\alpha_i^-(\varepsilon_1)}\right)^3\right) e^{-((\widetilde{\alpha}_i^-(\varepsilon_1))^2/2)}. \end{aligned}$$

Furthermore, we also obtain the following corollary with $\varepsilon_1 = \varepsilon + \eta$.

Corollary 2: Setting $\varepsilon_1 = \varepsilon + \eta$ in Theorem 4, we have

$$\begin{aligned} &P(\|\widehat{\theta}_k - \theta\|_1 \geq \varepsilon + \eta) \\ &\geq \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon+\eta)} - \left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon+\eta)}\right)^3\right) e^{-((\widetilde{\alpha}_i^+(\varepsilon+\eta))^2/2)} \\ &\quad + \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{\widetilde{\alpha}_i^-(\varepsilon+\eta)} + \left(\frac{1}{\widetilde{\alpha}_i^-(\varepsilon+\eta)}\right)^3\right) \\ &\qquad e^{-((\widetilde{\alpha}_i^-(\varepsilon+\eta))^2/2)}. \end{aligned}$$

D. Lower Bounds Based on Asymptotic Normality

The idea here is to approximate the underlying distribution by a normal random variable. It is easily seen that $\rho_k^i := \sqrt{k}(\xi_k^i - F(\widetilde{c}_i))/\sqrt{F(\widetilde{c}_i)(1-F(\widetilde{c}_i))}$ converges in distribution to the standard normal random variable. By virtue of the Berry–Esseen estimate [10, Th. 1, p. 542], the following lemma is in force.

Lemma 1: $|P(\rho_k^i \leq Z) - p(z \leq Z)| \leq \delta_k$, where $\Delta_k = O(1/\sqrt{k})$ as $k \to \infty$ and Z is the standard normal random variable.

Using this lemma, we obtain the following.

Theorem 5: We have the following lower bounds:

$$\begin{aligned} &P(\|\widehat{\theta}_k - \theta\|_1 \geq \varepsilon + \eta) \\ &\geq \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon+\eta)} - \left(\frac{1}{\widetilde{\alpha}_i^+(\varepsilon+\eta)}\right)^3\right) e^{-(\widetilde{\alpha}_i^+(\varepsilon+\eta))^2/2} \\ &\quad + \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{\widetilde{\alpha}_i^-(\varepsilon+\eta)} + \left(\frac{1}{\widetilde{\alpha}_i^-(\varepsilon+\eta)}\right)^3\right) \\ &\quad e^{-\frac{(\widetilde{\alpha}_i^-(\varepsilon+\eta))^2)}{2}} + \Delta_k \end{aligned}$$

where $\Delta_k = O(1/\sqrt{k})$ as $k \to \infty$.

Proof: Note that by Lemma 1

$$\begin{aligned} P\left(\xi_k^i \geq F\left(\widetilde{c}_i + \frac{(\varepsilon+\eta)\|\Phi_0\|_I}{n}\right)\right) \\ = P(\rho_k^i \geq \widetilde{\alpha}_i^+(\varepsilon+\eta)) \geq P(Z \geq \widetilde{\alpha}_i^+(\varepsilon+\eta)) - \Delta_k^+ \end{aligned}$$

where $\Delta_k^+ = O(1/\sqrt{k})$. Similarly

$$\begin{aligned} P\left(\xi_k^i \leq F\left(\widetilde{c}_i - \frac{(\varepsilon+\eta)\|\Phi_0\|_I}{n}\right)\right) \\ = P(\rho_k^i \leq \widetilde{\alpha}_i^-(\varepsilon+\eta)) \geq P(Z \leq \widetilde{\alpha}_i^-(\varepsilon+\eta)) - \Delta_k^- \end{aligned}$$

where $\Delta_k^- = O(1/\sqrt{k})$. Define $\Delta_k = -(\Delta_k^- + \Delta_k^+)$. Using the estimates of lower bounds as in Theorem 4 for the normal random variable Z, the desired result then follows. □

E. Dithers

When the disturbance has a finite support, i.e., the density $f_d(x) = 0, x < -\kappa$ or $x > \kappa$ with a finite κ, the corresponding $F(x)$ is not invertible outside the interval $[-\kappa, \kappa]$. The results in this section are not applicable if $\widetilde{c}_i \notin [-\kappa, \kappa]$ for some i. Consequently, the identification capability of the binary sensor will be reduced. In other words, it is possible that for a selected input, $s(t)$ is a constant (0 or 1) for all t, hence no information is obtained in observations.

One possible remedy for this situation is to add a dither to the sensor input. Hence, assume the disturbance d contains two parts: $d(t) = d_0(t) + h(t)$, where $d_0(t)$ is an i.i.d. disturbance with density f_0 and $h(t)$ is an i.i.d. stochastic dither, independent of d_0, with density f_h. In this case, the density f_d of d is the convolution: $f_d = f_0 * f_h$. By choosing an appropriate f_h, f_d will have a larger support and possess the desired properties for system identification.

IV. Deterministic Frameworks

This section will focus on deterministic representation of the disturbance. Since some results in this section will be valid under any l^p norm, the following assumption is given in general l^p norm. The norm will be further specified if certain results are valid only for some p values.

Assumption A2): For a fixed $p \geq 1$, to be specified later

1) the unmodeled dynamics $\widetilde{\theta}$ is bounded in the l^p norm by $||\widetilde{\theta}||_p \leq \eta$;
2) the disturbance d is uniformly bounded in the l^∞ norm by $||d||_\infty \leq \delta$;
3) the prior information on θ is given by $\Omega_0 = \mathrm{Ball}_p(\theta_0, \varepsilon_0) \subset \mathbb{R}^n$ for some known $\theta_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$.

For a selected input sequence u, let $s = \{s(t),\ t = t_0 + 1, \ldots, t_0 + N\}$ be the observed output. Define

$$\begin{aligned}\Omega_N(t_0, u, s) = &\\ \{\theta : s(t) = &I_{\{\phi^T(t)\theta + \widetilde{\phi}^T(t)\widetilde{\theta} + d(t) \leq C\}} \text{ for some}\\ &||\widetilde{\theta}||_p \leq \eta, ||d||_\infty \leq \delta \text{ and } t = t_0 + 1, \ldots, t_0 + N\}\end{aligned}$$

and

$$e_N = \inf_{\|u\|_\infty \leq u_{\max}} \sup_{t_0} \sup_{s} \mathrm{rad}_p \left(\Omega_N(t_0, u, s) \cap \mathrm{Ball}_p(\theta_0, \varepsilon_0)\right)$$

where rad_p is the radius of the set in l^p norm. e_N is the optimal worst-case uncertainty after N steps of observations. For a given desired identification accuracy ε, the time complexity of $\mathrm{Ball}_p(\theta_0, \varepsilon_0)$ is defined as

$$N(\varepsilon) = \min\{N : e_N \leq \varepsilon\}.$$

We will characterize e_N, determine optimal or suboptimal inputs u, and derive bounds on time complexity $N(\varepsilon)$.

A. Lower Bounds on Identification Errors and Time Complexity

We will show in this subsection that identification time complexity is bounded below by the Kolmogorov entropy of the prior uncertainty set.

Case 1: Disturbance-Free Observations and No Unmodeled Dynamics:

Theorem 6: Let $\delta = 0$ and $\eta = 0$. Suppose that for a given $p \geq 1$ the prior uncertainty $\Omega_0 = \mathrm{Ball}_p(\theta_0, \varepsilon_0)$. Then, for any $\varepsilon < \varepsilon_0$, the time complexity $N(\varepsilon)$ is bounded below by $N(\varepsilon) \geq n \log(\varepsilon_0/\varepsilon)$.

Proof: $\mathrm{Ball}_p(c, \varepsilon)$ in $\mathbb{R}^n$ has volume $a_{p,n}\varepsilon^n$, where the coefficient $a_{p,n}$ is independent of ε. To reduce the identification error from ε_0 to below ε, the volume reduction must be at least $a_{p,n}\varepsilon^n / a_{p,n}\varepsilon_0^n = (\varepsilon/\varepsilon_0)^n$.

Each binary sensor observation defines a hyperplane in the parameter space $\mathbb{R}^n$. The hyperplane divides an uncertainty set into two subsets, with the larger subset having volume at least half of the volume of the original set. As a result, in a worst-case scenario one binary observation can reduce the volume of a set by $1/2$ at best. Hence, the number N of observations required to achieve the required error reduction is at least $(1/2)^N \leq (\varepsilon/\varepsilon_0)^n$, or $N \geq n \log(\varepsilon_0/\varepsilon)$. □

It is noted that $n \log \varepsilon_0/\varepsilon$ is precisely the Kolmogorov ε-entropy of the prior uncertainty set Ω_0 [17], [43]. Hence, Theorem 6 provides an interesting new interpretation of the Kolmogorov entropy in system identification, beyond its application in characterizing model complexity [43]. Theorem 6 establishes a lower bound of exponential rates of time complexity. Upon obtaining an upper bound of the same rates in the next subsection, we will show that the Kolmogorov ε-entropy indeed defines the time complexity rates in this problem. Next, we present an identifiability result, which is limited to $p = 1$.

Proposition 1: The uncertainty set $\mathrm{Ball}_1(0, C/u_{\max})$ is not identifiable.

Proof: For any $\theta \in \mathrm{Ball}_1(0, C/u_{\max})$, the output

$$\begin{aligned}y(t) =& \phi^T(t)\theta\\ \leq& ||\phi(t)||_\infty ||\theta||_1\\ \leq& u_{\max} \times \frac{C}{u_{\max}} = C.\end{aligned}$$

It follows that $s(t) = 1, \forall t$. Hence, the observations could not provide further information to reduce uncertainty. □

Case 2: Complexity Impact of Bounded Disturbances: In the case of noisy observations, the input–output relationship becomes

$$y(t) = \phi^T(t)\theta + d(t) \quad s(t) = I_{\{y(t) \leq C\}} \tag{25}$$

where $|d(t)| \leq \delta$. For any given $\phi^T(t)$, an observation on $s(t)$ from (25) defines, in a worst-case sense, two possible uncertainty half planes

$$\begin{aligned}\Omega_1 =& \{\theta \in \mathbb{R}^n : \quad \phi^T(t)\theta \leq C + \delta\}, \qquad s(t) = 1\\ \Omega_2 =& \{\theta \in \mathbb{R}^n : \quad \phi^T(t)\theta > C - \delta\}, \qquad s(t) = 0.\end{aligned}$$

Uncertainty reduction via observation is possible only if the uncertainty set before observation is not a subset of each half plane (so that the intersection of the uncertainty set and the half plane results in a smaller set).

Theorem 7: If $\varepsilon \leq \delta/u_{\max}$, then for any $\theta_0 \in \mathbb{R}^n$ either $\mathrm{Ball}_1(\theta_0,\varepsilon) \subseteq \Omega_1$ or $\mathrm{Ball}_1(\theta_0,\varepsilon) \subseteq \Omega_2$. Consequently, in a worst-case sense $\mathrm{Ball}_1(\theta_0,\varepsilon)$ is not identifiable.

Proof: Suppose that $\mathrm{Ball}_1(\theta_0,\varepsilon) \not\subseteq \Omega_1$. Then, there exists $\theta_1 \in \mathrm{Ball}_1(\theta_0,\varepsilon)$ such that $\phi^T(t)\theta_1 > C+\delta$. $\theta \in \mathrm{Ball}_1(\theta_0,\varepsilon)$ satisfies $||\theta-\theta_1||_1 \leq 2\varepsilon$. We have

$$\begin{aligned}\phi^T(t)\theta =& \phi^T(t)\theta_1 + \phi^T(t)(\theta-\theta_1)\\ >& C+\delta+\phi^T(t)(\theta-\theta_1)\\ \geq& C+\delta-u_{\max}2\varepsilon\\ \geq& C-\delta\end{aligned}$$

for any $\theta \in \mathrm{Ball}_1(\theta_0,\varepsilon)$. This implies that $\mathrm{Ball}_1(\theta_0,\varepsilon) \subseteq \Omega_2$. The opposite case can be proved similarly. □

Theorem 7 shows that worst-case disturbances introduce irreducible identification errors of size at least $\delta/u_{\max}$. This is a general result. A substantially higher lower bound can be obtained in the special case of $n=1$.

Consider the system $y(t)=au(t)+d$. Suppose that at time t the prior information on a is that $a \in \Omega = [\underline{a},\overline{a}]$ with $\underline{a} > C/u_{\max}$ for identifiability (see Proposition 1). The uncertainty set has center $a_0=(1/2)(\underline{a}+\overline{a})$ and radius $\varepsilon=(1/2)(\overline{a}-\underline{a})$. To minimize the posterior uncertainty in the worst-case sense, the optimal $u(t)$ can be easily obtained as $u(t)=C/a_0$.

Theorem 8: If $\delta < C$, then the uncertainty set $[\underline{a},\overline{a}]$ cannot be reduced if $\varepsilon \leq \delta(u_{\max}/(1-\delta/C))$.

Proof: Let $\varepsilon = \delta/(u_{\max}(1-\delta/C))$. Then, $\delta = \varepsilon C/(C/u_{\max}+\varepsilon)$. For any $a \in [\underline{a},\overline{a}]$, noting $a_0=\underline{a}+\varepsilon$, we have $|a-a_0| \leq \varepsilon$, and

$$\begin{aligned}au(t) =& a\times\frac{C}{a_0} = (a_0+(a-a_0))\times\frac{C}{a_0}\\ =& C+(a-a_0)\times\frac{C}{a_0}\\ \leq& C+\frac{\varepsilon C}{\underline{a}+\varepsilon}\\ <& C+\frac{\varepsilon C}{\frac{C}{u_{\max}}+\varepsilon}\\ =& C+\delta.\end{aligned}$$

Hence, the observation $s(t)=1$ does not provide any information.

Similarly, if $s(t)=0$, we can show that all $\theta \in [\underline{a},\overline{a}]$ will result in $au(t) > C-\delta$. Again, the observation does not reduce uncertainty. □

At present, it remains an open question if Theorem 8 holds for higher order systems.

Case 3: Complexity Impact of Unmodeled Dynamics: When the system contains unmodeled dynamics, the input–output relationship becomes

$$y(t)=\phi^T(t)\theta+\widetilde{\phi}^T(t)\widetilde{\theta} \qquad s(t)=I_{\{y(t)\leq C\}} \tag{26}$$

where $||\widetilde{\theta}||_1 \leq \eta$. We will show that unmodeled dynamics will introduce an irreducible identification error on the modeled part.

For any $\widetilde{\phi}(t)$, the set $\{\widetilde{\phi}^T(t)\widetilde{\theta} : ||\widetilde{\theta}||_1 \leq \eta\} = [-\eta m_t, \eta m_t]$, where $m_t = ||\widetilde{\phi}(t)||_\infty$.

Theorem 9: If $\varepsilon \leq \eta$, then in a worst-case sense, for any θ_0, $\mathrm{Ball}_1(\theta_0,\varepsilon)$ is not identifiable.

Proof: Under (26), an observation on $s(t)$ provides observation information

$$\begin{aligned}\Omega_1 =& \{\theta\in\mathbb{R}^n : \phi^T(t)\theta \leq C+\eta m_t\}, \qquad s(t)=1\\ \Omega_2 =& \{\theta\in\mathbb{R}^n : \phi^T(t)\theta > C-\eta m_t\}, \qquad s(t)=0.\end{aligned}$$

In the worst-case sense, $\mathrm{Ball}_1(\theta_0,\varepsilon)$ can be reduced by this observation only if $\mathrm{Ball}_1(\theta_0,\varepsilon)$ is neither a subset of Ω_1 nor Ω_2.

Suppose that $\mathrm{Ball}_1(\theta_0,\varepsilon) \not\subseteq \Omega_2$. We will show that $\mathrm{Ball}_1(\theta_0,\varepsilon) \subseteq \Omega_1$. Indeed, in this case there exists $\theta_1 \in \mathrm{Ball}_1(\theta_0,\varepsilon)$ such that $\phi^T(t)\theta_1 \leq C-\eta m_t$. Since any $\theta \in \mathrm{Ball}_1(\theta_0,\varepsilon)$ satisfies $||\theta-\theta_1||_1 \leq 2\varepsilon$, we have

$$\begin{aligned}\phi^T(t)\theta =& \phi^T(t)\theta_1+\phi^T(t)(\theta-\theta_1)\\ \leq& C-\eta m_t+\phi^T(t)(\theta-\theta_1)\\ \leq& C-\eta m_t+m_t 2\varepsilon\\ \leq& C+\eta m_t.\end{aligned}$$

This implies $\mathrm{Ball}_1(\theta_0,\varepsilon) \subseteq \Omega_1$. □

B. General Upper Bounds

In this section, general upper bounds on identification errors or time complexity will be established. For a fixed $p \geq 1$, suppose that the prior information on θ is given by $\mathrm{Ball}_p(\theta_0,\varepsilon_0)$. For identifiability, assume that the signs of a_i have been detected and $\underline{a} = \min\{|\underline{a}_i|, i=1,\ldots,n\} > C/u_{\max}$.[3] Denote $\overline{a} = \max_{\theta\in\mathrm{Ball}_p(\theta_0,\varepsilon_0)} ||\theta||_\infty$. We will establish upper bounds on the time complexity $N(\varepsilon)$ for reducing the size of uncertainty from ε_0 to ε, in l^p norm.

Case 1: Disturbance-Free Observations and No Unmodeled Dynamics: Let $\eta=0$ and $\delta=0$ and consider $y(t)=\phi^T(t)\theta$.

Theorem 10: Suppose that $u_{\max} > C/\underline{a}$. Then, the time complexity to reduce uncertainty from ε_0 to ε is bounded by

$$N(\varepsilon) \leq (n^2-n+1)\left\lceil \frac{1}{p}\log n + \log\frac{\varepsilon_0}{\varepsilon}\right\rceil. \tag{27}$$

Since n is a constant independent of N, this result, together with Theorem 6, confirms that the Kolmogorov entropy defines the time complexity rates in binary-sensor identification. The accurate calculation for $N(\varepsilon)$ remains an open and difficult question, except for $n=1$ (gain uncertainty) which is discussed in the next subsection.

The proof of Theorem 10 utilizes the following lemma. Consider the first-order system $y(t)=au(t)$, $s(t)=I_{\{y(t)\leq C\}}$, where $a \in [\underline{a},\overline{a}]$ and $\underline{a} > C/u_{\max} > 0$. Let $\varepsilon_0=(1/2)(\overline{a}-\underline{a})$.

[3]The sign of a_i can be obtained easily by choosing an initial testing sequence of u. Also, those parameters with $|\underline{a}_i| < C/u_{\max}$ can be easily detected. Since uncertainty on these parameters cannot be further reduced (Proposition 1), they will be left as remaining uncertainty. $\underline{a}$ defined here will be applied to the rest of the parameters. The detail is omitted for brevity.

Lemma 2: There exists an input sequence u such that N observations on $s(t)$ can reduce the radius of uncertainty to $\varepsilon = 2^{-N}\varepsilon_0$.

Proof: Let $[\underline{a}(t), \overline{a}(t)]$ be the prior uncertainty before a measurement on $s(t)$. Then, $\varepsilon(t) = (1/2)(\overline{a}(t) - \underline{a}(t))$. By choosing $u(t) = C/(\underline{a}(t) + \varepsilon(t))$, the observation on $s(t)$ will determine uniquely either $a \in [\underline{a}(t), \underline{a}(t) + \varepsilon(t)]$ if $s(t) = 1$; or $a \in [\overline{a}(t) - \varepsilon(t), \overline{a}(t)]$ if $s(t) = 0$. In either case, the uncertainty will be reduced by $1/2$. Iteration on the number of observations leads to the conclusion. □

The proofs of this subsection rely on the following idea. Choose $u(t) = 0$ except those with index $j(n^2 - n + 1) + i, i = 1, n + 1, \ldots, (n - 1)n - n + 3, j = 0, 1, \ldots$. This input design results in a specific input–output relationship

$$\begin{aligned} & y\left(j(n^2 - n + 1) + n\right) \\ =& a_{n-1} u\left(j(n^2 - n + 1) + 1\right) \\ & y\left(j(n^2 - n + 1) + n + 1\right) \\ =& a_0 u\left(j(n^2 - n + 1) + n + 1\right) \\ & \vdots \\ & y\left(j(n^2 - n + 1) + (n - 1)n + 1\right) \\ =& a_{n-2} u\left(j(n^2 - n + 1) + (n - 1)n - n + 3\right). \end{aligned} \quad (28)$$

In other words, within each block of $n^2 - n + 1$ observations, each model parameter can be identified individually once. Less conservative inputs can be designed. However, they are more problem dependent and *ad hoc*, and will not be presented here.

Proof of Theorem 10: By Lemma 2, uncertainty radius on each parameter can be reduced by a factor 2^{-N_1} after N_1 observations. This implies that by using the input (28), after $N = (n^2 - n + 1)N_1$ observations, the uncertainty radius can be reduced to

$$\begin{aligned} \mathrm{rad}_p(\Omega_N) \le& n^{1/p}\mathrm{rad}_\infty(\Omega_N) \\ \le& n^{1/p}2^{-N/(n^2-n+1)}\mathrm{rad}_\infty(\Omega_0) \\ \le& n^{1/p}2^{-N/(n^2-n+1)}\mathrm{rad}_p(\Omega_0) \\ =& n^{1/p}2^{-N/(n^2-n+1)}\varepsilon_0. \end{aligned}$$

Hence, for $n^{1/p}2^{-N/(n^2-n+1)}\varepsilon_0 \le \varepsilon$ it suffices to have

$$N = (n^2 - n + 1)\left\lceil \frac{1}{p}\log n + \log\frac{\varepsilon_0}{\varepsilon} \right\rceil.$$

□

Case 2: Noisy Observations: Consider $y(t) = \phi^T(t)\theta + d(t)$, where $|d(t)| \le \delta$.

Theorem 11: Suppose $\delta < C$. Let $\rho = (1/2)(1 - \delta/C)$ and $\sigma = \delta\overline{a}/(2C(1 - \rho)) = \overline{a}\delta/(C(1 + \delta/C))$. If $\varepsilon_0 > \varepsilon > \sigma$ and $u_{\max} > C/\underline{a}$, the time complexity $N(\varepsilon)$ for reducing uncertainty from ε_0 to ε is bounded in l^p norm by

$$N(\varepsilon) \le (n^2 - n + 1)\left\lceil \frac{1}{p}\log n + \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho} \right\rceil. \quad (29)$$

Proof: Using the input in (28), the identification of the n parameters $a_0, \ldots, a_{n-1}$ is reduced to identifying each parameter individually. Now for identification of a single parameter $y(t) = au(t) + d(t)$, we can derive the following iterative uncertainty reduction relationship. If the prior uncertainty at t is $[a(t) - \varepsilon(t), a(t) + \varepsilon(t)]$, then the optimal worst-case input $u(t)$ can be shown as $u(t) = C/a(t)$.[4] The posterior uncertainty will be either $[a(t) - \varepsilon(t), a(t) + (\delta/C)a(t)]$, if $s(t) = 1$; or $[a(t) - (\delta/C)a(t), a(t) + \varepsilon(t)]$, if $s(t) = 0$. Both have the radius

$$\begin{aligned} \varepsilon(t+1) =& \frac{1}{2}\left(\varepsilon(t) + \frac{\delta}{C}a(t)\right) \\ =& \frac{1}{2}\left(1 - \frac{\delta}{C}\right)\varepsilon(t) + \frac{1}{2}\frac{\delta}{C}(a(t) + \varepsilon(t)) \\ \le& \rho\varepsilon(t) + \frac{\delta\overline{a}}{2C}. \end{aligned}$$

Starting from ε_0, after N_1 observations, we have

$$\begin{aligned} \varepsilon(N_1) \le& \rho^{N_1}\varepsilon_0 + \frac{\delta\overline{a}}{2C}\sum_{i=0}^{N_1-1}\rho^i \\ =& \rho^{N_1}\varepsilon_0 + \frac{\delta\overline{a}}{2C}\frac{1 - \rho^{N_1}}{1 - \rho} \\ =& \rho^{N_1}\varepsilon_0 + \sigma(1 - \rho^{N_1}) \\ =& \rho^{N_1}(\varepsilon_0 - \sigma) + \sigma. \end{aligned}$$

To achieve $\varepsilon(N_1) \le \varepsilon$, it suffices

$$\rho^{N_1}(\varepsilon_0 - \sigma) + \sigma \le \varepsilon \quad \text{or} \quad N_1 \ge \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho}.$$

Following the same arguments as in the proof of Theorem 10, we conclude that

$$N = (n^2 - n + 1)\left\lceil \frac{1}{p}\log n + \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho} \right\rceil$$

will suffice to reduce the uncertainty from ε_0 to ε in l^p norm. □

Case 3: Unmodeled Dynamics: Consider $y(t) = \phi^T(t)\theta + \widetilde{\phi}^T(t)\widetilde{\theta}$. The results of this case hold for $p = 1$ only. The unmodeled dynamics introduces an uncertainty on the observation on $y(t)$: $\{\widetilde{\phi}^T(t)\widetilde{\theta} : \|\widetilde{\theta}\|_1 \le \eta\} = [-\eta m_t, \eta m_t]$, $m_t = \|\phi(t)\|_\infty$.

Theorem 12: Suppose $0 < \eta < C/u_{max}$. Let $\rho_1 = (1/2)(1 - \eta u_{\max}/C)$, $\sigma_1 = \eta u_{\max}\overline{a}/(2C(1 - \rho_1))$.

$$N(\varepsilon) \le (n^2 - n + 1)\left\lceil \log n + \frac{\log\frac{\varepsilon-\sigma_1}{\varepsilon_0-\sigma_1}}{\log\rho_1} \right\rceil. \quad (30)$$

Proof: By using the input (28), the identification of θ is reduced to each of its components. For a scalar system $y(t) = au(t) + \widetilde{\phi}^T(t)\widetilde{\theta}$, since $|\widetilde{\phi}^T(t)\widetilde{\theta}| \le \eta u_{\max}$ we can apply Theorem 11 with δ replaced by $\eta u_{\max}$. Inequality (30) then follows from Theorem 11. □

C. Special Cases: Identification of Gains

In the special case $n = 1$, explicit results and tighter bounds can be obtained. When $n = 1$, the observation equation becomes

$$y(t) = au(t) + \widetilde{\phi}^T(t)\widetilde{\theta} + d(t).$$

[4]More detailed derivations are given in the next subsection.

Assume that the initial information on a is that $\underline{a}(0) \le a \le \overline{a}(0), \underline{a}(0) \neq 0, \overline{a}(0) \neq 0$, with radius $\varepsilon_0 = (1/2)(\overline{a}(0) - \underline{a}(0))$.

Case 1: $y(t) = au(t)$. It is noted that this is a trivial identification problem when regular sensors are used: After one input $u(0) \neq 0$, a can be identified uniquely.

Theorem 13:

1) Suppose the sign of a is known, say, $\underline{a}(0) > 0$, and $u_{\max} \ge C/\underline{a}(0)$. Then, the optimal identification error is $e_N = 2^{-N} e_0$ and the time complexity is $N(\varepsilon) = \lceil \log \varepsilon_0/\varepsilon \rceil$.

   If at $t-1$, the information on a is that $a \in [\underline{a}(t-1), \overline{a}(t-1)]$, then the optimal $u(t)$ is

$$u(t) = \frac{2C}{\underline{a}(t-1) + \overline{a}(t-1)} \tag{31}$$

   where $\underline{a}(t)$ and $\overline{a}(t)$ are updated by

$$\underline{a}(t) = \begin{cases} \frac{1}{2}(\underline{a}(t-1) + \overline{a}(t-1)), & \text{if } s(t) = 0 \\ \underline{a}(t-1), & \text{if } s(t) = 1 \end{cases}$$
$$\overline{a}(t) = \begin{cases} \overline{a}(t-1), & \text{if } s(t) = 0 \\ \frac{1}{2}(\underline{a}(t-1) + \overline{a}(t-1)), & \text{if } s(t) = 1 \end{cases}.$$

2) If $\underline{a}(0)$ and $\overline{a}(0)$ have opposite signs and

$$\delta_l = \max\{\underline{a}(0), -\frac{C}{u_{\max}}\} \quad \delta_h = \min\{\overline{a}(0), \frac{C}{u_{\max}}\}$$

   then the uncertainty interval (δ_l, δ_h) is not identifiable. Furthermore, in the case of $\underline{a}(0) \le \delta_l$ and $\overline{a}(0) \ge \delta_h$, if $\delta_h - \delta_l \le \varepsilon$ and $\varepsilon_0 \ge 2\varepsilon$ then the time complexity $N(\varepsilon)$ is bounded by

$$\left\lceil \log \frac{\varepsilon_0}{\varepsilon} \right\rceil \le N(\varepsilon) \le \left\lceil \log \frac{\varepsilon_0 - (\delta_h - \delta_l)}{\varepsilon} \right\rceil + 2.$$

Proof: In Appendix.

In this special case, the actual value $C > 0$ does not affect identification accuracy. This is due to noise-free observation. The value C will become essential in deriving optimal identification errors when observation noises are present. $C = 0$ is a singular case in which uncertainty on a cannot be reduced (in the sense of the worst-case scenario). Indeed, in this case one can only test the sign of a. It is also observed that the optimal $u(t)$ depends on the previous observation $s(t-1)$. As a result, $u(t)$ can be constructed causally and sequentially, but not offline.

Case 2: $y(t) = au(t) + d(t)$. Here, we assume $|d(t)| \le \delta < C$. Prior information on a is given by $a \in \Omega_0 = [\underline{a}(0), \overline{a}(0)]$, and $\underline{a}(0) > 0$.

Theorem 14: Suppose that $u_{\max} \ge C/\underline{a}(0)$ and $\overline{a}(0)/\underline{a}(0) \ge (1+\delta/C)/(1-\delta/C)$. Then

1) the optimal input $u(t)$ is given by the causal mapping from the available information at $t-1$

$$u(t) = \frac{2C}{\underline{a}(t-1) + \overline{a}(t-1)}.$$

   The optimal identification error satisfies the iteration equation

$$e(t) = \frac{1}{2} e(t-1) + \frac{1}{2}\frac{\delta}{C}(\overline{a}(t-1) + \underline{a}(t-1)) \tag{32}$$

   where $\overline{a}(t)$ and $\underline{a}(t)$ are updated by the rules

$$\overline{a}(t) = \overline{a}(t-1),\ \underline{a}(t) = \frac{C-\delta}{u(t)},\ \text{if } s(t) = 0$$
$$\underline{a}(t) = \underline{a}(t-1),\ \overline{a}(t) = \frac{C+\delta}{u(t)},\ \text{if } s(t) = 1.$$

2) $\overline{a}(t)/\underline{a}(t) \ge (1+\delta/C)/(1-\delta/C)$ for all $t \ge 1$; $\{\underline{a}(t)\}$ is monotone increasing, $\{\overline{a}(t)\}$ and $\{\overline{a}(t)/\underline{a}(t)\}$ are monotone decreasing; $\lim_{t\to\infty} \overline{a}(t)/\underline{a}(t) = (1+\delta/C)/(1-\delta/C)$.
3) At each time t, uncertainty reduction is possible if and only if $\overline{a}(t-1)/\underline{a}(t-1) > (1+\delta/C)/(1-\delta/C)$.

Proof: In the Appendix.

Theorem 15: Let $\alpha_1 = (1/2)(1-\delta/C)$ and $\alpha_2 = (1/2)(1+\delta/C)$. Then, under the conditions and notation of Theorem 14

1) for $t \ge 1$, the optimal identification error $e(t)$ is bounded by

$$\begin{aligned} &\alpha_1^t e(0) + \frac{\delta a(1-\alpha_1^t)}{C\alpha_2} \\ \le &\alpha_1^t e(0) + \frac{\delta \overline{a}(t-1)(1-\alpha_1^t)}{C\alpha_2} \\ \le &e(t) \le \alpha_2^t e(0) + \frac{\delta \underline{a}(t-1)(1-\alpha_2^t)}{C\alpha_1} \\ \le &\alpha_2^t e(0) + \frac{\delta a(1-\alpha_2^t)}{C\alpha_1}. \end{aligned} \tag{33}$$

2) Let $\varepsilon_0 = e(0)$ and $\varepsilon_0 > \varepsilon > \delta a/(C\alpha_1) = 2\delta a/(C-\delta)$. Then the time complexity $N(\varepsilon)$ for reducing uncertainty from ε_0 to ε is bounded by

$$\left\lceil \frac{\log \frac{\varepsilon - \frac{\delta a}{C\alpha_2}}{\varepsilon_0 - \frac{\delta a}{C\alpha_2}}}{\log \alpha_1} \right\rceil \le N \le \left\lceil \frac{\log \frac{\varepsilon - \frac{\delta a}{C\alpha_1}}{\varepsilon_0 - \frac{\delta a}{C\alpha_1}}}{\log \alpha_2} \right\rceil.$$

3) There exists an irreducible relative error

$$\frac{2\frac{\delta}{C}}{1+\frac{\delta}{C}} \le \frac{e(\infty)}{a} \le \frac{2\frac{\delta}{C}}{1-\frac{\delta}{C}}. \tag{34}$$

4) The parameter estimation error is bounded by

$$0 \le \frac{\overline{a}(\infty) - a}{\overline{a}(\infty)} \le \frac{2\frac{\delta}{C}}{1+\frac{\delta}{C}},\ 0 \le \frac{a - \underline{a}(\infty)}{\underline{a}(\infty)} \le \frac{2\frac{\delta}{C}}{1-\frac{\delta}{C}}. \tag{35}$$

Proof: In the Appendix.

Remark 5: It is noted that the bounds in item 2) of Theorem 15 can be easily translated to a sequential information bounds by replacing a with the online inequalities $\underline{a}(t-1) \le a \le \overline{a}(t-1)$.

Case 3: $y(t) = au(t) + \widetilde{\phi}^T(t)\widetilde{\theta}$. Let $u_t = \{u(\tau),\ \tau \le t\}$. Then, $\|u_t\|_\infty$ is the maximum $|u(\tau)|$ up to time t. Since we assume no information on $\widetilde{\theta}$, except that $\|\widetilde{\theta}\|_1 \le \eta$, it is clear that $\sup_{\|\widetilde{\theta}\|_1 \le \eta} |\widetilde{\phi}^T(t)\widetilde{\theta}| = \eta m_t$ where $m_t = \|\widetilde{\phi}(t)\|_\infty$.

Let $w(t) = \widetilde{\phi}^T(t)\widetilde{\theta}$. Then $\{\widetilde{\phi}^T(t)\widetilde{\theta} : \|\widetilde{\theta}\|_1 \le \eta\} = \{w(t) : |w(t)| \le \eta m_t\}$.

Theorem 16: Suppose that $\underline{a}(0) > 0$, $u_{\max} \ge C/\underline{a}(0)$, $\eta < \underline{a}(0)$.

1) The optimal input $u(t)$, which minimizes the worst-case uncertainty at t, is given by the causal mapping from the available information at $t-1$

$$u(t)=\frac{2C}{\underline{a}(t-1)+\overline{a}(t-1)}. \tag{36}$$

The optimal identification error at t satisfies the iteration equation

$$e(t)=\frac{1}{2}e(t-1)+\frac{\eta m_t}{2C}(\overline{a}(t-1)+\underline{a}(t-1)) \tag{37}$$

where $\overline{a}(t)$ and $\underline{a}(t)$ are updated by the rules

$$\overline{a}(t)=\overline{a}(t-1),\ \underline{a}(t)=\frac{C-\eta m_t}{u(t)},\ \text{if } s(t)=1$$
$$\underline{a}(t)=\underline{a}(t-1),\ \overline{a}(t)=\frac{C+\eta m_t}{u(t)},\ \text{if } s(t)=0.$$

2) The uncertainty is reducible if and only if $\overline{a}(t-1)>\underline{a}(t-1)+2\eta$.
3) For $t\geq 1$, the optimal identification error $e(t)$ is bounded by

$$\left(\prod_{j=1}^{t}\beta_1(j)\right)e(0)+\frac{\eta a}{C}\sum_{i=1}^{t}m_i\prod_{j=i+1}^{t}\beta_1(j)\leq e(t)\leq\left(\prod_{j=1}^{t}\beta_2(j)\right)e(0)+\frac{\eta a}{C}\sum_{i=1}^{t}m_i\prod_{j=i+1}^{t}\beta_2(j) \tag{38}$$

where $\beta_1(t)=(1/2)(1-\eta m_t/C)$ and $\beta_2(t)=(1/2)(1+\eta m_t/C)$.
4) Let $\varepsilon_0=e(0)$ and $\varepsilon_0>\varepsilon>2\eta\overline{a}(0)/(\underline{a}(0)-\eta)$. Also, denote $\beta_1=(1/2)(1-\eta/\underline{a}(0))$, $\beta_2=(1/2)(1+\eta/\underline{a}(0))$. Then, the time complexity $N(\varepsilon)$ for reducing uncertainty from ε_0 to ε is bounded by

$$\left\lceil\frac{\log\frac{\varepsilon-\frac{\eta a}{\overline{a}(0)\beta_2}}{\varepsilon_0-\frac{\eta a}{\overline{a}(0)\beta_2}}}{\log\beta_1}\right\rceil\leq N(\varepsilon)\leq\left\lceil\frac{\log\frac{\varepsilon-\frac{\eta a}{\underline{a}(0)\beta_1}}{\varepsilon_0-\frac{\eta a}{\underline{a}(0)\beta_1}}}{\log\beta_2}\right\rceil \tag{39}$$

Proof: In the Appendix.

Note that $\beta_2(t)\geq\beta_1(t)$ and $\beta_1(t)+\beta_2(t)=1$; and $\beta_1\to\beta_2$ as $\eta\to 0$, uniformly in t.

V. Discussions on Combined Stochastic and Deterministic Binary-Sensor Identification

The theoretical development of this paper highlights the distinctive underlying principles used in designing inputs and deriving posterior uncertainty sets in stochastic and deterministic information frameworks.

In the deterministic worst-case framework, the information on noise is limited to its magnitude bound. Identification properties must be evaluated against worst-case noise sample paths. As shown earlier, the optimal input is obtained on the basis of choosing an optimal worst-case testing point (a hyperplane) for the prior uncertainty set. When the prior uncertainty set is large, this leads to a very fast exponential rate of uncertainty reduction. However, when the uncertainty set is close to its irreducible limits due to disturbances or unmodeled dynamics, its rate of uncertainty reduction decreases dramatically due to its worst-case requirements. Furthermore, when the disturbance magnitude is large, the irreducible uncertainty will become too large for identification error bounds to be practically useful.

In contrast, in a stochastic framework, noise is modeled by a stochastic process and identification errors are required to be small with a large probability. Binary-sensor identification in this case relies on the idea of averaging. Typically, in stochastic identification the input is designed to provide sufficient excitation for asymptotic convergence, rather than fast initial uncertainty reduction. Without effective utilization of prior information in designing the input during the initial time interval, initial convergence can be very slow. This is especially a severe problem in binary-sensor identification since a poorly designed input may result in a very imbalanced output of the sensor in its 0 or 1 values, leading to slow convergence rate. In the case of large prior uncertainty, the selected input may result in nonswitching at the output, rendering the stochastic binary-sensor identification inapplicable. On the other hand, averaging disturbances restores estimate consistency and overcomes a fundamental limitation of the worst-case identification.

Consequently, it seems a sensible choice of using the deterministic framework initially to achieve fast uncertainty reduction when the uncertainty set is large, then using the stochastic framework to modify estimation consistency. In fact, we shall demonstrate by an example that these two frameworks complement each other precisely, in the sense that when one framework fails the other starts to be applicable. Consider the first-order system $y(t)=au(t)+d(t)$, where $d(t)$ is i.i.d. but with support on $[-\delta,\delta]$. Suppose that the prior information on a is given by $\Omega_0=[a_0-\varepsilon,a_0+\varepsilon]$, with $a_0-\varepsilon>0$.

First, we will show that if ε is large, then some subsets of Ω_0 cannot be identified by the stochastic averaging approach. More precisely, we note that the stochastic averaging method requires that one select a constant $u(t)=u_0$ such that the following condition is satisfied: $P(y(t)>C)>0$ and $P(y(t)\leq C)>0$. Under this condition, the distribution function F is invertible at the convergent point of the empirical distribution $\widehat{F}_k(x)$. Consequently, The results of Section III can be applied to identify a.

However, if $\varepsilon>a_0\delta/C$, then for any choice of u_0, either $P(y(t)>C)=0$ or $P(y(t)\leq C)=0$, that is, the above condition is always violated. Indeed, suppose that $P(y(t)>C)>0$, for all possible $a\in\Omega_0$. In particular, this implies that for $a=a_0-a_0\delta/C\in\Omega_0$, we have $(a_0-a_0\delta/C)u_0+\delta>C$ or equivalently $a_0u_0>C$.

Now, consider the subset $\Omega_1=[a_0+a_0\delta/C,a_0+\varepsilon]\subset\Omega_0$. For any $a\in\Omega_1$, we have

$$au_0+d(t)\geq au_0-\delta\geq\left(a_0+\frac{a_0\delta}{C}\right)u_0-\delta>\left(1+\frac{\delta}{C}\right)C-\delta=C.$$

This implies that $P(y(t)\leq C)=0$.

On the other hand, if we apply the deterministic identification first to reduce uncertainty on a first, by Section IV, the uncertainty can be precisely reduced to $\Omega_2 = [a_0 - a_0\delta/C, a_0 + a_0\delta/C]$. It is easy to show that for Ω_2 the stochastic binary-sensor identification is applicable since we have

$$\text{If } P(y(t) \le C) = 1, \text{ then } a = a_0 - \frac{a_0\delta}{C}$$

$$\text{If } P(y(t) > C) = 1, \text{ then } a = a_0 + \frac{a_0\delta}{C}$$

$$\text{Otherwise, } P(y(t) > C) > 0 \text{ or } P(y(t) \le C) > 0.$$

The following numerical example is devised to further illustrate these ideas. Consider the system $y(t) = au(t) + d(t)$, where $|d(t)| \le \delta$ with a uniform distribution in $[-\delta, \delta]$. The true value $a = 200$. Suppose that the threshold $C = 50$, disturbance bound $\delta = 10$, and prior information on a is that $a \in [1{,}1000]$. Deterministic identification starts with a fast uncertainty reduction, but settles to a final irreducible uncertainty set [167.6, 251.4], as shown in of Fig. 1(a).

On the other hand, if one elects to use stochastic framework, it is critical to find an input value that will cause the sensor to switch. The large prior uncertainty on a makes it difficult to find such an input. For instance, $a \in [1, 1000]$ and $C = 50$ imply possible values of u in [0.05,50]. A sample of 10 randomly selected values in [0.05,50] gives 20.8123, 47.1245, 0.5278, 19.4371, 18.7313, 0.5676, 25.4479,16.9107, 9.1140, 44.1065, all of them fail to be a viable input ($y(t) = 200u - 10 > 50$ for all t).

Next, we combine deterministic and stochastic approaches. First, the deterministic approach is used to reduce the uncertainty set to, say, [165, 255]. This is achieved after ten observations. We then switch to the stochastic framework. Select

$$u = \frac{C}{0.5(255 + 165)} = 0.2381.$$

This leads to $au = 200 \times 0.2381 = 47.62$, which satisfies the condition of stochastic binary-sensor identification (invertibility of F). Upon changing to stochastic identification, the output $s(t)$ is observed. The estimate on a is calculated by (9). The trajectory of the estimate is shown in Fig. 1(b).

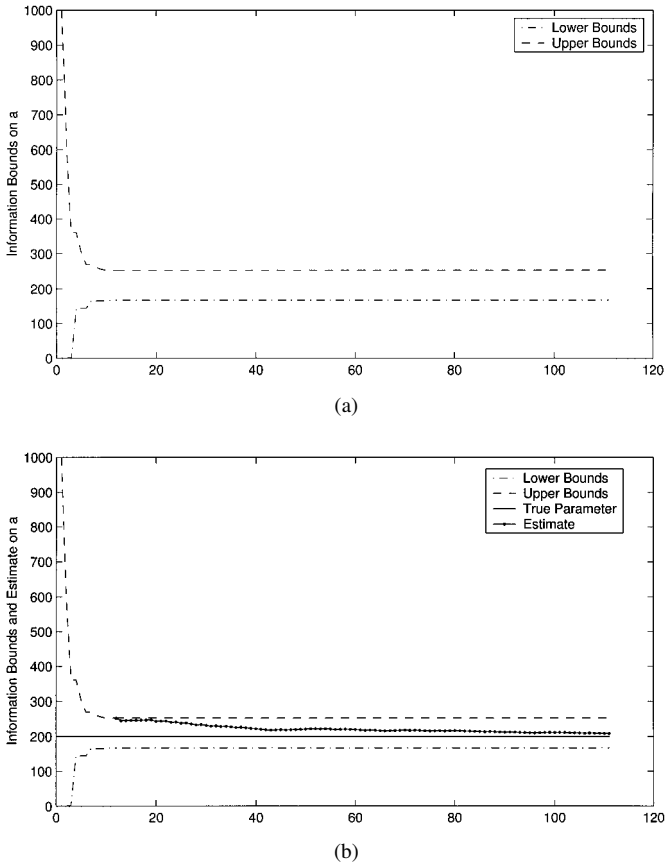


Fig. 1. Comparison of stochastic and deterministic frameworks. (a) Deterministic identification. (b) Combined identification.

VI. Illustrative Examples

In this section, we will use two examples to illustrate how the algorithms developed in this paper can be applied to address the motivating issues discussed in Section I. In example 1, we will show that by using binary sensors for identification one can achieve output tracking for reference set points that are different from the sensor switching point. Example 2 demonstrates that the common practice in industry applications, in which two binary sensors are used to force a controlled variable in the set bounds, does not impose additional difficulties in applying our results to output tracking control. Since online identification (especially persistent identification problems in which identification is needed beyond its initial parameter convergence) is only meaningful when system parameters drift slowly but substantially from their initial values, we use a slowly varying system to demonstrate our methods in Example 2.

Example 1: **Tracking Control Using One Binary Sensor.** Suppose that the goal of control is to set $y(t) = y^*$, where y^* is a desired reference point. A binary sensor is deployed with a threshold C. Traditional control in this problem is to design a feedback control that will maintain $y(t)$ close to C. However, if the sensor threshold is not equal to the target: $C \neq y^*$, the traditional feedback will fail to drive y to the target y^*. Using the identification approaches to estimate system parameters first, however, one can potentially control $y(t)$ to a small range around y^* after identification.

Let the true system be[5]

$$y(t) = a_1u(t) + a_2u(t-1) + d(t)$$

with $a_1 = 12$ and $a_2 = 5$. The disturbance d is i.i.d. with uniform distribution in $[-1, 1]$. The target output is $y^* = 30$. Suppose that the sensor has threshold $C = 20$. $u_{\max} = 40$. Let the prior information on the parameters be $a_i \in [1, 51], i = 1, 2$.

By the input (28) with $n = 2$, we have index $3j + i$, $i = 1$ or 3, $j = 0, 1, \ldots$. Hence, the input sequence is $u = \{0, u(1), 0, u(3), u(4), 0, u(6), u(7), 0, u(9), \ldots\}$. The corresponding input–output relation becomes

$$y(3j+2) = a_2u(3j+1), \qquad j = 0, 1, \ldots$$
$$y(3j+3) = a_1u(3j+3), \qquad j = 0, 1, \ldots.$$

By choosing $u(3j+1)$ and $u(3j+3)$ optimally for identification of individual parameters, we can reduce parameter uncertainty first to, say, a radius of 0.5 on each parameter.

[5]For simplicity, we use a minimum-phase system for this example. As a result, after identification, tracking control can be designed by simple inversion. In the case of nonminimum-phase plants, tracking design should be done by optimal model matching such as H^∞ control. In both cases, the identified model can be used to track an output that differs from the threshold.

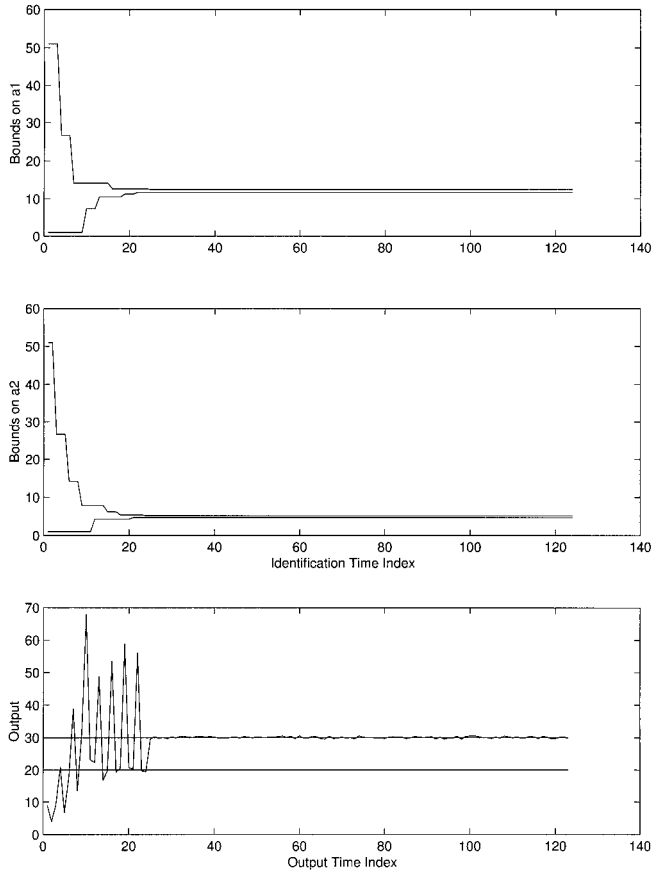


Fig. 2. Tracking control with one binary sensor.

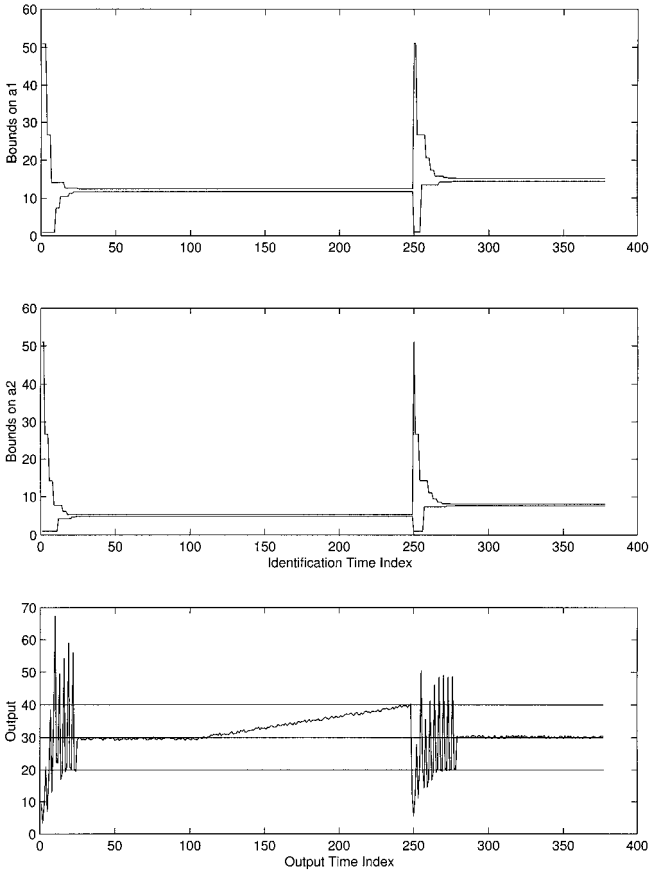


Fig. 3. Tracking control with two binary sensors.

At the end of the identification phase, the centers of the uncertainty sets for the parameters a_1 and a_2 are used as the estimates $\widehat{a}_1$ and $\widehat{a}_2$. Then, the control for tracking y^* is calculated by

$$u(t) = \frac{1}{\widehat{a}_1}(y^* - \widehat{a}_2 u(t-1)).$$

Fig. 2 shows the uncertainty sets (upper and lower bounds) on a_1 and a_2 during the identification phase, and outputs in both identification and tracking control phase. It is seen that after binary-sensor identification, one can achieve tracking control, even when the desired output value is far away from the sensor threshold. It should be noted that the large fluctuations on y during the identification phase is unavoidable due to the large prior uncertainty set [1,51] assumed on parameters.

Example 2: **Tracking Control Using Two Binary Sensors.** It is noted that if system parameters in Example 1 are varying with time, then parameter drifting may cause tracking performance to deteriorate without being detected by the sensor. In Example 1, escaping of y toward infinity will not be detected since $y^* > C$. One possible remedy is to employ two binary sensors with the thresholds $C_1 < y^* < C_2$. When parameter drifting causes $y(t)$ across these thresholds, reidentification will be employed. This is illustrated later.

Suppose that the parameters of the system change with time, drifting slowly from the current values $a_1 = 12$, $a_2 = 5$ to the new values $a_1 = 20$, $a_2 = 10$. A new binary sensor is added with threshold $C_2 = 40$. Fig. 3 shows the impact of this parameter variation on the output y. When y increases to cross the threshold C_2, identification is employed again. This identification captures the new values of the parameters and improves tracking performance.

VII. Conclusion

Identification with binary-valued sensors is of practical importance and theoretical interest. The main findings of this paper reveal many distinctive aspects in such identification problems from traditional identification. Furthermore, the relationships between time complexity and the Kolmogorov entropy and between stochastic and deterministic frameworks in their identifiability provide new understandings of fundamental links in identification problems.

Binary-sensor identification introduced in this paper was initially motivated by several typical industrial control problems. A limited investigation to different application areas has generated many examples in a broader range of applications, including biology, economy, medicine, and further links to other branches of information processing methodologies. It is important to motivate further investigations by studying these application areas vigorously and rigorously.

Potential extensions of the work reported in this paper are numerous, including from the MA models to ARMAX models, from linear systems to Wiener or Hammerstein nonlinear models, and from input–output observations to blind identification.

APPENDIX

A. Proof of Theorem 13

1) The identification error and time complexity follow directly from Theorems 6 and 10 with $n = 1$. As for the optimal input, notice that starting from the uncertainty $[\underline{a}(t), \overline{a}(t)]$ an input $u(t)$ defines a testing point $C/u(t)$ on a. The optimal worst-case input is then obtained by placing the testing point at the middle. That is
$$\frac{C}{u(t)} = \frac{1}{2}(\underline{a}(t) + \overline{a}(t))$$
which leads to the optimal input and result in posterior uncertainty sets.
2) When the input is bounded by $u(t) \in [-u_{\max}, u_{\max}]$, the testing points cannot be selected in the interval $[-C/u_{\max}, C/u_{\max}]$. Consequently, this uncertainty set cannot be further reduced by identification. Furthermore, by using $u(1) = -u_{\max}$ and $u(2) = u_{\max}$ as the first two input values, a can be determined as belonging uniquely to one of the three intervals: $[\underline{a}(0), -C/u_{\max})$, $[-C/u_{\max}, C/u_{\max}]$, $[C/u_{\max}, \overline{a}(0)]$. By taking the worst-case scenario of $\overline{a}(0) - C/u_{\max} = \varepsilon_0 - (\delta_h - \delta_l)$, the time complexity for reducing the remaining uncertainty to ε is $\lceil \log((\varepsilon_0 - (\delta_h - \delta_l))/\varepsilon) \rceil$. This leads to the upper bound on $N(\varepsilon)$. The lower bound follows from Theorem 6 with $n = 1$. □

B. Proof of Theorem 14

1) Since $u(t) > 0$, the relationship (25) can be written as $a = (y(t) - d(t))/u(t)$.
The observation outcome $y(t) \geq C$ will imply that
$$a \geq \frac{C - d(t)}{u(t)} \geq \frac{C - \delta}{u(t)}$$
which will reduce uncertainty from $a \in [\underline{a}(t-1), \overline{a}(t-1)]$ to $[(C-\delta)/u(t), \overline{a}(t-1)]$ with error $e_1(t) = \overline{a}(t-1) - (C-\delta)/u(t)$. Similarly, $y < C$ implies $a < C + \delta/u(t)$ and $a \in [\underline{a}(t-1), (C+\delta)/u(t)]$ with $e_2(t) = (C+\delta)/u(t) - \underline{a}(t-1)$. In a worst-case scenario, $e(t) = \max\{e_1(t), e_2(t)\}$. Consequently, the optimal $u(t)$ can be derived from $\inf_{u(t)} e(t)$. Hence, the optimal $u(t)$ is the one that causes $e_1(t) = e_2(t)$, namely
$$\frac{C+\delta}{u(t)} - \underline{a}(t-1) = \overline{a}(t-1) - \frac{C-\delta}{u(t)}$$
or
$$u(t) = \frac{2C}{\underline{a}(t-1) + \overline{a}(t-1)}.$$
The optimal identification error is then
$$\begin{aligned} e(t) &= \frac{(C+\delta)(\overline{a}(t-1) + \underline{a}(t-1))}{2C} - \underline{a}(t-1) \\ &= \left(\frac{1}{2} + \frac{\delta}{2C}\right)(\overline{a}(t-1) + \underline{a}(t-1)) - \underline{a}(t-1) \\ &= \frac{1}{2}e(t-1) + \frac{\delta}{2C}(\overline{a}(t-1) + \underline{a}(t-1)). \end{aligned}$$
2) We prove $\overline{a}(t)/\underline{a}(t) \geq (1+\delta/C)/(1-\delta/C)$ by induction. Suppose that $\overline{a}(t-1)/\underline{a}(t-1) \geq (1+\delta/C)/(1-\delta/C)$. Then, we have $u(t)\overline{a}(t-1) \geq C + \delta$ and $u(t)\underline{a}(t-1) \leq C - \delta$, which, respectively, leads to $\overline{a}(t)/\underline{a}(t) = (u(t)\overline{a}(t-1))/(C-\delta) \geq (1+\delta/C)/(1-\delta/C)$ in the case where $s(t) = 1$, and $\overline{a}(t)/\underline{a}(t) = (C+\delta)/(u(t)\underline{a}(t-1)) \geq (1+\delta/C)/(1-\delta/C)$ in the case where $s(t) = 0$. Thus, by the initial condition that $\overline{a}(0)/\underline{a}(0) \geq (1+\delta/C)/(1-\delta/C)$ we have $\overline{a}(t)/\underline{a}(t) \geq (1+\delta/C)/(1-\delta/C)$ for all $t \geq 1$.
By $\overline{a}(t-1)/\underline{a}(t-1) \geq (1+\delta/C)/(1-\delta/C)$, we have $u(t)\underline{a}(t-1) \leq C - \delta$ and $u(t)\overline{a}(t-1) \geq C + \delta$, which gives $\overline{a}(t) = \overline{a}(t-1)$ and $\underline{a}(t)/\underline{a}(t-1) = (C-\delta)/(u(t)\underline{a}(t-1)) \geq 1$ in the case where $s(t) = 1$, and $\underline{a}(t) = \underline{a}(t-1)$ and $\overline{a}(t)/\overline{a}(t-1) = (C+\delta)/(u(t)\overline{a}(t-1)) \leq 1$ in the case where $s(t) = 0$. Thus, $\{\underline{a}(t)\}$ is monotonely increasing and $\{\overline{a}(t)\}$ is monotonely decreasing.
Furthermore, by $\underline{a}(t)/\underline{a}(t-1) \geq 1$ and $\overline{a}(t-1)/\overline{a}(t) \geq 1$ we obtain $(\underline{a}(t)\overline{a}(t-1))/(\overline{a}(t)\underline{a}(t-1)) \geq 1$, i.e., $\overline{a}(t-1)/\underline{a}(t-1) \geq \overline{a}(t)/\underline{a}(t)$. Hence, $\{\overline{a}(t)/\underline{a}(t)\}$ is monotonely decreasing.
The dynamic expression (32) can be modified as
$$e(t) = \frac{1}{2}\left(1 - \frac{\delta}{C}\right)e(t-1) + \frac{\delta}{C}\overline{a}(t-1) \tag{40}$$
or
$$e(t) = \frac{1}{2}\left(1 + \frac{\delta}{C}\right)e(t-1) + \frac{\delta}{C}\underline{a}(t-1). \tag{41}$$
By taking $t \to \infty$ on both sides of (40) and (41) we get $\overline{a}(\infty) = ((C+\delta)/2\delta)e(\infty)$ and $\underline{a}(\infty) = ((C-\delta)/2\delta)e(\infty)$. This leads to $\lim_{t\to\infty}\overline{a}(t)/\underline{a}(t) = (1+\delta/C)/(1-\delta/C)$.
3) From (32), it follows that the uncertainty is reducible if and only if $(\delta/C)(\overline{a}(t-1)+\underline{a}(t-1)) < e(t-1) = \overline{a}(t-1) - \underline{a}(t-1)$. This is equivalent to $\overline{a}(t-1)/\underline{a}(t-1) > (1+\delta/C)/(1-\delta/C)$. □

C. Proof of Theorem 15

1) From (40) and the monotone decreasing property of $\overline{a}(t)$, we have
$$e(t) \geq \alpha_1^t e(0) + \frac{\delta\overline{a}(t-1)}{C}\sum_{i=0}^{t-1}\alpha_1^i$$
and from (41) and the monotone increasing property of $\underline{a}(t)$
$$e(t) \leq \alpha_2^t e(0) + \frac{\delta\underline{a}(t-1)}{C}\sum_{i=0}^{t-1}\alpha_2^i.$$
The results follow from $\sum_{i=0}^{t-1}\alpha_1^i = (1-\alpha_1^t)/(1-\alpha_1)$, $\sum_{i=0}^{t-1}\alpha_2^i = (1-\alpha_2^t)/(1-\alpha_2)$, $1-\alpha_1 = \alpha_2$ and $\underline{a}(t) \leq a \leq \overline{a}(t)$.
2) From item 2 of Lemma 14, it follows that the error $e(t) = \overline{a}(t) - \underline{a}(t)$ is monotonely decreasing. Thus, the upper bound on the time complexity is obtained by solving the inequality for the smallest N satisfying
$$e(N) \leq \alpha_2^N\varepsilon_0 + \frac{\delta a(1-\alpha_2^N)}{C\alpha_1} \leq \varepsilon.$$
Similarly, the lower bound can be obtained by calculating the largest N satisfying
$$\varepsilon \leq \alpha_1^N\varepsilon_0 + \frac{\delta a(1-\alpha_1^N)}{C\alpha_2} \leq e(t).$$

3) This follows from (33) and Item 2 of Lemma 14, which implies the existence of $\lim_{t\to\infty} e(t)$.
4) From the last two lines of the proof of Item 2 of Lemma 14 it follows $\overline{a}(\infty) = ((C+\delta)/2\delta)e(\infty)$ and $\underline{a}(\infty) = ((C-\delta)/2\delta)e(\infty)$. This, together with (34), gives (35). □

D. Proof of Theorem 16

1) The results follow from the definition of m_t and Theorem 15, with δ replaced by ηm_t.
2) From (37) and (36), it follows that the uncertainty is reducible if and only if $\eta m_t/u(t) < (1/2)e(t-1) = (1/2)(\overline{a}(t-1) - \underline{a}(t-1))$. This is equivalent to $\eta < (1/2)(\overline{a}(t-1) - \underline{a}(t-1))$ or $\overline{a}(t-1) > \underline{a}(t-1) + 2\eta$, since $m_t/u(t) \geq 1$.
3) By (37), we have

$$e(t) = \frac{1}{2}\left(1 + \frac{\eta_t}{C}\right) e(t-1) + \frac{\eta m_t}{C}\underline{a}(t-1) \tag{42}$$

and

$$e(t) = \frac{1}{2}\left(1 - \frac{\eta m_t}{C}\right) e(t-1) + \frac{\eta m_t}{C}\overline{a}(t-1). \tag{43}$$

Further, from $\underline{a}(t) \leq a \leq \overline{a}(t)$ for all $t \geq 0$

$$e(t) \leq \beta_2(t)e(t-1) + \frac{\eta m_t}{C}a$$

and

$$e(t) \geq \beta_1(t)e(t-1) + \frac{\eta m_t}{C}a.$$

Then, the inequalities in (38) can be obtained by iterating the previous two inequalities in t.
4) Since for all $t \geq 1$, $\overline{a}(0) \geq \overline{a}(t) \geq \underline{a}(t) \geq \underline{a}(0)$

$$u(t) = \frac{2C}{\underline{a}(t-1) + \overline{a}(t-1)} \leq \frac{C}{\underline{a}(0)}$$

which implies that $C/\overline{a}(0) \leq k_t \leq C/\underline{a}(0)$. This leads to $\beta_1(t) \geq \beta_1 = (1/2)(1 - \eta/\underline{a}(0))$ and $\beta_2(t) \leq \beta_2 = (1/2)\,(1 + \eta/\underline{a}(0))$. Hence

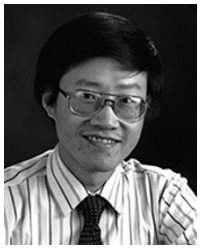
$$\beta_1 e(t-1) + \frac{\eta a}{\overline{a}(0)} \leq e(t) \leq \beta_2 e(t-1) + \frac{\eta a}{\underline{a}(0)} \quad \text{for all t} \geq 1. \tag{44}$$

As a result, the inequalities of Theorem 15 can be adopted here to get (39). □

REFERENCES

[1] H. Abut, Ed., "Vector Quantization," in *IEEE Press Selected Reprint Series*. Piscataway, NJ: IEEE Press, 1990.
[2] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, 1968.
[3] E. R. Caianiello and A. de Luca, "Decision equation for binary systems: Application to neural behavior," *Kybernetik*, vol. 3, pp. 33–40, 1966.
[4] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhäuser, 1991.
[5] H. F. Chen and G. Yin, "Asymptotic properties of sign algorithms for adaptive filtering," *IEEE Trans. Automat. Contr.*, vol. 48, pp. 1545–1556, Sept. 2003.
[6] M. A. Dahleh, T. Theodosopoulos, and J. N. Tsitsiklis, "The sample complexity of worst-case identification of FIR linear systems," *Syst. Control Lett.*, vol. 20, 1993.
[7] C. R. Elvitch, W. A. Sethares, G. J. Rey, and C. R. Johnson Jr, "Quiver diagrams and signed adaptive fiters," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 30, pp. 227–236, Feb. 1989.
[8] E. Eweda, "Convergence analysis of an adaptive filter equipped with the sign-sign algorithm," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1807–1811, Oct. 1995.
[9] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1968, vol. I.
[10] ——, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1971, vol. II.
[11] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 191–199, Mar. 1984.
[12] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
[13] K. Gopalsamy and I. K. C. Leung, "Convergence under dynamical thresholds with delays," *IEEE Trans. Neural Networks*, vol. 8, pp. 341–348, Apr. 1997.
[14] R. G. Hakvoort and P. M. J. Van den Hof, "Consistent parameter bounding identification for linearly parameterized model sets," *Automatica*, vol. 31, pp. 957–969, 1995.
[15] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, pp. 500–544, 1952.
[16] P. R. Kumar, "Convergence of adaptive control schemes using least-squares parameter estimates," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 416–424, Apr. 1990.
[17] A. N. Kolmogorov, "On some asymptotic characteristics of completely bounded spaces," *Dokl. Akad. Nauk SSSR*, vol. 108, pp. 385–389, 1956.
[18] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.
[19] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1987.
[20] M. Milanese and G. Belforte, "Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors: Linear families of models and estimators," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 408–414, Apr. 1982.
[21] M. Milanese and A. Vicino, "Information-based complexity and nonparametric worst-case system identification," *J. Complexity*, vol. 9, pp. 427–446, 1993.
[22] ——, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, pp. 997–1009, 1991.
[23] K. Pakdaman and C. P. Malta, "A note on convergence under dynamical thresholds with delays," *IEEE Trans. Neural Networks*, vol. 9, pp. 231–233, Feb. 1998.
[24] D. Pollard, *Convergence of Stochastic Processes*. New York: Springer-Verlag, 1984.
[25] K. Poolla and A. Tikku, "On the time complexity of worst-case system identification," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 944–950, May 1994.
[26] K. Sayood, *Introduction to Data Compression*, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2000.
[27] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.
[28] L. Schweibert and L. Y. Wang, "Robust control and rate coordination for efficiency and fairness in ABR traffic with explicit rate marking," *Comput. Commun.*, vol. 24, pp. 1329–1340, 2001.
[29] J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski, *Information-Based Complexity*. New York: Academic, 1988.
[30] D. C. N. Tse, M. A. Dahleh, and J. N. Tsitsiklis, "Optimal asymptotic identification under bounded disturbances," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 1176–1190, Aug. 1993.
[31] S. R. Venkatesh and M. A. Dahleh, "Identification in the presence of classes of unmodeled dynamics and noise," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1620–1635, Dec. 1997.
[32] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[33] ——, *The Nature of Statistical Learning Theory*, 2nd ed, ser. Statistics for Engineering and Information Science. New York: Springer-Verlag, 1999.
[34] M. Vidyasagar, *Learning and Generalization: With Applications to Neural Networks*, 2nd ed. New York: Springer-Verlag, 2003.
[35] L. Y. Wang, "Persistent identification of time varying systems," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 66–82, Jan. 1997.
[36] L. Y. Wang, I. V. Kolmanovsky, and J. Sun, "On-line identification lean NOx trap in GDI engines," presented at the 2000 Amer. Control Conf., Chicago, IL, June 2000.
[37] L. Y. Wang, Y. Kim, and J. Sun, "Prediction of oxygen storage capacity and stored NOx using HEGO sensor model for improved LNT control strategies," presented at the 2002 ASME Int. Mechanical Engineering Congr. Exposition, New Orleans, LA, Nov. 17–22, 2002.

[38] L. Y. Wang and L. Lin, "On metric dimensions of discrete-time systems," *Syst. Control Lett.*, vol. 19, pp. 287–291, 1992.
[39] L. Y. Wang and G. Yin, "Toward a harmonic blending of deterministic and stochastic frameworks in information processing," in *Robustness in Identification and Control*. New York: Springer-Verlag, 1999, LNCS, pp. 102–116.
[40] ——, "Persistent identification of systems with unmodeled dynamics and exogenous disturbances," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 1246–1256, July 2000.
[41] ——, "Closed-loop persistent identification of linear systems with unmodeled dynamics and stochastic disturbances," *Automatica*, vol. 38, no. 9, 2002.
[42] G. Yin, V. Krishnamurthy, and C. Ion, "Iterate-averaging sign algorithms for adaptive filtering with applications to blind multiuser detection," *IEEE Trans. Inform. Theory*, vol. 49, pp. 657–671, Mar. 2003.
[43] G. Zames, "On the metric complexity of causal linear systems ε-entropy and ε-dimension for continuous time," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 222–230, Apr. 1979.
[44] G. Zames, L. Lin, and L. Y. Wang, "Fast identification n-widths and uncertainty principles for LTI and slowly varying systems," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 1827–1838, Sept. 1994.



Le Yi Wang (S'85–M'89–SM'01) received the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1990.

He is a Professor of Electrical and Computer Engineering at Wayne State University, Detroit, MI. His research interests are in the areas of robust control, system identification, complexity and information, adaptive systems, hybrid and nonlinear systems, information processing and learning, as well as automotive and medical applications of control and information processing methodologies.

Dr. Wang was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and is an Editor of the *Journal of System Sciences and Complexity*.

Ji-Feng Zhang (M'92–SM'97) received the B.S. degree in mathematics from Shandong University, Shandong, China, in 1985 and M.S. and Ph.D. degrees in control theory and stochastic systems from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), Beijing, China, in 1988 and 1991, respectively.

Since 1985, he has been with ISS-CAS, where he is now a Professor. His current research interests are adaptive control, stochastic systems and descriptor systems.

Dr. Zhang received the National Science Fund for Distinguished Young Scholars from the National Science Foundation of China in 1997 and the First Prize of the Young Scientist Award of the CAS in 1995.

G. George Yin (S'87–M'87–SM'96–F'02) received the M.S. degree in electrical engineering and Ph.D. in applied mathematics from Brown University, Providence, RI, in 1987.

He jointed the Mathematics Department at Wayne State University, Detroit, MI, in 1987. He served on the Editorial Boards of *Stochastic Optimization & Design*, *Mathematical Review Date Base Committee*, and various conference program committees. He was the Editor of the *SIAM Activity Group on Control and Systems Theory Newsletters*, SIAM Representative to the 34th Confeence on Decision and Control, Chair of the 1996 AMS-SIAM Summer Seminar in Applied Mathematics. He is Chair of 2003 AMS-IMS-SIAM Summer Research Conference "Mathematics of Finance."

Dr. Yin was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.